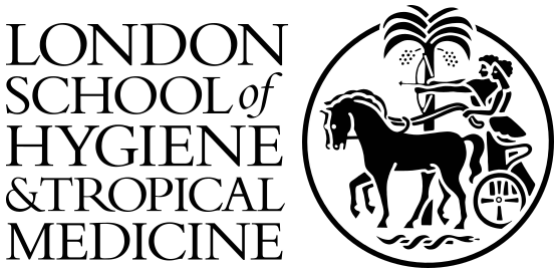


Rediscovering the natural history of tuberculosis using modelling to combine historical and contemporary data

Alexandra Suzanne Richards

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

September 2023

Department of Infectious Disease Epidemiology

Faculty of Epidemiology and Population Health

London School of Hygiene & Tropical Medicine

Funded by ERC

Declaration by the candidate

I, Alexandra Suzanne Richards, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed

21 September 2022

Abstract

Tuberculosis has been a major cause of morbidity and mortality globally for hundreds of years and remains so today. In spite of evidence pointing to significant burdens of asymptomatic but infectious (subclinical) disease, there is very little knowledge on the rates and directions of progression. Models reflect this lack of information by simplifying the structure of disease, with very little data informing the direct transitions between states.

This thesis seeks to understand the natural history of tuberculosis from first infection to death or recovery with a focus on accurately representing the spectrum of disease through pathology, bacteriology, and symptom presentation in order to better parameterise models.

A systematic review of TB research following untreated cohorts found research from the first half of the 1900s describing both progression and regression of disease in terms of bacteriology, radiology, and symptoms.

I used this data to construct and parameterise a model of pulmonary TB disease, extending the commonly used "active disease" compartment into three stages of disease: minimal, subclinical, and clinical. I then simulated cohorts through disease to discover that 50% of people with subclinical disease may never present symptomatically, and that, despite finding a median duration of infectious disease of 12 months, after 5 years up to 20% of people starting with infectious disease could still be living with infectious TB.

I then used the new model structure to compare the impact different screening tests could provide. With the current limited data on test performance for minimal and subclinical disease, mass x-ray screenings, as used in the first half of the 20th century, are the most impactful without screening every person bacteriologically. A test that could accurately confirm diagnosis of minimal disease could increase the impact of x-ray screening even further.

Finally, I parameterised an extension to the TB disease model, including progression from infection to disease. The data for this came from previously unused cohorts in the original systematic review. I found that rates of progression to disease are low in comparison to rates of recovery from infection. The best parameterisations maintained a well-used TB model structure of an intermediate state between infection and disease, along with rapid progression that by-passes this intermediate state.

In conclusion, I have created a more complex TB model structure that differentiates between presentations of disease, and based the parameterisation, from infection to death, on real world data. I have used this new structure to evaluate the impact of different screening tests, but there

are countless other applications of this model structure for future implementations, including calculating better estimates of the global burden of TB and designing trials.

Acknowledgments

The last four years have had ups and downs and there have been so many people who have been there for me, and I owe them all thanks.

To my primary supervisor, Rein Houben, for giving me this opportunity, guiding me through these four years of PhD and helping me grow as a researcher. I've learnt so much and I have appreciated your support through everything this PhD has thrown at me. To Nicky McCreesh and Alison Grant, for your expertise and guidance, pushing me to improve my work. To Jon Emery and Katherine Horton, for being happy to answer all my questions. To the review team, who spent hours outside of their own work to help me trawl through hundreds of studies from the first half of the 20th century. To all members of the TB modelling group and CMMID for being friendly, supportive, and approachable.

To my friends who have provided a place of sanity outside the work. To Constanza, who has been unwaveringly supportive from thousands of miles away, first in California, now in New Zealand. You'll be started and done with your PhD before you know it, and you'll do amazingly. To Catherine, for introducing me to the best chocolate fudge cake in the New Forest through our cycling adventures, and for always being up for a chat. To Alex, for music, book, knitting, and food recommendations, for being happy to sign up to crazy sporting events with me and cheerleading our way through training for them. To my friends at LSHTM in my office and from my cohort, Rutendo, Judy, Will, Sophie, and Ellie, who helped break up the early days with chats and hot drinks. To all the other friends I've made and stayed in touch with throughout my PhD, for being supportive and providing an outlet. A special shout out to the friends who've sat on voice chats, all around the world, whilst we work together.

To my family, who have supported me throughout my life. To my aunt and uncle, who regularly welcomed me to Sunday lunch when I lived in London. To my siblings, who put on their best interested faces when I talk about an issue I'm having with my work for the third time in a day. For the comfort food they have made and bought me during the writing up process, and for putting up with my questionable film choices (that I know you love really). To my parents, who have always encouraged me to pursue interests and believed that I can achieve whatever I put my mind to. For their support over the course of this degree, and all my prior education, I wouldn't have been able to do this without you. And finally, to Benji, despite some very early mornings as a puppy, for the walk and play breaks, and the company whilst I wrote this thesis, however late or early I was working.

Table of Contents

List of Figures

(Figures in appendices not included)

List of Tables

(Tables in appendices not included)

List of abbreviations

ABM	Agent Based Model
CI	confidence interval
CRP	C-reactive protein
CT	computed tomography
cxr	chest radiography
DOTS	directly observed treatment, short-course
HIV	human immunodeficiency virus
IBM	Individual Based Model
IGRA	inferon-gamma release assay
M.tb	Mycobacterium tuberculosis
MRI	magnetic resonance imaging
NNT	number needed to test
PET	positron emission tomography
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
SEIR	Susceptible, Exposed, Infectious, Recovered
TB	Tuberculosis
TNF	tumour necrosis factor
TPP	target product profile
TST	tuberculin skin test
UI	uncertainty interval
WHO	World Health Organization
Xpert	Xpert MTB/RIF

Chapter 1 Introduction

In this chapter I will briefly explain the background and rationale for this thesis, through a summary of the research gaps that justify the aims and objectives of this thesis. I will then outline the thesis structure and specify my contribution to all pieces of work.

1.1 Background

Tuberculosis (TB) is a major source of morbidity and mortality, with an estimated 10 million new cases a year and causing 1.5 million deaths.[1] Until 2020, and the start of the coronavirus (COVID-19) pandemic, TB was the leading cause of death from a single agent, exceeding deaths from HIV/AIDS.[1]

Recent prevalence surveys have used a combination of x-ray and symptom screening to determine eligibility for bacteriological testing, and found that a significant proportion of people who tested positive bacteriologically, had screened negative on symptoms.[2,3] Whilst the proportion of people with asymptomatic bacteriologically positive disease varied widely (36% to 80%), there was no correlation between factors such as TB or HIV prevalence, or the patient diagnostic rate within the country.[2,3] Asymptomatic disease is well documented within clinical presentations of disease, although has lapsed in and out of definitions for case detection.[3,4]

Most models for pulmonary TB are based around a structure that assumes a single state for exposure (latent infection) and a single state for the resulting illness (active disease).[5,6] Often, the latent infection is split into fast and slow compartments to represent the known artefact of TB infection that the risk of progression is fastest in the first few years with a much slower risk in later years.[6] The active disease compartment is also often split into two, based on smear status.[11]

These natural history models are parameterised using data from a very select group of papers, or calibrated to external targets such as the number of people receiving treatment or dying from disease.[12–15] However with the effective treatment regimens available today, it is not possible (or ethical) to collect data to directly parameterise the models. There are many more studies from the pre-treatment era that contain data on how infection and disease progresses, but no systematic review of the available literature has been conducted to collect this data.

Although it is known clinically that disease is a spectrum and allows for natural regression, this is not well represented in models, in part due to a lack of sufficient data. However, there is also little agreement on structure or terminology to refer to different points within the spectrum.[16–18]

If all case detection and treatment was offered on smear status, and people were offered smears regardless of symptom status, based instead on risk factors such as local or occupational prevalence, known family members with TB, or known previous infection, then creating policy based on a model structure that is defined by those differences is useful. However, in reality, tests are restricted due to cost and likely (or unlikely) positivity. Therefore, people presenting with TB specific symptoms are offered a test, whereas people without symptoms will not even seek healthcare with any query that will initiate a TB test. The fact that most models do not then differentiate between disease with and without symptoms means that predictions of impact of healthcare interventions on "active disease" will have half the effect in real life due to half the people with active disease not presenting to healthcare facilities.[3]

1.2 Aims

This thesis has two main aims:

1. Create a model of pulmonary TB in adults from infection to death that incorporates the spectrum of disease with data-driven parameter estimates
2. Describe how an updated model structure could affect our understanding of best practices for finding and treating people with disease

1.3 Objectives

To address aim 1, there are three objectives:

3. Conduct a systematic review of pre-treatment literature to find data on progression of pulmonary tuberculosis in adult cohorts
4. Develop a natural history model of the spectrum of TB disease and calibrate the parameters to data from the systematic review
5. Extend the natural history model to include progression from infection and calibrate to other data from the systematic review

To address aim 2, there are two further objectives:

4. Simulate an untreated population through disease to quantify the progression and regression pathways taken
5. Compare the effectiveness of different screening programmes to detect infectious disease on simulated treated populations

1.4 Thesis structure

This thesis is structured in a research paper format in accordance with the guidelines of the London School of Hygiene and Tropical Medicine. Three research papers are included.

Chapter 1 contains an introduction to the research gaps, a statement of the aims and objectives of the work, and information on author contributions, ethics, and funding.

Chapter 2 is a background literature review providing the relevant information needed for each subsequent section of research and how they link together. This chapter identifies the research gaps highlighted in section 1.1.

In Chapter 3, I address objective 1 with a report on a systematic review of historical data to understand the spectrum of the natural history of disease.

Chapter 4 contains paper 1, which has been uploaded as a preprint on medRxiv. It is a report on the estimation of parameters for the spectrum of TB disease and an analysis of what these parameters mean for pathways through disease. This chapter addresses objectives 2 and 4.

Chapter 5 contains an unpublished final draft of paper 2. It is a comparison of different screening tools as an alternative to population level symptom screening and addresses objective 5.

Chapter 6 is a description of the fitting work undertaken to join the spectrum of disease with the progression from initial infection to disease, addressing objective 3.

Finally, Chapter 7 is a discussion of all work presented throughout the thesis and the future steps that can be taken following this work.

1.5 Contributions of the author

Chapter 2: I conducted the literature review and drafted the chapter that presents the background material for this thesis.

Chapter 3: The systematic review of historical literature was originally conceived by Hanif Esmail and Rein Houben. I helped develop the aims and the protocol with Bianca Sossen, Hanif Esmail, and Rein Houben. I have done the largest share of the work at all stages of the review, screening the titles, finding and reviewing the full texts, assessing bias and extracting data. This work has been shared with the original group and through a team of nine other people who I guided and managed; three helped source the full texts and search the index medicus, six assisted with the full text screening, bias screening, and data extraction. Every study that has

had data extracted, I either extracted myself or approved the extraction by someone else. I prepared the tables and figures, and wrote the chapter.

Chapter 4: This chapter uses the data from Chapter 3 and analyses it through a model. The initial idea was conceived by Rein Houben. I developed the aim of the manuscript, and the structure of the model with Rein Houben. I designed the fitting methodology with the help of Jon Emery and Katherine Horton. I coded the models, excluded the appropriate data from the systematic review and produced the figures and results. I drafted the manuscript and incorporated comments from all co-authors and submitted to medRxiv prior to being able to submit to a journal. This paper will be submitted to a journal alongside a separate write up of Chapter 3 but I will oversee the submission process and respond to peer review with the necessary revisions of the work.

Chapter 5: I created the aims and objectives of this chapter with Rein Houben. I designed the model framework, coded the model, conducted the analyses, created the figures, and drafted the manuscript. I will oversee the submission process to a journal and revise the manuscript as necessary in response to peer review.

Chapter 6: This work is an extension of my work in Chapter 4 but with additions that I helped conceive jointly with Katherine Horton, Jon Emery, and Rein Houben. I identified potential data for Katherine Horton to use and was involved in discussions of how that data should be used. Jon Emery ran initial tests to understand the best model structure to fit the new data. I joined the structure Jon created with my work from Chapter 4 and calibrated the model structure to the data. I have created all tables and figures included here, and drafted the chapter. This will not be submitted to a journal in this form, but is written in the same style to maintain consistency between chapters.

Chapter 7: I summarised the findings from chapters 3-6 and discussed the strengths, limitations, and implications for future works.

1.6 Ethical considerations

All data are from publicly available sources, presented in published books or peer reviewed journals.

1.7 Funding

This research was funded primarily through an ERC Starting Grant, with external funding from TB MAC to support the wider systematic review.

1.8 References

1 World Health Organization. *Global tuberculosis report 2021*. World Health Organization: Geneva, 2021https://apps.who.int/iris/handle/10665/346387.

2 Onozaki I, Law I, Sismanidis C, Zignol M, Glaziou P, Floyd K. National tuberculosis prevalence surveys in asia, 1990-2012: An overview of results and lessons learned. *Trop Med Int Health* 2015; **20**: 1128–1145.

3 Frascella B, Richards AS, Sossen B, Emery JC, Odone A, Law I *et al.* Subclinical tuberculosis disease - a review and analysis of prevalence surveys to inform definitions, burden, associations and screening methodology. *Clinical Infectious Diseases* 2020. doi:10.1093/cid/ciaa1402.

4 National Tuberculosis Association. *Diagnostic standards and classification of tuberculosis.* 1940 ed. New York, N.Y., 1940http://hdl.handle.net/2027/coo.31924089435949.

5 Colijn C, Cohen T, Murray M. International Symposium on Mathematical and Computational Biology. WORLD SCIENTIFIC: Manaus, Brazil, 2007, pp 123–148.

6 Menzies NA, Wolf E, Connors D, Bellerose M, Sbarra AN, Cohen T *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; **18**: e228–e238.

7 Ragonnet R, Flegg JA, Brilleman SL, Tiemersma EW, Melsew YA, McBryde ES *et al.* Revisiting the Natural History of Pulmonary Tuberculosis: a Bayesian Estimation of Natural Recovery and Mortality rates. *Clinical Infectious Diseases* 2020. doi:10.1093/cid/ciaa602.

8 Menzies NA, Cohen T, Hill AN, Yaesoubi R, Galer K, Wolf E *et al.* Prospects for Tuberculosis Elimination in the United States: Results of a Transmission Dynamic Model. *American Journal of Epidemiology* 2018. doi:10.1093/aje/kwy094.

9 Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce de Leon A, Daley CL *et al.* Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet (London, England)* 1999; **353**: 444–449.

10 Tostmann A, Kik Sandra V, Kalisvaart Nico A, Sebek Maruschka M, Verver S, Boeree Martin J *et al.* Tuberculosis Transmission by Patients with Smear-Negative Pulmonary Tuberculosis in a Large Cohort in The Netherlands. *Clinical Infectious Diseases* 2008; **47**: 1135–1142.

11 Houben RMGJ, Lalli M, Sumner T, Hamilton M, Pedrazzoli D, Bonsu F *et al.* TIME impact a new user-friendly tuberculosis (TB) model to inform TB policy decisions. *BMC Medicine* 2016; **14**. doi:10.1186/s12916-016-0608-4.

12 Ferebee SH. Controlled chemoprophylaxis trials in tuberculosis. A general review. *Bibl Tuberc* 1970; **26**: 28–106.

13 Comstock GW. Epidemiology of tuberculosis. *The American Review of Respiratory Disease* 1982; **125**: 8–15.

14 Styblo K. In: Flenley DC, Petty TL (eds). *Advances in Respiratory Medicine*. Churchill Livingstone, Edinburgh, 1986, pp 77–108.

15 Styblo K. Selected papers (vol 24). *KNCV Tuberculosis Foundation - Selected papers* 1991.

16 Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG *et al.* Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. *Clinical Microbiology Reviews* 2018; **31**: 24.

17 Lin PL, Flynn JL. The End of the Binary Era: Revisiting the Spectrum of Tuberculosis. *The Journal of Immunology* 2018; **201**: 2541–2548.

18 Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC *et al.* Tuberculosis. *Nat Rev Dis Primers* 2016; **2**: 16076.

Chapter 2 Background

2.1 Current understanding of TB

It is estimated that annually, 10 million people develop tuberculosis disease (TB), and 1.5 million people die from TB, despite effective treatments existing.[1] This burden is not evenly spread globally, with southern Africa and eastern Asia bearing the greatest burden.[1] Approximately one quarter of the world's population are estimated to have been, either recently or distally, infected with the causative agent *Mycobacterium tuberculosis* (*M.tb*), again not uniformly distributed globally.[2] After infection, there is about a 10% lifelong risk of progressing to disease.[3] About 90% of the people who develop TB disease are adults, and most develop pulmonary disease, although TB can develop in almost any part of the body.[1]

M.tb infection and TB disease are understood to exist on a spectrum.[3–5] While terminology can be inconsistent, there is some consensus on the description of stages within the spectrum.[3–5] Some people have no viable *M.tb* in their body, either through never having been infected, or through clearing an infection, with or without having ever progressed to, and subsequently recovered from disease, and they cannot develop disease without being (re)infected.[3–5] Then there is a state of infection where progression to disease is possible without reinfection.[3–5] Conceptually, disease is then split into any number of states to best describe the spectrum; Pai *et al.* use two states, Drain *et al.* use three, and Lin and Flynn use four.[3–5] However many states are used to represent disease, the consensus is that there is an asymptomatic, but bacteriologically-positive state, in which people may not realise they have TB, but are still potentially infectious.[3–5] There is also emerging consensus that progress through the spectrum of disease is not necessarily unidirectional, but can take trajectories undulating through more and less severe disease states.[3–5]

Prevalence surveys have highlighted how common asymptomatic, but bacteriologically-positive disease (usually referred to as subclinical disease) is.[6,7] Some prevalence surveys find as many as 80% of people with bacteriologically-positive disease do not report TB symptoms, with a median of 50% across all national and sub-national surveys from 1990 to 2020.[6,7] Symptoms are typically defined as a cough of at least two or three weeks, but vary widely between surveys including weight loss, fever, night sweats, and haemoptysis.[7] The recently reported prevalence survey in South Africa found that 43% of people with bacteriologically-positive disease reported symptoms at the initial screen, meaning 57% of all disease was subclinical.[8] In contrast, 93% of people detected with bacteriologically-positive disease screened positive with a chest x-ray.[8] Most prevalence surveys use culture for bacteriological confirmation however,

some use molecular tests, and others (mostly those from before 2010) use smear microscopy.[7] Further details on bacteriological testing are discussed in section 2.3.

Whilst the clinical understanding of disease recognises the importance of the spectrum of disease, and policy is starting to change alongside (as seen by the renewed focus on symptom presentation in prevalence surveys), modelling and analysis have been slow to follow.[1,7] For example, global estimates for prevalence and incidence are based on a two-stage model of infection and disease, often referred to as latent infection and active disease.[9–12] However, there are limited data in use at the moment that can describe the progression of disease through the spectrum, hindering uptake of these insights in mathematical models of TB.

2.2 Changing conceptualisation of TB

Recent research in TB has centred on a re-discovery of a spectrum of infection and disease. Whilst it is a different perspective to the current paradigm of a binary of latent infection and active disease, it is not a new one.

TB has been present in human populations for millennia, with claims of cures (of varying successes) dating back to the 14th century, and pathological descriptions of the course of disease dating back to the 17th century.[13] It was only 1720 that Benjamin Martin hypothesised the infectious nature of TB, with another 162 years until Robert Koch managed to isolate the causative agent, *M.tb*.[13] Within the first year of this discovery, improved techniques for identifying *M.tb* had been developed and research into the progression of disease continued throughout the 20th century, with observational studies initiated from occupational and mass screening, and latterly, trials for the efficacy of drug treatments at varying stages of disease.

In 1904, the National Association for the Study and Prevention of Tuberculosis was formed in the USA, and, within a year, a committee was formed to create a standardised definition for the "classification of cases of pulmonary tuberculosis".[14] The definitions were adjusted over time with new understanding and testing, and by June 1917, the first edition of a formal *Diagnostic Standards in Tuberculosis* was published, followed by editions two and three before the end of the year. In parallel, the American Sanatorium Association was also developing a set of diagnostic criteria and in 1912 had already laid out a proposed update to terminology to highlight the difficulty in ascertaining whether a person had truly been cured of TB. They substituted in the phrase "arrested" in place of "cured", and in place of the previous use of "arrested" they used "quiescent".[14] At the point these definitions were first discussed, there was no requirement for radiography, but, in 1922, the American Sanatorium Association's definitions officially included radiography.[15] The definitions first introduced by the American

Sanatorium Association were subsequently used by the National Association from 1913 whilst creating the first edition of their diagnostic standards.[14]

Although there were changes over time, these descriptions persisted, and by the 1940 version of the National Association's standards, the definitions were laid out clearly in terms of radiology, bacteriology, and symptoms, highlighting the variety in disease observed.[14,16]

Arrested: "Constitutional symptoms absent. Sputum, if any, must be concentrated and found microscopically negative for tubercle bacilli. Lesions stationary and apparently healed according to X-ray examination; no evidence of pulmonary cavity."

Quiescent: "No constitutional symptoms. Sputum, if any, may contain tubercle bacilli. Lesions stationary or retrogressive according to X-ray examination; cavity may be present."

Active: "Symptoms unchanged, worse or less severe, but not completely abated. Lesions not completely healed or progressive according to X-ray examination. Sputum almost always contains tubercle bacilli."

Beyond defining disease, the National Association definitions also provided in depth classifications of the extent of pulmonary lesions (minimal, moderately advanced, and far advanced) and symptoms (none, slight, moderate, severe).[16] Instructions were also provided on best practices to demonstrate the presence of tubercle bacilli, through smear, culture, and animal inoculation with a variety of sampling techniques. Following this, they provide a clear definition of what constitutes both arrest and cure of disease, in terms of the number and duration of negative bacteriological results.[16] Further instructions were included on performing and reading tuberculin skin tests (TSTs). A description of the progression of disease was also included, and a shortened description has been included here:[16]

> *"The onset may be symptomless, devoid of abnormal physical signs, and denoted only by the appearance in the roentgenogram … Abnormal physical signs are elicited in less than half the cases at the onset, and then may consist merely of a few rales heard after expiratory cough…In some cases, … the patient may raise a mere fleck of mucopurulent sputum in the morning, …which may reveal tubercle bacilli. … Absorption and fibrosis of this lesion may follow; or rapid, slow or intermittent progression … Local symptoms of cough, expectoration and hemoptysis are, with very few exceptions, indicative of necrosis and ulceration of the lesion into the bronchi (cavity formation). The finding of tubercle bacilli in the sputum has the same significance. … The constitutional symptoms of pulmonary tuberculosis are related closely as a rule to the rate and degree of extension of the lesions…. A disparity of symptoms sometimes is observed. Thus,*

progressive excavation of a pulmonary lesion may be demonstrated while the fever lessens and the patient gains weight. … Extensive pulmonary tuberculosis may heal satisfactorily, leaving the patient more or less disabled …. The numberless clinical episodes of tuberculosis all have a basis and the explanation should always be sought."

This describes a progression of disease prior to the development of symptoms, along with symptoms waning, despite progression in other areas of disease.

Although the National Association definitions were widespread, they were not the only definitions in use and were sometimes expanded upon to create individualised categorisations. Opie and McPhedran split the stages of *M.tb* infection into three groups; tuberculosis recognisable by a tuberculin test, tuberculosis recognisable only by radiography, and clinically manifest tuberculosis.[17] Tuberculosis only recognisable by a tuberculin test would fit with what is now commonly referred to as a latent infection, with "no lesions demonstrable by roentgenographic examination".[17] Tuberculosis recognisable only by radiography was split by age (adult or child type) and severity (calcified nodules or "latent" tuberculosis, split this time by radiography status as defined by the national association).[16,17] With the explicit requirement that this form of disease be detectable only through radiography, it does not fit into the two-stage model of *M.tb* infection and progression where nothing exists between a latent infection and infectious disease.[17] The final type, clinically manifest tuberculosis, is also split into subgroups (without bacilli, with bacilli, non-pulmonary, and suspected), all with the baseline of exhibiting "well-known symptoms and physical signs".[17] This definition also refers back to the National Association definitions, both in the stage of x-ray progression and severity of symptoms to determine the stage of disease.[16,17]

These descriptions and categorisations of TB show that those treating patients with TB knew that there was more to TB than just a state of infection and a state of disease. However somewhere between 1940 and the 2000s the two-stage model became widely used.

From 1964, in the WHO eighth expert committee report on TB, recommendations were made to scale back mass radiography screening, in favour on focussing first on those with symptoms.[18] By 1974, in the WHO ninth expert committee report on TB, the statement on mass radiography was even stronger; it is very expensive and "contributes only a small proportion of the total number of cases found".[19] Thus the focus should remain on people with symptoms and developing outpatient clinics where people could receive free TB treatment.[19] The ten years between 1964 and 1974 also marked a change in the approach to testing for disease.[18,19] While in 1964, they define a "case" as any bacteriologically-confirmed TB with an emphasis that a positive smear result did not need any further investigation, in 1974, part of the rejection of

mass radiography was on account of the few smear-positive cases found through the schemes, implying that only smear testing was used to determine cases.[18,19] Culture testing, and treatment provision based on culture, were not lost, but this change marks the start of the differentiation between smear-positive and smear-negative disease that still persists today.[18,19] This framing of "cases" set up the thinking for the two-stage approach for infection and disease.

Whilst the global approach to the understanding of disease simplified, there were still people discussing how disease could change and fluctuate between presentations. In 1977, Gothi gave a presentation on his understanding of the natural history of TB.[20] His visualisation of this natural history is included as Figure 2.1. He split the natural history into five sections, with the first being the undetectable infection, then the first signs of infection appear in the form of a response to TST.[20] Sometimes this stage comes with symptoms, sometimes not, sometimes this stage progresses rapidly to severe disease or even death, sometimes it recovers to nothing detectable.[20] Beyond this there is potential to develop either "adult-type" pulmonary disease or extra-pulmonary disease, both of which can lead to severe disease and death or recovery, even to the point of no evidence of lesions on the lung.[20] In line with other descriptions, Gothi also believed there was no simple two-stage process of infection and disease, and that the path through infection and disease could be extremely variable.[20]

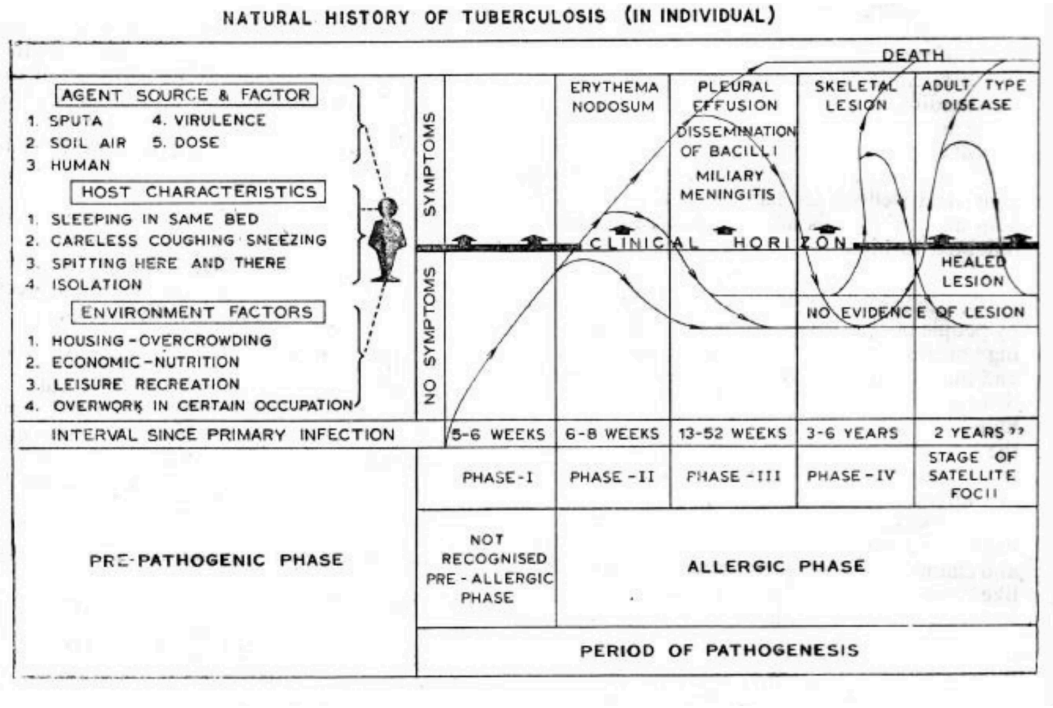


FIGURE 2.1: THE NATURAL HISTORY STRUCTURE DESCRIBED BY GOTHI IN HIS 1977 PRESENTATION[20]

After the recognition of HIV in the early 1980s, and its epidemic spread, years of reductions in TB prevalence were reversed, particularly in Africa, where prevalence of both diseases was high.[21,22] Untreated HIV infection makes a person much more likely to progress to disease either

through a long term infection or re-infection.[21–23] Between 1990 and 2008 there was a global 11% increase in incidence and 40% increase in prevalence of TB, mainly driven by the growing HIV epidemic.[24] This growth threatened to overwhelm healthcare systems.[25] At the same time, the dissolution of the Soviet Union led to a collapse of health systems in the former Soviet republics, creating an increase in both TB incidence and mortality, alongside an increase in drug-resistant TB.[25]

DOTS, short for 'Directly Observed Treatment, Short course', was first implemented in 1993 and as the challenge of controlling the HIV epidemic and drug-resistant TB grew, DOTS was extended around the world.[25] DOTS is a framework which includes providing treatment to people with sputum smear-positive disease who report to healthcare centres with TB symptoms.[25] There is standardised regimen of drugs and observation of treatment is maintained for at least two months.[25] It is likely that, through the success of the programmes, the two-stage model of TB infection and disease was cemented into global communication and policy. However, despite the success of DOTS, TB prevalence isn't falling at the rate required to meet the UN Sustainable Development Goals in 2030 or the END TB goals in 2035.[1,26]

2.3 Tests for *M.tb* infection and TB disease

With different stages of *M.tb* infection and TB disease, the immunological response of the body is different.[3] At each stage of disease, there are different tools that can be used to infer, detect and confirm the existence of *M.tb* within a person.[3,5] Detection of infection and disease have the primary objective of allowing treatment to be correctly targeted. Preventative therapy is aimed at people with an infection, to prevent progression to disease.[27] Longer course, multi-drug regimens are recommended for people with disease, to cure as well as to minimise further lung damage and reduce potential for onwards infection.[28]

2.3.1 Standard tests and investigations

A panel of experts have collated recommended and possible investigations in the process to diagnose TB disease.[29] Here, I will discuss each of microbiology, imaging, and symptoms and signs, along with tests for infection.

Microbiology

Mycobacterial culture on liquid media is widely considered the reference standard test for diagnosing and confirming TB disease.[3,30,31] However, at about four weeks to culture *Mtb* and six to eight weeks to declare a negative result, it is a slow process.[30,32] Sputum smear microscopy is not as sensitive, nor as specific, as culture, and so will likely only be positive with higher

bacterial loads than culture can detect, however, results can be available on the same day.[3] The lower sensitivity of sputum smear microscopy leads to the distinction often seen between smear-positive and smear-negative disease, with the assumption that a higher load correlates with higher infectiousness and more severe disease.[33]

There are also molecular tests that utilise nucleic acid amplification for detection of *M.tb* DNA.[3] Xpert MTB/RIF (Xpert) is the most widely used of these and, with the correct infrastructure, can provide a same day result with a similar performance to that of liquid culture.[3] Alongside testing for disease, Xpert also tests for rifampicin resistance with a 95% sensitivity and 98% specificity.[3]

Imaging

The most common form of imaging for TB is chest radiography, first included in diagnostic guidelines in 1922.[15] Interpretation of radiographic films is subjective, leading to intra- and inter-reader variation.[34] Although experience reduces discrepancies between readers, when asked "is this film normal?" there was a 34% disagreement in responses from a panel of 90 experts, and similarly when asked "is there a cavity present" there was a 28% disagreement.[34] Performance of radiography is also affected by other factors outside of interpretation, like the quality of the film , the process of developing the film, the functionality of the machine, and the skill of the radiographer.[34] However, improvements in technology mean that radiography machines, once significant pieces of infrastructure, can be carried in a backpack, and produce a digital image that can be analysed by an algorithm to give a score between 0 and 99 on the likelihood of TB.[35] These technologies can achieve 98% specificity compared to expert readers, and 76% specificity when compared to Xpert as the reference, both when sensitivity is fixed at 90%.[35–37] Although more uncommon, and needing more significant infrastructure, other imaging technologies can be used for detection of TB in the lungs, including combined positron emission tomography/computed tomography (PET-CT), standalone CT, and magnetic resonance imaging (MRI).[3,31,38]

Mtb Infection

There is no test in use that can confirm viable *M.tb* infection. Instead, the two tests used for *M.tb* infection actually detect an immune response to *M.tb,* which is considered a proxy for infection.[39] The two tests for infection are interferon-gamma release assays (IGRA) and TST. A positive TST or IGRA test does not indicate future progression to disease, and it is believed that people who have fully recovered from an infection, or even disease, may still return a positive TST or IGRA with no risk of further progression.[40] Similarly, a positive TST could also be as a

response to receiving the BCG vaccine or from the booster effect of having two TSTs close together.[3,40] Meta-analyses to establish the performance of these tests for detecting infection find TST sensitivity ranges from 65% to 77% and IGRA sensitivity ranges from 61% to 89% with newer versions at the upper end of this range.[41] IGRA can have specificity between 86% and 99%, which is higher than can be expected for TST in a BCG vaccinated population.[41] When considering each test as a predictor for progression to disease, the positive predictive value is low (2.7% for IGRA and 1.5% for TST), meaning that a high number of people who screen positive will not progress, but the negative predictive value is high (99.7% for IGRA and 99.4% for TST), meaning that a low number of people who screen negative will progress.[42] Therefore, if used to target preventative therapy, many people would be treated who would never have progressed to disease.

Symptoms

Symptoms are used as another screening tool for TB, particularly in prevalence surveys.[6,7,43] The guidelines for systematic screening highlight five symptoms that are indicative of TB; cough of any duration, haemoptysis, weight loss, fever, or night sweats.[43] The same symptoms without haemoptysis form the four symptom screening guidelines for people living with HIV.[44] These are not an exhaustive list of symptoms caused by TB, but some prevalence surveys use these as their baseline for a symptom screen.[7] Some prevalence surveys reduce the list, most commonly to a cough or at least two weeks, but some extend the list to any possible TB symptom.[7] Symptom screening cannot definitively diagnose disease and so will be followed by some form of bacteriological testing and potentially radiology or other imaging.

2.3.2 Novel tests

Diagnostic tests

The WHO has laid out Target Product Profiles (TPPs) for novel screening and diagnostic tests that move away from the classic tools of radiography, bacteriology, and symptoms.[45] For a novel triage test for TB disease they proposed an optimal target of 95% sensitivity and 80% specificity, with minimal requirements of 90% and 70% respectively.[45] For the diagnostic test, the target is an optimal sensitivity of 98% in people with smear-positive disease, and 68% in people with smear-negative (culture-positive) disease.[45] These standards are comparable to those of the Xpert assay.[45,46]

C-reactive protein (CRP), an inflammatory marker, is not specific to TB, however it has shown a higher sensitivity for “active tuberculosis” than many other inflammatory markers.[47,48] This is primarily being considered as a screening test for people living with HIV, and in a healthcare

setting, CRP can achieve a sensitivity of 89% and specificity of 72% in people living with HIV - very close to the minimal requirements for a triage test.[47,49] In a meta-analysis of nine studies, two studies evaluated the performance in the context of screening, with sensitivity between 81% and 85% and specificity between 58% and 81%.[50]

Predictive tests

The WHO also released TPPs for tests that can predict progression from infection to disease.[51] These would ideally identify the 5 - 15% of those infected with *M.tb* who were likely to progress to disease, as opposed to the currently available tests of TST and IGRA that only show whether a person has been exposed to *M.tb*.[51] This could help focus preventative therapy interventions to only those at risk of progression.[52] The optimal sensitivity and specificity for this test was set at 90%, with the minimal performance set at 75% each.[51] These performances were expected to predict progression within two years of the initial test.[51]

A study in South African adolescents found a blood-biomarker, a 16-gene signature of risk with a reported sensitivity of 66% and specificity of 81%.[53] Further validation was performed on a cohort from another study finding a sensitivity of 54% and specificity of 83%.[53] This analysis only accounted for disease that developed within 12 months of the blood test, which could mean that the sensitivity at two years would be higher.[53] However, with the results published, this would not meet the TPP for a test for progression from infection to disease.[51]

Another recently completed trial in South Africa, CORTIS, by the same team, refined the blood-biomarker to an 11-gene signature.[53,54] With two different score thresholds for the biomarker, at 60% and 26%, sensitivity and specificity for both progressive infection and prevalent disease were reported. Prediction of progressive infection performed more poorly using the 11-gene signature rather than the 16-gene signature with the 60% threshold reporting a sensitivity of 25% and specificity of 91%, and the 26% threshold reporting a 48% sensitivity and 75% specificity.[54] At face value, the sensitivity and specificity for prevalent TB disease also do not perform sufficiently well for the triage or diagnostic test TPPs, with sensitivities of 35% and 72%, and specificities of 91% and 75% for the 60% and 26% thresholds respectively.[54] However, when these are stratified by symptom presentation, the performance of the 26% threshold surpasses the minimum performance for a triage test for people with symptomatic, bacteriologically-positive disease.[45,54]

As a predictor for progression from infection, blood-biomarkers still need significant refinement.[51,53,54] Beyond the need to improve sensitivity and specificity, these biomarkers are still based around the two-stage infection and disease paradigm and a change towards

understanding risk of progression to stages on the spectrum may well improve their prognostic power.[53,54] However, there is potential for use of blood-biomarkers as a triage test for symptomatic, bacteriologically-positive disease.[45,54] The downside of blood-biomarkers is that they require a collection of venous blood, and specialist lab equipment, creating barriers to testing.[54]

It is not only the blood-biomarkers that are designed around the two-stage model of infection and disease. These tests are used in a way that assumes a single performance across people with active disease, which is considered a single state. However, if there is an understanding that not all people with active disease report symptoms, for example, then it is clear to see that a symptom screen will not perform equally in those with and without symptoms. Accepting the different performances of screening tests across the spectrum means that the impact of screening a population is dependent on both the distribution of disease within the population, as well as the performance of the test being used. This is a research gap I will investigate in Chapter 5.

2.4 Modelling

Mathematics has been used to model and understand infectious diseases since at least the 18th century.[55] Daniel Bernoulli, a mathematician, physicist, and physician, tried to estimate the increase in life expectancy through variolation against smallpox by considering competing risks, a methodology now better known in actuarial than epidemiological literature.[55,56] The foundation of compartmental models as we know them today was laid by four physicians in the early 20th century; Ross, Hamer, McKendrick, and Kermack.[55] The first formalisation of a TB model was in 1962 by Hans Waaler *et al.*.[57,58] Taking inspiration from other fields, such as meteorology, demography, and economics, he argued that epidemiology would be a more promising field to use modelling in, as compared to demography or economics, because the unpredictability of humans does not matter, just the underlying behaviour of the disease.[57] Whilst this statement does not hold true in reality, good approximations can be made without taking human choice into consideration, as is also the case for demography and economics.

2.4.1 Early TB models

As is the case for all models, the model Waaler *et al.* used was not intended to be wholly realistic, instead providing enough representation to improve the use of available data.[57] Acknowledging this simplicity, Waaler *et al.* split the population into three states affected by TB, the "Non-infected", the "Infected, non-cases", and the "Infected, cases".[57] These were set out as a series of first-order difference equations with assumptions that newborns are always free from

infection, death is possible from any state, and once infected, always infected, but healing is possible from "Infected, cases" back to "Infected, non-cases".[57] Parameterisation of this model was based on data from studies in South India, led by Frimodt-Moller.[59] The prevalence of TB infection in India at the time was estimated to be 52%, but due to "non-specific sensitivity" (presumably higher than expected proportions of the population with a positive TST result), this was assumed to be an overestimate, with 25% used as a more reasonable approximation - matching recent estimates.[2,57] Healing, from "cases" to "non-cases" was estimated to occur at a rate of 10% per year.[57] This model was used to compare interventions and show that a continuous BCG programme that effectively reduces the infection rate by 50% has a better impact on prevalence long term than either passive case finding or a single active case finding intervention with sufficient treatment for 67% of the "cases".[57] This study was, however, only intended as a proof of concept that the modelling mechanisms in other areas could successfully be utilised to understand epidemiological trends.[57]

By 1968 Waaler had extended his model structure further, as can be seen in Figure 2.2.[60] Firstly waning BCG protection was added for those non-infected but vaccinated.[60] Then the compartment for infection was split by time at 5 years from infection.[60] From infection it was possible to progress to active disease that was classified as either infectious or non-infectious, with progression to infectious from non-infectious possible.[60] The differentiation between the two active disease compartments had two purposes: to calculate the force of infection based only on those with infectious disease, and to apply different treatments to the two groups.[60] Recovery from disease was similarly split based on the previous disease episode type (infectious or non-infectious), to allow for different relapse rates.[60] The whole model was then split into five-year age groups.[60] This model was intended to be closer to reality than the original model, although introduced significant complexity that it could not be solved analytically.[60] Although presented as a theoretical structure with no implementations, it was later used with data in both Europe and India to understand the impact of different control measures.[61,62]

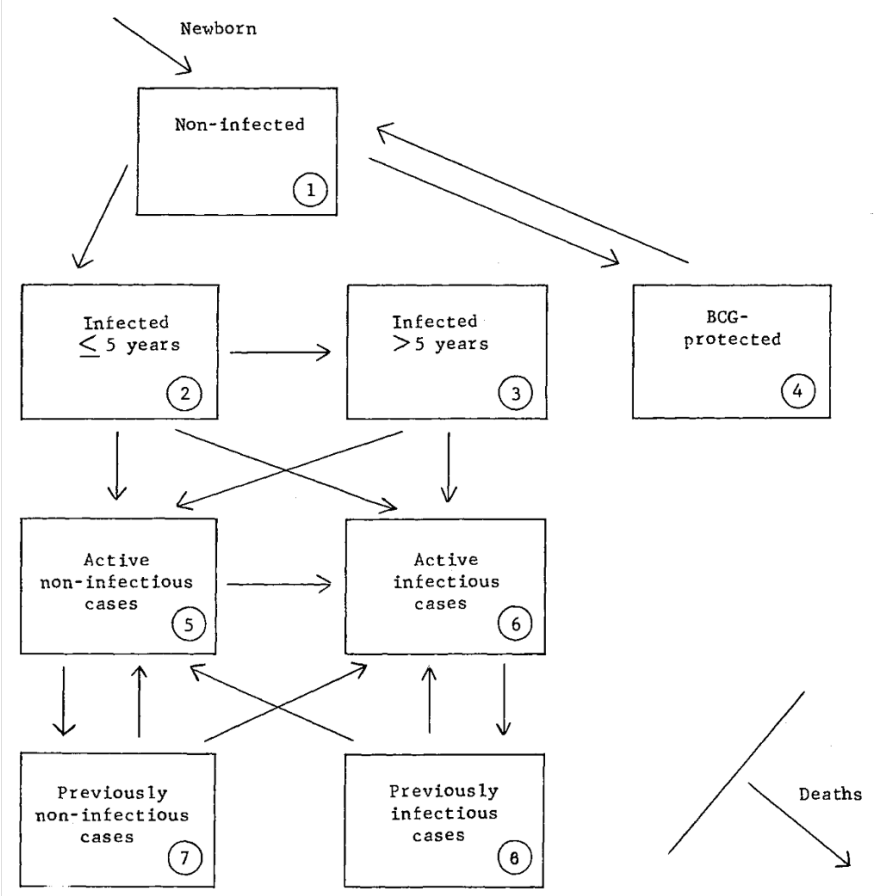


FIGURE 2.2: THE UPDATED MODEL STRUCTURE DESCRIBED BY WAALER IN 1968[60]

A second study, by Revelle and Male, also looked to determine optimal case detection and treatment pathways.[63] The model structure and population heterogeneity was simplified, by splitting the population into five groups, based on the outcome of four tests; TST, chest radiography, sputum culture, and sputum smear.[63] The assumption was that a positive reaction to a higher test (smear>culture>radiography>TST) implied positivity in all tests below.[63] Active disease was defined as all those with a positive culture, and the aim was to maximise the number of people with active disease who received treatment.[63] Rather than discussing the progression between disease, and how the effects of different screening tests progressed over time, this was purely an economic analysis on prevalent disease, maximising the number of people with active disease detected and treated, whilst minimising costs incurred per each active case treated.[63]

An alternative structure, proposed by Chorba and Sanders in 1971 considered eight states; susceptibles, new infection, dormant low-risk, dormant high-risk, active infection, natural recovery, therapy, recovery with therapy, as seen in Figure 2.3.[64] Susceptibles were those who have never had a significant infection, however, due to the limitations on testing, this was defined as all those with a negative TST.[64] Active infection was defined as all those with

infectious disease, therapy was for those who were undergoing treatment, recovery with therapy was for those who are non-infectious after treatment, and natural recovery was those who have recovered to a non-infectious state from active infection without treatment.[64] Re-progression to active infection was possible from the states of natural recovery and recovery with therapy.[64] The three states of new infection and dormancy were interconnected with all three containing people who had a positive TST.[64] Dormant high-risk were those who had had a positive TST for some time, and also had a positive chest x-ray, but had never reached the active infectious state.[64] Dormant low-risk was similar, but with nothing visible on chest x-rays, and new infections were those who had not progressed to either dormancy state or active disease.[64] This model was simulated for the years 1953 to 1989, and validated against data up until 1969.[64] In comparison to Waaler's estimate, natural recovery from active infection is doubled to 20% annually.[57,64]

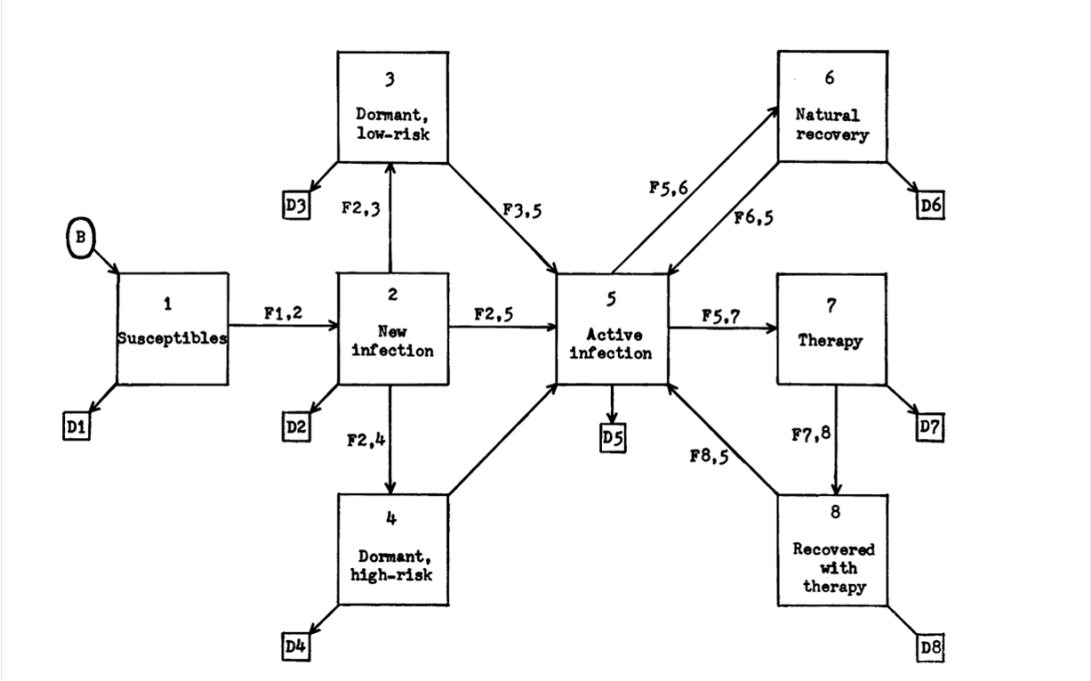


FIGURE 2.3: THE MODEL STRUCTURE DESCRIBED BY CHORBA AND SANDERS IN 1971[64]

Horwitz, in 1973, presented another model structure, with slight differences to those seen before, as seen in Figure 2.4.[65] Infection risk, as with all other models, was determined by the number of people with active disease, but recovery from infection through both natural processes or through chemoprophylaxis was included.[65] Infection was followed by "active cases", defined as people who are sick or receiving treatment, followed either by recovery (through treatment or otherwise) or death.[65] Each state was assigned a number based on the relevant prevalence in Denmark at the beginning of 1969, with flows assigned a number based on the changes during the subsequent year.[65] Based on these numbers, rates were then approximated for each flow.[65]

As can be seen from these initial models, definitions and structures all have an underlying similarity, but the specifics of the models have key differences. For example, some define disease as bacteriological positivity, but Horwitz, using the data available, describes active cases as those who are sick with no mention of bacteriology.[57,61–65] However, Horwitz's structure does not differentiate between those undergoing treatment and those still with disease, unlike Chorba and Sanders.[64,65] Each of these models is parameterised by data from a single location and source, and these data are taken with no acknowledgement of the uncertainty behind them. With the exception of Waaler's 1968 split between non-infectious and infectious, one concept none of the transmission models incorporate is the heterogeneity in "active disease".[57,60,64,65] Revelle and Male, although not a transmission model, provided the most differentiation between disease states, for an economic analysis of screening prevalent disease, but there was no need to consider progression to or between disease states.[63] Even the differentiation within active cases in Waaler's model does not allow for regression from infectious to non-infectious disease without recovery to non-active disease, which restricts the ability to understand the undulations in disease that are known to happen.[20,60]

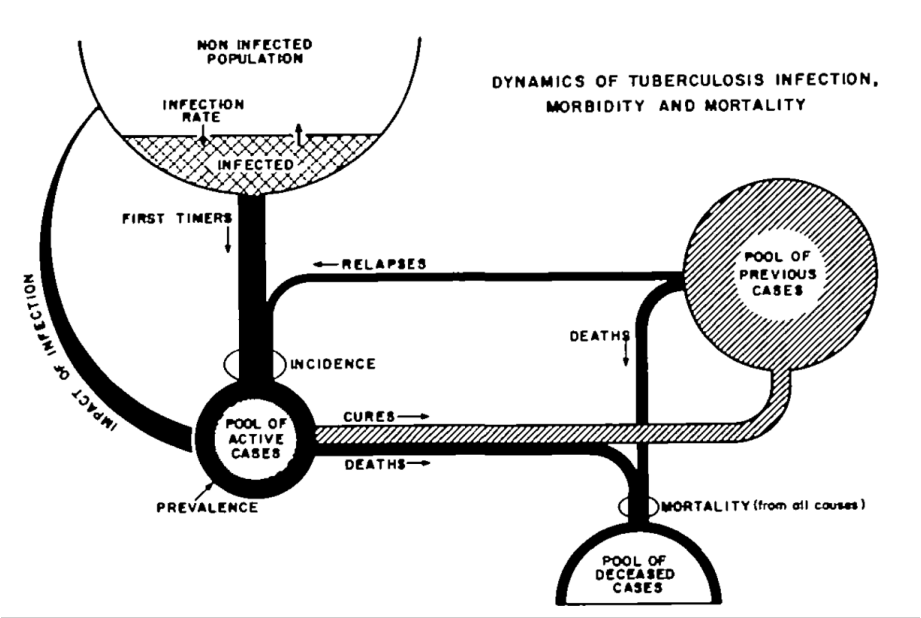


FIGURE 2.4: THE MODEL STRUCTURE DESCRIBED BY HORWITZ IN 1973[65]

These first 15 years of TB modelling yielded a small number of models, but almost nothing was published between the early 1970s and the mid-1990s.[58] However, from 1995 and onwards, there were over 300 studies published in 20 years, and more have been published since.[66] A number of studies have summarised and categorised TB models that have been published during this later period.[58,66,67] As it is not feasible to describe all other models on TB natural

history in detail, I will use these summaries to aid the discussion of further progressions in TB natural history modelling.[58,66,67]

2.4.2 Progression of models over time

2.4.2.1 Model structures

Simplistically, TB model structures usually follow an SEIR (susceptible, exposed, infectious, recovered) framework.[58,66,67] However, as even the earliest models started to show, restricting the progression through infection and disease to two compartments (exposed and infectious) is unusual.

Menzies *et al.* classified 12 model structures for progression from infection to disease.[66] Over half of the generic structures had at least two compartments for "latent" infection.[66] Two more of the structures create a similar effect by using time dependent parameter progressions, although these were only implemented in individual based models where each individual could be tracked by the time spent in each compartment.[66] Of the final three model structures, one does not include an infection compartment, one progresses from susceptible to infection, and then to disease, and the final structure has two progression routes from susceptible, one direct to disease and the other to infection.[66] This final model structure was, in fact, the most widely used, with almost half of the studies using some variation around it.[66]

For comparison against real world data, Menzies *et al.* used data on progression from infection to disease from the MRC vaccine trials by Sutherland to test the accuracy of each of these structures.[66,68] The model structure used in almost half the studies found within the review was outperformed by eight of the other structures, all of which included at least two possible stages of infection, with different progression rates.[66] This is largely due to the fact that the data show a sharp decrease in the number of people progressing from infection to disease over time, which cannot be replicated by a single parameter.[66,68]

Menzies *et al.* only classify the model structure differences in the space between initial infection and progression to disease.[66] There is neither discussion on recovery from infection or disease, or on differentiations within the structure of disease.[66] Most models do not incorporate recovery from infection, instead assuming lifelong infection, although some do include recovery, either naturally, or through treatment (preventative therapy).[2,69–72] Natural recovery from disease is more commonly included. This can either be modelled as a return to latent infection, or to a separate "recovered" compartment.[69–71] For the model structure chosen by Menzies *et al.* with both natural recovery from slow latent and from active TB, means that it is possible to

completely recover from infection and disease with a simulated partial immunity to re-infection.[71] This is not a possibility in most other model structures.

Many model structures choose to split disease between smear-negative and smear-positive active disease, on the basis that smear-positive is more infectious than smear-negative.[75] Sometimes the distinction within the model is labelled instead as low- and high-transmissibility disease, with the underlying assumption of a split by smear status.[76] All these structures are very similar to the model proposed by Waaler, and the subsequent work by Blower *et al.*, where disease was split into infectious and non-infectious strata.[60,77]

Population level strata for risk factors enable models to incorporate factors that impact the progression of disease. In TB models, HIV is the most common stratification, with almost 30% of the models collated in the review by Menzies *et al.* including this stratification.[66] The 11 country-specific analyses collated by Houben *et al.*, nine included some form of stratification for HIV; from two strata to 11 strata.[67] Age and drug resistance were the next most common stratifications, although less common over all models than HIV.[66,67]

2.4.2.2 Data sources

Of the transmission models included in the review by Menzies *et al.*, the highest number of studies, almost 25%, gave no citations for the source of their parameters of progression from infection to active disease.[66] Otherwise, for the remaining 236 studies, the most cited sources for parameters were 14 studies; four modelling studies, five empirical studies, four review studies, and one expert report.[66] This is not an exhaustive list of data sources, but highlights the similarities between the parameterisation choices.

The four review studies in the list of most cited sources from Menzies *et al.* were cited 55 times by the studies in the review.[66] Ferebee's review of controlled chemoprophylaxis trials was the most used source.[78] There were 13 studies included in this review, with six as part of a series of US Public Health Service trials, and the other seven from around the world (Japan, Kenya, Philippines, the Netherlands, USA, Greenland, Tunisia) and results were presented on prevention and eradication of infection along with tuberculosis morbidity and adverse reactions to isoniazid.[78] The definition of "disease" in these studies is variable, sometimes based on bacteriology, sometimes radiology, sometimes symptoms, which makes the threshold for what is defined as "disease" difficult to draw and compare.[78] Infection, determined by TSTs, is not always defined from time of conversion, more often from the time a positive reaction was recorded (whether new or old).[78] It is understood that the rate of progression to disease changes over time and so it is difficult to determine any true rate from infection to disease, as

the time from infection to positive TST is unknown. The review by Comstock, the next most cited review, used similar data to Ferebee (e.g. the US Public Health Service trials, data from Greenland) but also introduced data from other sources (e.g. South India and Canada).[79] Again, summaries on risk of infection and risk of progression to disease were reported, with the same limitations as those from the Ferebee review.[78,79] Neither provided a definitive value for these risks, however, with many different countries and settings, an appropriate estimate from a single study could be extracted.[78,79] The final two reviews are by Styblo and these are even more extensive than those of Ferebee and Comstock.[78–81] Although extensive, none of these reviews is systematic which could introduce bias from the selection process. Also, the definitions in the reviews are limited, with the progression to disease defined from a post-infection TST to a mixture of radiology, bacteriology, and symptoms, meaning that these are progressions are to disease all across the spectrum and have a truncated duration.

Empirical studies, where the data are collected first hand, make five of the top fifteen most cited sources as compiled by Menzies *et al.*, and account for 74 citations.[66] Horsburgh *et al.* was the most cited empirical study, in determining the rate of reactivation TB, among people with a latent infection.[82] The second most cited empirical study reported on a TB outbreak within a facility for people living with HIV.[23] It describes the progression to disease for 11 people who are presumed to have contracted TB from one other person at the facility. Rather than the standard assumption of disease behaviour taking months or even years to develop, in this scenario, one person developed disease within four weeks of initial exposure.[23] These data are invaluable when stratifying a model by HIV status, as the reviews by Comstock, Ferebee, and Styblo are mostly too old to contain studies on TB/HIV co-infection.[78–80] The next two most commonly cited empirical studies are those by Sutherland, one analysing infection and disease in populations in the Netherlands, from 1880 to 1970, and the other following participants from the BCG vaccine trial in the UK.[66,68,83] The final empirical study listed, by Di Perri *et al.* is another small study of a TB outbreak among people living with HIV.[84] The varying definitions for disease seen in the reviews is also seen in these studies.

A modelling study by Vynnycky and Fine is the most popular source of parameters, when a source is cited.[66,85] This is a modelling study to estimate age-specific risks of developing disease, split into primary, endogenous, and exogenous.[85] As such, the model and parameters are stratified by age.[85] This model uses data from the MRC vaccine trial (Sutherland *et al.*), case fatality data from India as used in the review by Comstock, and notification data from Norway.[68,79,85,86] Blower *et al.*, in the second most cited modelling study, cite both Styblo and Comstock as the source of progression parameters.[66,77,79,81] Mortality and recovery rates are

sourced from a review paper by Gzyrbowski and Enarson, compiling data from England, Canada, the U.S.A., India, Singapore, Korea, Taiwan, and Kenya, some of which have been used in the other reviews.[86,87] Dye *et al.* have parameterised their model using data from each of the WHO regions, and validated the model against data from the Netherlands.[88] They have also referred back to both the Vynnycky and Fine study and the Blower *et al.* study in the construction and analysis of the model.[77,85,88] The final widely cited modelling study for parameters is an analysis on the control of drug-resistant tuberculosis by Dye and Williams.[66,89] Data to inform the reproduction number for drug-susceptible disease was sourced, again, from work by Styblo.[80,89]

This is not an exhaustive list of modelling papers or sources for parameterisation. However, as can be seen from this summary, there is substantial cross-referencing, and ultimately, most references lead back to a subset of the four review studies, or the studies reviewed within them. As previously mentioned, these data have not been systematically reviewed, lack a well defined description of what constitutes disease, assume a two-stage model with no spectrum of disease, and have truncated the duration of infection without a record of when infection or TST conversion occurred.[78,79] I intend to approach this research gap in my work in Chapters 3, 4, and 6.

A parameter that was not discussed in the review by Menzies *et al.* is that of TB mortality.[66] A systematic review by Tiemersma *et al.* collated data on disease mortality in smear-positive and smear-negative disease, followed later by an analysis to convert the reported values into estimates of mortality rate and recovery rate suitable for use in a model.[33,90] For comparison, Ragonnet *et al.* identified 22 modelling studies that split TB disease by smear status, and presented specific TB mortality and recovery rates.[33] Point estimates for the annual rate of TB mortality from smear-positive disease varied from almost 0 to 0.6, and the corresponding smear-negative value ranged from 0 to 0.4.[33] In comparison, the rates Ragonnet *et al.* attributed to TB mortality from smear-positive and smear-negative disease were almost 0.4 and almost 0 respectively.[33] The studies by Tiemersma *et al.* and Ragonnet *et al.* have drawbacks; they conflate the historical term of "open disease" with smear-positive disease, and have assumed a fixed smear status within disease.[33,90] The concept of smear-positive and smear-negative disease only formalised as a concept in the mid-1960s, and descriptions of open disease often imply a differentiation based on symptom presentation, sometimes with no mention of bacteriological status so the conflation of smear-positive with open TB may be flawed.[18,19,91] However, there is still value in their data and analysis, and the comparison of mortality rates

from different studies shows that there is no consensus on the understanding of disease mortality.[33,90]

2.4.2.3 Latest developments

Despite the evolving understanding of disease and the WHO and prevalence surveys moving away from smear status, much of the present modelling work still includes a differentiation on the basis of smear alone with no mention of presentation of symptoms.[70,76,92] Most of these models are re-formulations of previously published structures, with occasional complexities added. For example, Arregui *et al.* have reworked the model structure from one used by Dye *et al.*, with susceptible, latent, disease and treatment stages, split into pulmonary smear-positive, pulmonary smear-negative, and non-pulmonary disease.[70,88] Zwick *et al.* have used the same baseline model, however combined smear-negative and extra-pulmonary disease into one non-infectious disease compartment.[69,70]

Understanding of the need to model across the spectrum of disease has been growing, and whilst working on this thesis, a few studies and preprints have been published that incorporate some form of spectrum of disease. Ku *et al.* have used data from prevalence surveys to determine the duration of symptomatic and asymptomatic TB disease.[93]This has been followed by an analysis by Emery *et al.* on the relative infectiousness of subclinical disease as compared to clinical disease.[94] In a full transmission model, Clark *et al.* have incorporated subclinical disease into a model used to analyse different vaccine delivery strategies.[95] Finally, Ryckman *et al.* have extended the work by Tiemersma *et al.* and Ragonnet *et al.* to create a model that encompasses both smear and symptom statuses.[33,90,96] I will discuss these works in more detail throughout the thesis, as they pertain to the content in each chapter. An in-depth discussion of each work, and how they relate to my findings will be included in the Discussion chapter (Chapter7).

2.4.3 Modelling methodologies

2.4.3.1 Representation of chance

There is a choice between a deterministic system or a stochastic system and whether or not to include chance within a mathematical model.[97] A stochastic system contains an element of chance within event selection, implemented by using random number generators to determine whether, or when, an event will occur given the prior probability.[97] Stochasticity is important to determine dynamics of a system at low prevalences (e.g. emergence or elimination of a disease).[97,98] Stochasticity is implemented by choosing events based on the probability of the event happening and the outcome of a random number generator, although this can be

implemented in a number of different ways within systems.[97,98] With the same input parameters, the output of a stochastic system will vary each time. Of interest, is the overall picture of the system after multiple evaluations with the same, or similar, starting parameters.[97] The output after multiple evaluations means that variation due to chance can be quantified, however the computation required for each evaluation, and the number of evaluations required means that these systems can be computationally expensive.[97]

A deterministic system, on the other hand, has no variation due to chance; for a given parameter set, one evaluation of the system will return the same as 100 evaluations.[97] Using parameters that represent the behaviour within a population will result in an output that represents the average outcome.[97] For example, for a disease with a modelled reproduction number of just less than one, the expected behaviour is that the disease would not lead to an outbreak. With stochasticity included in the model, there is a small chance that an outbreak can occur, but in a deterministic model, this will never happen. Computationally, deterministic systems are easier and quicker to use.[97] However the lack of chance within the framework can lead to predictions that are not, and cannot be, realistic, particularly when small numbers (e.g. of cases) are involved.[97]

2.4.3.2 Representation of a population

Compartmental models split a population into characteristic groups; a simplistic disease model will split a population into three groups by whether they are susceptible to an infection, infected, or recovered.[99] There are then parameters that describe the rate into and out of each compartment.[99] The reciprocal of the parameter describes the mean duration within the relevant compartment.[99] Within each compartment, the population is considered one homogeneous group, there is no tracking of how long each individual stays in a compartment (beyond the mean duration), and at most points in time, the number of people within a compartment is not an integer.[97] To represent more heterogeneity within populations, strata can be added to models.[99] For example, when a disease has different rates of progression in different age groups, the model structures can incorporate forwards disease progression alongside sideways ageing (remaining in the same state but ageing up to a new age group).[75,85,99] With stratifications, it is possible to gain a well-rounded understanding of disease dynamics within a population. Compartmental models can be either deterministic or stochastic.[98]

Compartmental models can be used to explore population level trends, and describe a "typical" disease process.[97] However, there are some instances where a different approach is preferred or even required.[97,100] For example, populations where important heterogeneities in disease pathways are an outcome of interest, or a specific community is being simulated, where

between-person heterogeneities are important to consider.[97,100,101] Individual based models (IBMs), sometimes referred to as Agent based models (ABMs), differ from compartmental models by instead tracking individuals over time.[100–102] Each individual holds the information on their current state and attributes (e.g. disease, age, risk factors, location, etc.), along with their individual history, and progression in any direction at each time point is individualised.[97] IBMs, are stochastic, and are frequently more complex than compartmental models.[97,102] They have potential to become complex beyond understanding and useful analysis and are often computationally intensive, however, when used properly, they can be invaluable sources of analysis.[97,102] .

2.5 Research gaps

This chapter has expanded on the brief background in Chapter 1 to introduce the research gaps that this thesis intends to fill, along with the chapters where I discuss the work to fill them. The research gaps identified are:

- There was a wealth of published knowledge on the progression of M.tb infection and TB disease that has been forgotten (Chapter 3)
- Data to parameterise models originate from four review studies, which each overlap in the data collected, do not differentiate between smear status or symptoms, and poorly define the time of infection (Chapter 3, Chapter 4, Chapter 6)
- Models are still structured around smear status despite the move to classify disease by symptom status (Chapter 4, Chapter 6)
- Screening tests assume a homogeneous disease state to define performance and so predictions of population interventions do not take into account the varying performance of tests across the spectrum of disease (Chapter 5)

I will address these research gaps through the aims and objectives introduced in Chapter 1.

2.6 References

1. World Health Organization. Global tuberculosis report 2021. (World Health Organization, 2021).

2. Houben, R. M. G. J. & Dodd, P. J. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS medicine* **13**, e1002152 (2016).

3. Pai, M. *et al.* Tuberculosis. *Nat Rev Dis Primers* **2**, 16076 (2016).

4. Drain, P. K. *et al.* Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. *Clinical Microbiology Reviews* **31**, 24 (2018).

5. Lin, P. L. & Flynn, J. L. The End of the Binary Era: Revisiting the Spectrum of Tuberculosis. *The Journal of Immunology* **201**, 2541–2548 (2018).

6. Onozaki, I. *et al.* National tuberculosis prevalence surveys in asia, 1990-2012: An overview of results and lessons learned. *Trop Med Int Health* **20**, 1128–1145 (2015).

7. Frascella, B. *et al.* Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. *Clinical Infectious Diseases* **73**, e830–e841 (2021).

8. Moyo, S. *et al.* Prevalence of bacteriologically confirmed pulmonary tuberculosis in south africa, 201719: A multistage, cluster-based, cross-sectional survey. *The Lancet Infectious Diseases* **22**, 1172–1180 (2022).

9. Glaziou, P., Sismanidis, C., Zignol, M. & Floyd, K. *Methods used by WHO to estimate the global burden of TB disease*. https://www.who.int/tb/publications/global_report/gtbr2018_online_technical_appendix_global_disease_burden_estimation.pdf?ua=1 (2016).

10. Glaziou, P., Dodd, P. J., Dean, A. & Floyd, K. Methods used by WHO to estimate the global burden of TB disease. (2020).

11. Tuberculosis. *Institute for Health Metrics and Evaluation* (2021).

12. Ledesma, J. R. *et al.* Global, regional, and national sex differences in the global burden of tuberculosis by HIV status, 19902019: Results from the global burden of disease study 2019. *The Lancet Infectious Diseases* **22**, 222–241 (2022).

13. Barberis, I., Bragazzi, N. L., Galluzzo, L. & Martini, M. The history of tuberculosis: From the first historical records to the isolation of koch's bacillus. *Journal of Preventive Medicine and Hygiene* **58**, E9–E12 (2017).

14. Long, E. R. & Hopkins, F. D. History of diagnostic standards and classification of tuberculosis of the national tuberculosis association. *Am Rev Tuberc* **65**, 494–504 (1952).

15. Skavlem, J. H. Arrested Pulmonary Tuberculosis. *Diseases of the Chest* **4**, 8–11 (1938).

16. National Tuberculosis Association. Diagnostic standards and classification of tuberculosis. (New York, N.Y., 1940).

17. Opie, E. L. & McPhedran, F. E. The organization of an out-patient clinic for epidemiological investigation. *Am J Hyg* **22**, 539 (1935).

18. World Health Organization. *WHO expert committee on tuberculosis - eighth report*. https://apps.who.int/iris/bitstream/handle/10665/40606/WHO_TRS_290.pdf?sequence=1&isAllowed=y (1964).

19. World Health Organization. *WHO expert committee on tuberculosis - ninth report*. https://apps.who.int/iris/bitstream/handle/10665/41095/WHO_TRS_552_eng.pdf?sequence=1&isAllowed=y (1974).

20. Gothi, G. D. Natural History of Tuberculosis. *Indian Journal of Tuberculosis* **25**, (1977).

21. Corbett, E. L. *et al.* The growing burden of tuberculosis: Global trends and interactions with the HIV epidemic. *Archives of Internal Medicine* **163**, 1009–1021 (2003).

22. Cantwell, M. F. & Binkin, N. J. Tuberculosis in sub-Saharan Africa: a regional assessment of the impact of the human immunodeficiency virus and National Tuberculosis Control Program quality. *Tubercle and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease* **77**, 220–225 (1996).

23. Daley, C. L. *et al.* An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. *New England Journal of Medicine* **326**, 231–235 (1992).

24. Getahun, H., Gunneberg, C., Granich, R. & Nunn, P. HIV infection-associated tuberculosis: the epidemiology and the response. *Clin Infect Dis* **50 Suppl 3**, S201–7 (2010).

25. World Health Organization Communicable Diseases. *What is DOTS? : A guide to understanding the WHO-recommended TB control strategy known as DOTS*. WHO/CDS/CPC/TB/99.270 (1999).

26. World Health Organization. The end TB strategy. (2015).

27. Organizat, W. H. WHO consolidated guidelines on tuberculosis. Module 1: Prevention. Tuberculosis preventive treatment. *Tuberculosis, Lung Diseases, HIV Infection* 86–92 (2021) doi:10.30978/TB2021-2-86.

28. World Health Organization. WHO consolidated guidelines on tuberculosis: module 4: treatment: tuberculosis care and support. (World Health Organization, 2022).

29. Akkerman, O. W. *et al.* Clinical standards for drug-susceptible pulmonary TB. *The International Journal of Tuberculosis and Lung Disease* **26**, 592–604 (2022).

30. Goletti, D., Lee, M.-R., Wang, J.-Y., Walter, N. & Ottenhoff, T. H. M. Update on tuberculosis biomarkers: From correlates of risk, to correlates of active disease and of cure from disease. *Respirology* **23**, 455–466 (2018).

31. Ankrah, A. O. *et al.* PET/CT imaging of mycobacterium tuberculosis infection. *Clinical and Translational Imaging* **4**, 131–144 (2016).

32. Davies, P. D. O. & Pai, M. The diagnosis and misdiagnosis of tuberculosis [state of the art series. Tuberculosis. Edited by i. D. Rusen. Number 1 in the series]. *The International Journal of Tuberculosis and Lung Disease* **12**, 1226–1234 (2008).

33. Ragonnet, R. *et al.* Revisiting the Natural History of Pulmonary Tuberculosis: A Bayesian Estimation of Natural Recovery and Mortality Rates. *Clinical Infectious Diseases* **73**, e88–e96 (2021).

34. Cleeff, M. van, Kivihya-Ndugga, L., Meme, H., Odhiambo, J. & Klatser, P. The role and performance of chest x-ray for the diagnosis of tuberculosis: A cost-effectiveness analysis in nairobi, kenya. *BMC Infectious Diseases* **5**, 111 (2005).

35. Murphy, K. *et al.* Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. *Scientific Reports* **10**, 5492 (2020).

36. Digital x-ray systems.

37. FIND Radiology LandscapeProject Team. *Digital chest radiography and computer-aided detection (CAD) solutions for tuberculosis diagnostics: Technology landscape analysis*. (2021).

38. Bomanji, J. B., Gupta, N., Gulati, P. & Das, C. J. Imaging in tuberculosis. *Cold Spring Harbor Perspectives in Medicine* **5**, a017814 (2015).

39. Pai, M. *et al.* Gamma Interferon Release Assays for Detection of Mycobacterium tuberculosis Infection. **27**, 3–20 (2014).

40. Behr, M. A., Edelstein, P. H. & Ramakrishnan, L. Is Mycobacterium tuberculosis infection life long? *BMJ* **367**, l5770 (2019).

41. Thillai, M., Pollock, K., Pareek, M. & Lalvani, A. Interferon-gamma release assays for tuberculosis: Current and future applications. *Expert Review of Respiratory Medicine* **8**, 67–78 (2014).

42. Diel, R., Loddenkemper, R. & Nienhaus, A. Predictive value of interferon-γ release assays and tuberculin skin testing for progression from latent TB infection to disease state. *Chest* **142**, 63–75 (2012).

43. WHO. *Systematic screening for active tuberculosis: Principles and recommendations*. (2013).

44. World Health Organization. *Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource constrained settings*. (World Health Organization, 2011).

45. *High-priority target product profiles for new tuberculosis diagnostics: Report of a consensus meeting*. (2014).

46. World Health Organization. WHO consolidated guidelines on tuberculosis Module 2: Screening Systematic screening for tuberculosis disease. (World Health Organization, 2021).

47. Yoon, C. *et al.* Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *The Lancet Infectious Diseases* **17**, 1285–1292 (2017).

48. Yoon, C., Davis, J. L. & Cattamanchi, A. C-reactive protein and tuberculosis screening: a new trick for an old dog? *Int J Tuberc Lung Dis* **17**, 1656 (2013).

49. World Health Organization. *Rapid communication on systematic screening for tuberculosis*. https://extranet.who.int/iris/restricted/bitstream/handle/10665/337372/9789240016552-eng.pdf (2020).

50. Yoon, C. *et al.* Diagnostic accuracy of C-reactive protein for active pulmonary tuberculosis: a meta-analysis. *The International Journal of Tuberculosis and Lung Disease* **21**, 1013–1019 (2017).

51. World Health Organization. *Development of a target product profile (TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease*. https://apps.who.int/iris/bitstream/handle/10665/259176/WHO-HTM-TB-2017.18-eng.pdf;jsessionid=2C75724D4E5B1EEA18B98FBF55D5661B?sequence=1 (2017).

52. Fiore-Gartland, A. *et al.* Considerations for biomarker-targeted intervention strategies for tuberculosis disease prevention. *Tuberculosis* **109**, 61–68 (2018).

53. Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: A prospective cohort study. *The Lancet* (2016) doi:10.1016/S0140-6736(15)01316-1.

54. Scriba, T. J. *et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): A randomised controlled trial. *The Lancet Infectious Diseases* **0**, (2021).

55. Brauer, F., Castillo-Chavez, C. & Feng, Z. Introduction: A prelude to mathematical epidemiology. *Mathematical Models in Epidemiology* **69**, 3–19 (2019).

56. Dietz, K. & Heesterbeek, J. a. P. Daniel Bernoulli's epidemiological model revisited. *Mathematical Biosciences* **180**, 1–21 (2002).

57. Waaler, H., Geser, A. & Andersen, S. The Use of Mathematical Models in the Study of the Epidemiology of Tuberculosis. *American Journal of Public Health and the Nations Health* **52**, 1002–1013 (1962).

58. Colijn, C., Cohen, T. & Murray, M. International Symposium on Mathematical and Computational Biology. in 123–148 (WORLD SCIENTIFIC, 2007). doi:10.1142/9789812708779_0008.

59. Frimodt-M, J. A Community-Wide Tuberculosis Study in a South Indian Rural Population, 1950-1955. *Bulletin of the World Health Organization* **22**, 61–170 (1960).

60. Waaler, H. T. A dynamic model for the epidemiology of tuberculosis. *American Review of Respiratory Disease* **98**, 591–600 (1968).

61. Waaler, H. T. & Piot, M. A. The use of an epidemiological model for estimating the effectiveness of tuberculosis control measures. *Bulletin of the World Health Organization* **41**, 75–93 (1969).

62. Waaler, H. T., Gothi, G. D., Baily, G. V. J. & Nair, S. S. Tuberculosis in rural south india. A study of possible trends and the potential impact of antituberculosis programmes. *Bulletin of the World Health Organization* **51**, 263–271 (1974).

63. Revelle, C. & Male, J. A mathematical model for determining case finding and treatment activities in tuberculosis control programs. *The American Review of Respiratory Disease* **102**, 403–411 (1970).

64. Chorba, R. W. & Sanders, J. L. Planning models for tuberculosis control programs. *Health Services Research* **6**, 144–164 (1971).

65. Horwitz, O. Disease, cure, and death: epidemiologic and clinical parameters for chronic diseases illustrated by a model–tuberculosis. *American Journal of Epidemiology* **97**, 148–159 (1973).

66. 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* **18**, e228–e238 (2018).

67. Houben, R. M. G. J. *et al.* Feasibility of achieving the 2025 WHO global tuberculosis targets in south africa, china, and india: A combined analysis of 11 mathematical models. *The Lancet Global Health* **4**, e806–e815 (2016).

68. Sutherland, I. *The ten-year incidence of clinical tuberculosis following 'conversion' in 2,550 individuals aged 14 to 19 at time of conversion*. (1968).

69. Zwick, E. D., Pepperell, C. S. & Alagoz, O. Representing tuberculosis transmission with complex contagion: An agent-based simulation modeling approach. *Medical Decision Making* 0272989X2110078 (2021) doi:10.1177/0272989x211007842.

70. 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* **115**, E3238–E3245 (2018).

71. Menzies, N. A. *et al.* Prospects for Tuberculosis Elimination in the United States: Results of a Transmission Dynamic Model. *American Journal of Epidemiology* **187**, 2011–2020 (2018).

72. Goodell, A. J. *et al.* Outlook for tuberculosis elimination in California: An individual-based stochastic model. *PLOS ONE* **14**, e0214532 (2019).

73. Behr, M. A. *et al.* Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet (London, England)* **353**, 444–449 (1999).

74. Tostmann, A. *et al.* Tuberculosis Transmission by Patients with Smear-Negative Pulmonary Tuberculosis in a Large Cohort in The Netherlands. *Clinical Infectious Diseases* **47**, 1135–1142 (2008).

75. Houben, R. M. G. J. *et al.* TIME impact a new user-friendly tuberculosis (TB) model to inform TB policy decisions. *BMC Medicine* **14**, (2016).

76. Tuite, A. R., Gallant, V., Randell, E., Bourgeois, A.-C. & Greer, A. L. Stochastic agent-based modeling of tuberculosis in canadian indigenous communities. *BMC Public Health* **17**, 73 (2017).

77. Blower, S. M. *et al.* The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine* **1**, 815–821 (1995).

78. Ferebee, S. H. Controlled chemoprophylaxis trials in tuberculosis. A general review. *Bibl Tuberc* **26**, 28–106 (1970).

79. Comstock, G. W. Epidemiology of tuberculosis. *The American Review of Respiratory Disease* **125**, 8–15 (1982).

80. Styblo, K. Selected papers (vol 24). *KNCV Tuberculosis Foundation - Selected papers* (1991).

81. Styblo, K. in *Advances in Respiratory Medicine* (eds. Flenley, D. C. & Petty, T. L.) 77–108 (Churchill Livingstone, Edinburgh, 1986).

82. Horsburgh, C. R. *et al.* Revisiting rates of reactivation tuberculosis: a population-based approach. *American Journal of Respiratory and Critical Care Medicine* **182**, 420–425 (2010).

83. Sutherland, I., Svandova, E. & Radhakrishna, S. The development of clinical tuberculosis following infection with tubercle bacilli. *Tubercle* **63**, 255–68 (1982).

84. DiPerri, G. *et al.* NOSOCOMIAL EPIDEMIC OF ACTIVE TUBERCULOSIS AMONG HIV-INFECTED PATIENTS. *The Lancet* **334**, 1502–1504 (1989).

85. Vynnycky, E. & Fine, P. E. M. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiology and Infection* **119**, 183–201 (1997).

86. National Tuberculosis Institute, B. Tuberculosis in a rural population of South India: a five-year epidemiological study. *Bulletin of the World Health Organization* **51**, 473–88 (1974).

87. Grzybowski, S. & Enarson, D. A. The fate of cases of pulmonary tuberculosis under various treatment programmes. *Bull. I.U.A.T.* **53**, 70–75 (1978).

88. Dye, C., Garnett, G. P., Sleeman, K. & Williams, B. G. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *The Lancet* **352**, 1886–1891 (1998).

89. Dye, C. & Williams, B. G. Criteria for the control of drug-resistant tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 8180–8185 (2000).

90. Tiemersma, E. W., van der Werf, M. J., Borgdorff, M. W., Williams, B. G. & Nagelkerke, N. J. D. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PloS One* **6**, e17601 (2011).

91. Szucs, E. Sputum Examination in So-called Closed Tuberculosis. *JAMA: The Journal of the American Medical Association* **86**, 946 (1926).

92. Tovar, M., Aleta, A., Sanz, J. & Moreno, Y. Modeling the impact of COVID-19 on future tuberculosis burden. *Communications Medicine* **2**, 1–10 (2022).

93. Ku, C.-C. *et al.* Durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease with a Bayesian analysis of prevalence survey and notification data. *BMC Medicine* **19**, 298 (2021).

94. Emery, J. C. *et al.* Estimating the contribution of subclinical tuberculosis disease to transmission an individual patient data analysis from prevalence surveys. doi:10.1101/2022.06.09.22276188.

95. Clark, R. A. *et al.* The impact of alternative delivery strategies for novel tuberculosis vaccines in low- and middle-income countries: a modelling study. doi:10.1101/2022.04.16.22273762.

96. Ryckman, T. S., Dowdy, D. W. & Kendall, E. A. Infectious and Clinical Tuberculosis Trajectories: Bayesian modeling with case finding implications. doi:10.1101/2022.06.27.22276965.

97. Garnett, G. P., Cousens, S., Hallett, T. B., Steketee, R. & Walker, N. Mathematical models in the evaluation of health programmes. *The Lancet* **378**, 515–525 (2011).

98. Allen, L. J. S. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling* **2**, 128–142 (2017).

99. Blackwood, J. C. & Childs, L. M. An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics* **5**, 195–221 (2018).

100. Prats, C. *et al.* Individual-based modeling of tuberculosis in a user-friendly interface: Understanding the epidemiological role of population heterogeneity in a city. *Frontiers in Microbiology* **6**, (2016).

101. Nepomuceno, E. G., Takahashi, R. H. C. & Aguirre, L. A. INDIVIDUAL-BASED MODEL (IBM): AN ALTERNATIVE FRAMEWORK FOR EPIDEMIOLOGICAL COMPARTMENT MODELS. 30 (2016).

102. Grimm, V. *et al.* The ODD protocol: A review and first update. *Ecological Modelling* **221**, 2760–2768 (2010).

Chapter 3 Systematic Review of Historical Literature

3.1 Introduction

As highlighted in Chapter 2, our growing clinical understanding of a spectrum of TB disease is not often reflected in models beyond the difference between smear positive and smear negative bacteriologically positive disease.[1–5] Similarly, the parameters that describe any transitions are not based on direct observations but are instead calibrated to more widely available estimates such as incidence and prevalence.[6,7] With effective treatments having existed for more than 70 years, there is very little recent data that can be used to inform these transitions.[8,9]

Before effective treatments were discovered in the 1950s, scientists and physicians were still trying to fully understand the nature of TB disease and how best to treat it.[8] Robert Koch's discovery of *M.tb* was presented in 1882, followed by the discovery by Röentgen of x-rays in 1895 which together still form a basis of disease detection.[1,4,10] Radiography only became widely encouraged in the diagnostic criteria for TB disease in the 1920's, however both of these discoveries were vital to improving the understanding of TB disease.[11]

Whilst accessing data over 70 years old presents challenges, two studies have previously searched through older literature to understand segments of the natural history of TB.[12–14] The first review aimed to understand the natural history of intra-thoracic tuberculosis in children, and identified seven studies that followed cohorts of children for up to 25 years between 1928 and 1961.[12] The second review aimed to understand the duration of, and mortality from pulmonary tuberculosis in the general population.[13] They identified 22 studies following different populations between 1905 and 1970.[13] This second review was followed by a modelling analysis to convert the data presented into parameters for the TB specific mortality from smear-positive and smear-negative disease.[14] Smear-positive and smear-negative were not phrases used in literature of that age, so they have conflated open and closed TB with smear-positive and smear-negative respectively.[13,14] It is more likely, instead, that open TB was classified based on symptoms rather than bacteriology.[15] Whilst these two reviews and the subsequent modelling analysis have improved understanding of intra-thoracic disease in children and mortality from pulmonary TB in adults, there is still not any data collected on how pulmonary disease progresses and changes between infection and death or recovery.

Given the success of the previous reviews to identify usable historical data on aspects of the natural course of TB disease, data that can inform progression of pulmonary TB disease across the spectrum for use in modelling may exist in the 60 year gap between discoveries of *M.tb* and effective treatments. Here I present the work undertaken to search the literature that was

written both during and after the pre-treatment gap highlighted above in preparation for use in modelling.

3.1.1 Aims of the Review

This systematic review was registered with PROSPERO (CRD42019152585) and aimed to find and report data on the progression and regression of TB disease, measured by changes in chest radiography (cxr), bacteriology, and symptoms. This data could then be synthesised within a modelling framework to parameterise a structure that discretised the spectrum of disease. The systematic review itself is discussed in this chapter, and the proceeding modelling is discussed in Chapter 4.

3.1.2 My contribution

This systematic review of historical literature was originally conceived by Hanif Esmail (HE) and Rein Houben (RH). I (AR) helped develop the aims and the protocol with Bianca Sossen (BS), Hanif Esmail, and Rein Houben. I have done the largest share of the work at all stages of the review, screening the titles, finding and reviewing the full texts, assessing bias, extracting data, and coordinating the team of reviewers. This work has been shared with the original group and through a team of nine other people whom I guided and managed; three helped source the full texts and search the Index Medicus (Bridget Chivers, Adrienne Burrough, Rohisha Luchun), six assisted with the full text screening, bias screening, and data extraction (Torben Heinsohn (TH), Beatrice Frascella (BF), Aurea Oradini (AO), Federica Balzarini (FB), Brit Häcker (BH), and Katarina Kranzer (KK)). For every study that has had data extracted, I either extracted myself or approved the extraction by someone else. I prepared the tables and figures, and wrote the chapter. A manuscript for publication has also been written by Bianca Sossen, and is undergoing peer review.[16] I will be joint first author on this publication, and it can be found as a preprint on medRxiv (https://www.medrxiv.org/content/10.1101/2022.08.30.22279374v1), and in Appendix A.

3.2 Methods

3.2.1 Search Strategy

This review was registered with PROSPERO (CRD42019152585) (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019152585). We searched three online databases, Medline (via PubMed), Web of Science, and EMBASE, from their earliest records in the search period (1895, 1900, and 1947 respectively) to 1960. The search terms, in both English and German, included various phrases for TB disease ("Pulmonary tuberculosis" OR "Incipient tuberculosis" OR Phthisis OR Minimal tuberculosis OR Moderate tuberculosis OR Advanced tuberculosis) alongside phrases that indicated observation, diagnosis, or progression ("Follow-up stud*" OR "Follow up stud*" OR Course OR "Biological

Evolution" OR "After-history" OR Supervision OR Epidemiological Stud* OR Epidemiology OR Prognosis OR Prospective OR Longitudinal OR "Roentgenographic survey" OR Radiography OR Observation* OR Progress* OR Diagnosis). To supplement the online database search, a broad title screen of the Index Medicus from 1900 - 1945 was conducted, focusing similarly on TB disease that was measured and observed over time. Each volume of the Index Medicus was only scanned by one person before relevant titles were included for screening, and so inclusion over exclusion was the goal. To supplement each of these searches, collaborators each scanned personal libraries they had access to, and added any relevant articles to the review, regardless of time period.

All titles found through these methods were imported into Covidence, which also provided an initial de-duplication of titles picked up from multiple sources.[17] Where duplications were missed by Covidence, these were later manually excluded when discovered, and the number manually excluded was added to the initial de-duplication count.

3.2.2 Title and full text screening

Due to the age of the articles, very few had abstracts, so the initial screen was based on the title alone. Any title that was clearly not a description of untreated pulmonary tuberculosis in adults was excluded. If this was not clear, the article was included to the next stage. Every article was screened independently by two people (AR, BS, TH, KK, RH) with a third person independently resolving conflicts (HE). The titles in English were reviewed by AR and BS, the titles in German were reviewed by TH and KK, the titles in other languages were excluded and this process was completed by AR and RH.

Full texts were sourced through online journal archives. When this was not possible, archive.org and hathitrust.org were searched for scanned copies of old journals. If these sources did not yield the full text, the paper archives at the British Library and the Wellcome Library in London were both searched. If the text could not be sourced through any of these methods, the article was excluded from the review. Any articles in German went through the same process with the addition of searching the Heidelberg library and the library of the German Central Committee against tuberculosis (Deutches Zentralkomitee zur Bekämpfung der Tuberkulose, DZK).

Full text screening aimed to exclude articles that were unable to provide data for the purposes of the review. Therefore, the process focussed on eliminating articles where:

- there was no microbiological testing
- all participants underwent surgical or drug interventions
- some participants (>5%) underwent surgical or drug interventions and the results were not separated

- all participants were under 10 years old
- some participants (>5%) were under 10 years old and the results were not separated
- fewer than 25 participants
- there was no longitudinal follow-up
- it was only a review article
- the results were reported in a language other than English or German

These criteria were chosen at the start of the review process to help select the studies that would best fit the extraction criteria whilst being unambiguous for the review team. The intention was to remove as many studies as possible that would not be relevant or extractable, whilst keeping all those that were extractable. This meant hard limits for example no microbiological testing or fewer than 25 participants. Respectively these were intended to remove studies where there was insufficient testing to confirm TB, although in reality the ideal study had more than just a mention of microbiological testing, and remove small studies with a high risk of bias.

Each full text was independently screened by two reviewers, and any conflicts were resolved between the reviewers. If a conflict could not be resolved between the two, it was taken to a third person for a final say (AR, HE, BS), provided that person was not one of the initial two reviewers. For English articles, the group of reviewers expanded to six (AR, BS, TH, FB, BF, AO) due to the size of the review and time constraints of some of the reviewers. The group of German reviewers also expanded to three (TH, KK, BH).

Many of the articles in the literature search were part of multi-article studies of one population or cohort. Where possible, these multi-article studies were treated as one entity and as such, when the term study is used, this refers to all publications discussing one cohort or trial, whether a single article or multiple articles. Once an article passed full text screening, any others within the same series were brought through and grouped together, wherever possible.

3.2.3 Bias screening

The bias screening took two forms, referred to as general and specific. The general screening was based on an adjusted Newcastle-Ottawa scale, whilst the specific screening tested the reliability of the testing mechanisms used for each of radiography, microbiology, and symptom screening. Both screenings were assessed independently by two reviewers (two of AR, BS, TH, FB, BF, AO for English studies or TH and BH for German studies), with conflicts resolved between the two, or taken to a third person for a final say (AR, HE, BS).

The general bias screen was split into three sections: study selection, comparability, and outcome. Study selection covers four questions; the representativeness of the cohort, the selection of the non-exposed cohort, the ascertainment of exposure, and the demonstration that the outcome of interest was not present at the start of the study. Representativeness of the cohort was relevant for all studies, with the highest score awarded to studies where the cohort was identified through mass screening or similar. As there was no comparator group, the selection of non-exposed cohorts was not important. The final two sections were only applicable to the subset of studies that observed cohorts where individuals had recently been exposed to, or infected by, *M.tb*. All other studies included required proof of progression from infection in individuals, either through radiography or microbiology at baseline so did not require any proof of exposure. There was no comparator cohort, so comparability within the study was not a question included in the bias. Then the outcome was assessed in four points; how the outcome was assessed, whether any intervention could have changed the outcome, whether the follow-up was long enough to see the outcome, and whether the follow-up was complete. All studies had to be assessed against all points. Between the study selection and outcome selections, a maximum of two points could be lost by a single study for it to remain in the review. As stated in the PROSPERO registration, the studies were required to report at least 12 months follow-up of the cohorts to ensure sufficient elapsed time within the study to see changes in disease state. The criteria for the general bias screening can be seen in Figure 3.1.

The specific bias screen was also split into three sections: radiography, microbiology, and symptoms. Radiography was scored on the technology used, the diagnostic criteria used, the number of independent readers, and how many people received radiography at follow-up. Microbiology was scored on whether they mention how *M.tb* was detected, how the samples were collected, and how many people were tested at follow-up. Symptoms were scored on whether they discussed symptoms and how many people they were reported for at follow up. Each section was considered a pass if they lost no more than one star. Two of the three sections had to pass for the study to be included. If a study failed to explicitly discuss the methods of testing in a way that more than one section failed, but referred to a standardised diagnostic criteria with descriptions that included the missing sections to high enough quality (e.g. the National Tuberculosis Association's (NTA) diagnostic criteria), then the study could be included on the basis of those standards.[18] The criteria for the specific bias screening process can be seen in Figure 3.2.

	GENERAL QUALITY ASSESSMENT	MAXIMUM STARS
	STUDY SELECTION	
Answer for all studies	1. Representativeness of 'exposed' cohort	
	Truly representative of general population e.g. identified through prospective recruitment following mass screening or household screening	1 star
	Somewhat representative of general population e.g. identified through occupational screening	
	Selected group e.g. sanatorium or hospital patients	0 stars
	No description of the derivation of the cohort	
	2. Selection of non-exposed cohort	
	NOT APPLICABLE	
Only answer for PICO 1 studies	3. Ascertainment of exposure	
	Documented TST conversion	1 star
	Exposure appropriately inferred from significant exposure to active TB e.g. in household or occupational setting	
	No clear significant exposure to active TB	0 stars
	No description	
	4. Demonstration that outcome of interest not present at start of study	
	Yes - evidence of TB excluded by CXR at beginning of follow-up, or documented as not being present on CXR within previous 12 months	1 star
	Not stated	0 stars
	COMPARABILITY	
	NOT APPLICABLE	
	OUTCOME	
Answer for all studies	5. Assessment of outcome	
	Disease defined on the basis of microbiological demonstration of organism	1 star
	Disease defined without microbiological evidence	0 stars
	No description	
	6. Could intervention prevent development of outcome	
	No medical or surgical treatment provided	1 star
	Medical or surgical treatment provided to a subgroup of clearly identified individuals analysed separately and no selection bias	
	Medical or surgical treatment provided to a subgroup (<10%) and that treatment deemed unlikely to influence outcome	
	Medical or surgical treatment provided to a large number of individuals OR unable to be analysed separately OR selection bias	0 stars
	7. Was total follow-up long enough for outcomes to occur	
	Yes - follow-up equal to or longer than 12 months	1 star
	No	0 stars
	8. Adequacy of follow-up of cohorts	
	Complete follow-up	1 star
	Participants lost to follow up unlikely to introduce bias: <20% of participants lost to follow-up per annum, or description provided for those lost, making introduction of bias unlikely	
	Lost to follow-up rate > 20% per annum and no description of those lost	0 stars
	No statement	

FIGURE 3.1: THE GENERAL BIAS SCREENING CRITERIA

	ADDITIONAL SPECIFIC QUALITY ASSESSMENT *- If study meets general quality criteria above, used to determine if microbiological, radiographic and symptoms are recorded to sufficient quality*	MAXIMUM STARS
Answer for all studies	**RADIOGRAPHY**	
	1. Technology	
	Radiography (X-Ray, Roentgenography)	1 star
	Mass Miniature radiography (Photofluorography, Aburography)	
	Other (please specify)	0 stars
	2. Diagnostic criteria used	
	Clear use of recognised diagnostic criteria (e.g. ATS)	1 star
	Clear description of study specific methodology (not recognised diagnostic criteria)	
	No or limited description of criteria or methodology	0 stars
	No imaging	
	3. Methodology	
	Double or triple independent read	1 star
	Single reader	0 stars
	Not specified	
	4. Follow-up	
	Complete follow-up	1 star
	<20% of participants per annum in follow-up did not receive imaging OR description for those not receiving imaging, making introduction of bias unlikely	
	<80% followed up received imaging and no description of those lost	0 stars
	No statement	
	MICROBIOLOGY	
	1. Technology	
	Culture - solid	1 star
	Culture - liquid	
	Smear - with concentration	
	Smear - without concentration	
	Guinea pig innoculation	
	Other (please specify)	
	Not specified	0 stars
	2. Sampling method	
	24-hour sputum collection	1 star
	48-hour sputum collection	
	72-hour sputum collection	
	Induced sputum (e.g. physiotherapy or saline)	
	Spot sputum collection (including <24 hour collection	
	Gastric aspirate	
	Nasal/Larangeal/Tracheal swab	
	Other (please specify)	
	Not specified	0 stars
	Number of samples sent per participant:	
	3. Follow-up	
	Complete follow-up of those able to produce sputum	1 star
	<20% of per annum of participants had sputum collected (of those able to produce) for microbiological investigation OR description provided for that lost, making introduction of bias unlikely	
	Follow-up rate <80% and no description of those lost	0 stars
	No statement	
	SYMPTOMS	
	1. Methodology	
	Standardised symptom screen described in methodology	1 star
	Clear statement of symptomatic vs asymptomatic	
	Standardised definitions from published criteria used with symptoms included in definition	
	No clear statement	0 stars
	2. Follow-up	
	Complete follow-up	1 star
	<20% of participants per annum in follow-up did not get symptom assessment OR description provided for those lost, making introduction of bias unlikely	
	Follow-up rate <80% and no description of those lost	0 stars
	No statement	

FIGURE 3.2: THE SPECIFIC BIAS SCREENING CRITERIA FOR TESTS

3.2.4 Data extraction

For studies in English, data was extracted by one reviewer (AR, BS, TH, FB, BF, AO), with a second reviewer checking the extractions (AR, BS). For the German studies, data was extracted by one reviewer and checked by the other (TH, BH) and was then quality checked against the English extractions (AR). Any differences in interpretation of the extracted data were discussed between the reviewers initially and taken to a wider group (AR, BS, HE, RH) if no consensus could be reached.

Data extraction was split into three sections: a situational description of the study, the tests, methods, and diagnostic criteria used, and the progression of sub-cohorts. For the situational description of the study, data were collected, where possible, on the local burden of infection and disease, the source of the cohort and how the cohort was selected, and demographics of the cohort. The tests used were recorded, split by radiography, bacteriology, and symptoms. For each the method of measurement and frequency of measurement was recorded along with any extra information on specific diagnostic criteria used.

Finally, the main focus of the data extraction was to understand the progression through disease. A cohort where all individuals were classified as having the same disease state was identified, through records of each of radiography, bacteriology, and symptoms. The disease state was described by each of these three tests groups, based on the definitions extracted in the previous section (e.g. cxr pos, micro neg, sympt neg for a cohort where everyone was deemed, by definitions, to have a positive radiography result, but negative bacteriology and no symptoms). This group was considered the number at risk of progression at the start of follow-up. Regardless of the total length of follow-up, the baseline group size stayed the same as, through the bias screening process, loss to follow-up was low. Over time, the disease state of some individuals in the cohort changed, and grouping together individuals with the same change, provided the transition group. The number of people who made a similar transition, the time it took for all these transitions to be recorded, and the states they transitioned to (e.g. cxr pos, micro pos, sympt pos) were all recorded. The final point recorded was how the data were collected, recorded as either cumulative or cross-sectional follow-up. In cumulative follow-up studies, individuals were closely followed up with cumulative recording of whether they had transitioned to a new state.[19] After transitioning, individuals were excluded from follow-up. At each time point, the number of people recorded to have transitioned are over the entire study period until that point, rather than from the previous time point. In cross-sectional follow-up studies, individuals were followed up at the single reported timepoint; only their final state was recorded, without knowledge of any additional transitions that occurred before the end of the study.[19] For stable transitions, where there is no further movement, this distinction is not

necessary, however if the transition is not stable and movement between states continues, over time the cumulative count would continue to increase whereas the cross-sectional count would reduce.

Sometimes the test states were not explicitly mentioned, and instead phrases from the NTA diagnostic criteria were used.[18] These definitions split disease into three states; arrested, quiescent, and active.[18] We translated these into test results based on the description within the criteria.[18]

Arrested = cxr pos, micro neg, sympt neg
"Constitutional symptoms absent. Sputum, if any, must be concentrated and found microscopically negative for tubercle bacilli. Lesions stationary and apparently healed according to X-ray examination; no evidence of pulmonary cavity."

Quiescent = cxr pos, micro pos, sympt neg
"No constitutional symptoms. Sputum, if any, may contain tubercle bacilli. Lesions stationary or retrogressive according to X-ray examination; cavity may be present."

Active = cxr pos, micro pos, sympt pos
"Symptoms unchanged, worse or less severe, but not completely abated. Lesions not completely healed or progressive according to X-ray examination. Sputum almost always contains tubercle bacilli."

During the data extraction or checking process one of the reviewers (AR, BS), scanned the reference list and, using the title and context with which it was cited, snowballed further references for inclusion within the review. These articles were then considered at the full text review for further inclusion in the review.

Any studies that reached this point but could not provide extractable data, on the agreement of both reviewers, were excluded for not having sufficient data to extract.

3.2.5 Modelling preparation

In preparation for the modelling analysis in Chapter 4, data that could be used was labelled with the disease transitions represented.[2,3,20–24] The model structure, built from descriptions in multiple studies and evidence from the papers found within the review, can be seen in Figure 3.3A.[25] The model structure spilt TB disease into three discrete states; minimal, subclinical, and clinical.[25] These states are defined as:

- *Minimal disease (min/M)*: individuals with pathological changes in their lungs, due to *M.tb* with all bacteriologic tests negative (likely not infectious)[24,25]

- *Subclinical disease (sub/S)*: individuals with positive bacteriologic tests (likely infectious) but asymptomatic, also likely to have pathological changes in their lungs[4,25]
- *Clinical disease (clin/C)*: individuals with positive bacteriologic tests (likely infectious) and exhibiting symptoms of TB, also likely to have pathological changes in their lungs[25]

Two further definitions that will be used are:

- *TB disease*: any of the three disease states above
- *Infectious disease (inf/I)*: a combination of subclinical and clinical disease, under the assumption that positive bacteriological tests reflect the possibility of infectivity[26]

The model structure uses these definitions to help characterise disease transitions. Within TB disease, there are eight possible transitions, as shown in Figure 3.3.

- *Minimal to Subclinical (min-sub)* or *Subclinical to Minimal (sub-min)*: shown in Figure 3.3B, data for these transitions show a change between bacteriologically negative and positive (or vice versa) with no recorded symptoms when bacteriology positive
- *Subclinical to Clinical (sub-clin)* or *Clinical to Subclinical (clin-sub)*: shown in Figure 3.3C, data for these transitions show a change between symptom negative and positive (or vice versa) whilst always bacteriologically positive
- *Minimal to Clinical (min-clin)* or *Clinical to Minimal (clin-min)*: shown in Figure 3.3D, data for these transitions show a change between bacteriologically negative and positive (or vice versa) with symptoms when bacteriology positive. Although the data seemingly skips the subclinical stage, the underlying model structure assumes disease has progressed through subclinical without being recorded.
- *Minimal to Infectious (min-inf)* or *Infectious to Minimal (inf-min)*: shown in Figure 3.3E, data for these transitions show a change between bacteriologically negative and positive (or vice versa) with unknown symptoms when bacteriology positive. The underlying model structure assumes that disease progresses through the three states and that infectious disease is one of Subclinical or Clinical disease.

The defining disease states or groups were based on baseline and follow-up test results and the transitions were then defined as the baseline state to the follow-up state. Test results for minimal disease required radiologically positive and bacteriologically negative, regardless of symptoms. Test results for subclinical disease required bacteriologically positive, and symptom negative, regardless of radiologic state. Test results for clinical disease required bacteriologically positive, and symptom positive, regardless of radiologic state. Where bacteriology was positive but symptoms were not reported, this was assumed to include a mix of people with and without symptoms and was defined as infectious disease. These states are summarised in Table 3.1.

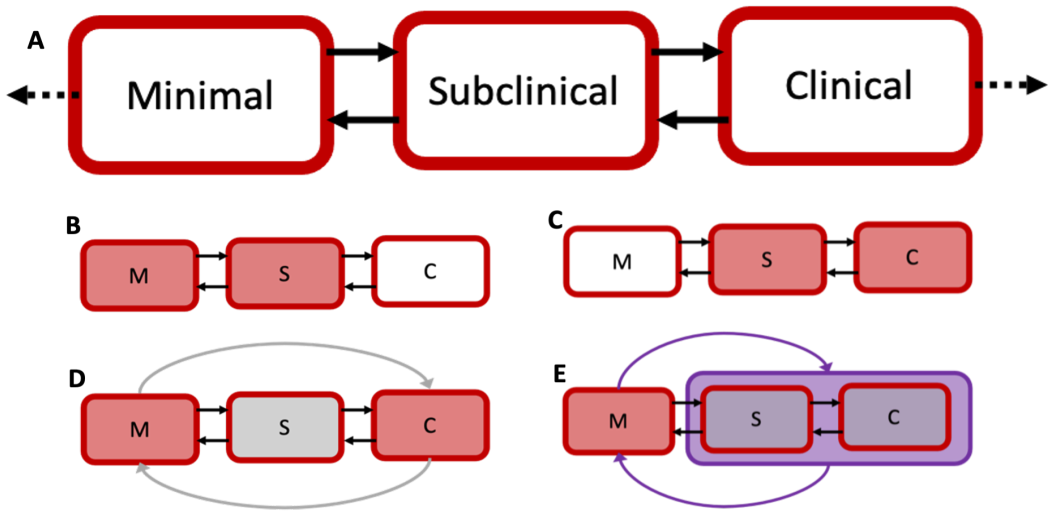


FIGURE 3.3: THE DISCRETISED MODEL STRUCTURE FOR DISEASE USED IN CHAPTER 4. A SHOWS THE THREE STATES AND THE PROPOSED TRANSITIONS BETWEEN EACH STATE, ALONG WITH RECOVERY FROM MINIMAL AND DEATH. B AND C HIGHLIGHT DATA THAT CAN DIRECTLY PARAMETERISE TRANSITIONS, WITH B SHOWING TRANSITIONS BETWEEN MINIMAL AND SUBCLINICAL, AND C SHOWING TRANSITIONS BETWEEN SUBCLINICAL AND CLINICAL. D SHOWS THE EFFECT OF DATA THAT MOVES BETWEEN MINIMAL AND SUBCLINICAL, THE UNDERLYING MODEL (BLACK LINES) STILL MOVES THROUGH SUBCLINICAL, EVEN THOUGH THE DATA (GREY LINES) SEEMINGLY SKIPS THE SUBCLINICAL STAGE. E, SIMILAR TO D, SHOWS TRANSITIONS BETWEEN MINIMAL AND INFECTIOUS (PURPLE BOX, SUBCLINICAL AND CLINICAL), WHERE THE UNDERLYING MODEL STRUCTURE STILL MOVES BETWEEN THREE STATES, BUT THE DATA EFFECTIVELY MOVES BETWEEN TWO.

TABLE 3.1: A SUMMARY OF THE TEST RESULTS THAT DEFINE THE DIFFERENT DISEASE STATES

Disease state	Minimal	Subclinical	Clinical	Infectious
X-ray	cxr pos	cxr pos	cxr pos	cxr pos
Microbiology	micro neg	micro pos	micro pos	micro pos
Symptoms	sympt neg/pos/unk/mix	sympt neg	sympt pos	sympt unk/mix

3.3 Results

3.3.1 Literature search

There were 12,769 titles imported from all sources (including snowballing) which combined down to 12,719 different studies. Initial de-duplication by Covidence found and excluded 2096 titles. A further 146 articles were identified as duplicates and removed by reviewers later in the process. This left 10,477 study titles to screen, of which 1,648 (15.7%) were deemed eligible for full text screening.

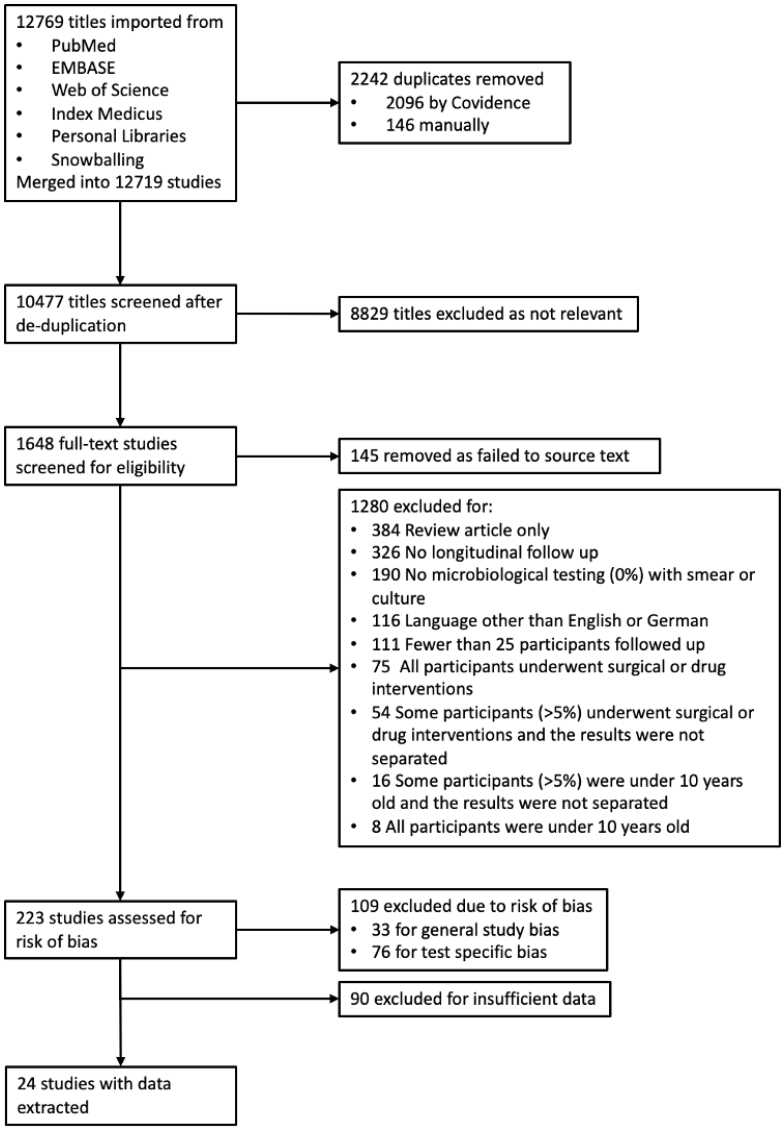


FIGURE 3.4: A PRISMA DIAGRAM DESCRIBING THE PROCESS OF THE SYSTEMATIC REVIEW

Of the 1,648 eligible for full-text screening, we were unable to source the text for 145 (8.8%). A further 1,280 (77.7%) were ineligible based on the exclusion criteria. Of the 223 that were eligible for data extraction, only 24 (10.7%) passed the bias screen (109 excluded) and had sufficient data (90 excluded) presented in a way that was conducive for extraction. Two of those excluded could have passed both the general and specific bias checks, but followed the cohort for less than one year so were not eligible to continue.[27,28] Assuming the 145 that had no full text were similarly distributed, there may have been two more studies with data to extract. The full detail of the process, in the form of a PRISMA diagram, can be seen in Figure 3.4. The studies and some of the associated data collected can be seen in Table 3.2.

3.3.2 Description of the studies included

The studies where data were extracted followed cohorts from the early 1920s until 2004. As noted previously, effective treatments became widely available in the 1950s and 1960s, and the study types and data available in each study reflect this transition. The study types can be broadly classified into two categories; observational studies where no treatment was available, and controlled trials where the efficacy of expanding treatment was being tested. There were 13 studies which were observational, 10 of which completed follow-up before 1960. The final three observational studies completed follow-up in 1967, 1968, and 2004 with all three offering standard of care in the local area for the infection or disease state suspected.[29–31] The remainder of the studies were controlled trials. The first was conducted in the 1940s to uncover the efficacy of Streptomycin as a treatment for bacteriologically positive disease.[32] The remaining studies were all controlled trials for bacteriologically negative disease and/or preventative therapies, mainly focussed on suspicions of TB disease through radiography results. The earliest of these studies concluded in 1959.

Follow-up types were more loosely defined by date or study type with 13 cumulative, 10 cross-sectional studies, and one study reporting both cross-sectional and cumulative data. More of the cross-sectional studies were observational (7/10 - 70%) and more of the cumulative studies were controlled trials (8/13 - 62%). Splitting instead by time of study, eight (80%) of the cross-sectional studies concluded before 1960, while five (38%) of the cumulative studies were concluded within the same time. Five of the cumulative studies were observational studies, with the other eight being controlled trials. The study with a mix of cumulative and cross sectional was an observational study following a prevalence survey.[31]

Of the 24 studies included, one study presented only a cohort from presumed infection, with no signs of disease through radiography, bacteriology, or symptoms.[29] One study presented a cohort with long term stable lesions, with limited expectations of progression.[33] Another six studies presented cohorts with presumed infection without disease, alongside cohorts with TB disease.[28,30,31,34–45] All cohorts without evidence of disease were not suitable for the modelling presented in Chapter 4 but are presented in Table 3.2 for completeness of the data.

TABLE 3.2: THE DATA FOR EACH STUDY INCLUDED IN THE SYSTEMATIC REVIEW AND, WHERE RELEVANT, THE MODEL TRANSITION THIS REPRESENTS GOING FORWARD. THE START AND END STATES ARE DEFINED BY TEST (CXR – RADIOGRAPHY, MICRO – BACTERIOLOGY, SYMPT – SYMPTOMS) AND RESULT (POS – POSITIVE, NEG – NEGATIVE, UNK – UNKNOWN). THE MODEL TRANSITIONS ARE REPRESENTED AS START AND END STATE (MIN – MINIMAL, SUB – SUBCLINICAL, CLIN – CLINICAL, INF – INFECTIOUS). THE "TRANSITIONS" ARE THE NUMBER OF PEOPLE WHOSE DISEASE PROGRESSED FROM THE START STATE TO THE END STATE, AND THE "COHORT SIZE' IS THE NUMBER AT RISK AT BASELINE.

Author	Continent	Study end	Study type	Data type	Start state	End state	Total transitions recorded	Cohort size at start	Months since enrolment	Model transition
Downes[45]	North America	1935	Observational - all known cases in population	cumulative	cxr pos micro pos sympt pos	cxr pos micro neg sympt neg	27	342	12.0	Clin-Min
							104	342	24.0	
							140	342	36.0	
							158	342	48.0	
							171	342	60.0	
Beeuwkes[33]	North America	1938	Observational - household contact study	cross sectional	cxr neg micro neg sympt neg	cxr unk micro pos sympt pos	1	784	33.0	
						cxr unk micro neg sympt pos	3	784	33.0	
					cxr pos micro neg sympt neg	cxr unk micro pos sympt pos	16	122	33.0	Min-Clin
					cxr pos micro pos sympt pos	cxr unk micro neg sympt unk	10	28	33.0	Clin-Min
Puffer[46]	North America	1943	Observational - household contact study	cross sectional	cxr pos micro neg sympt neg	cxr pos micro pos sympt pos	19	528	62.0	Min-Clin
					cxr pos micro pos	cxr pos micro neg	92	384	62.0	Clin-Min

					sympt pos	sympt neg				
Orrego Puelma[47]	South America	1945	Observational - patients	cross sectional	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	18	67	24.0	Min-Inf
Bobrowitz[48,49]	North America	1945	Observational - sanatoria patients	cross sectional	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	26	191	60.0	Min-Inf
Lincoln[50,51]	North America	1947	Observational - hospital TB patients admitted to start rest treatment	cumulative	cxr pos micro mix sympt unk	cxr pos micro neg sympt unk	45	134	24.0	Clin-Min
							71	134	36.0	
							80	134	48.0	
							83	134	60.0	
							86	134	72.0	
							87	134	84.0	
					cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	15	314	24.0	Min-Clin
							25	314	36.0	
							32	314	48.0	
							35	314	60.0	
							36	314	72.0	
Alling[50,52]	North America	1948	Observational - hospital TB patients admitted to start rest treatment	cumulative	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	8	58	60.0	Min-Clin
							10	58	156.0	
Marshall[31]	Europe	1948	Controlled trial - bacteriologically positive treatment	cumulative	cxr pos micro pos sympt pos	cxr pos micro neg sympt unk	2	52	6.0	Clin-Min

Borgen[34]	Europe	1949	Observational - occupational screening	cross sectional	cxr neg micro unk sympt unk	cxr pos micro pos sympt pos	4	6684	30.0	
					cxr pos micro neg sympt pos	cxr pos micro pos sympt pos	4	144	30.0	Min-Clin
Manser[53]	Europe	1951	Observational - hospital patients	cross sectional	cxr pos micro pos sympt unk	cxr pos micro neg sympt unk	15	40	6.0	Clin-Min
Breu[54]	Europe	1952	Observational - hospital patients	cross sectional	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	48	904	25.5	Min-Inf
Sikand[35]	Asia	1958	Observational - occupational screening	cumulative	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	38	319	12.0	Min-Inf
					cxr neg micro unk sympt unk	cxr pos micro pos sympt unk	89	11268	69.0	
						cxr pos micro neg sympt unk	251	11268	69.0	
Tuberculosis Society of Scotland[55]	Europe	1959	Controlled trial - bacteriologically negative treatment	cross sectional	cxr pos micro neg sympt neg	cxr pos micro pos sympt unk	9	95	24.0	Min-Sub
Groth-Petersen[36,37]	Europe	1959	Controlled trial -	cumulative	cxr neg micro unk sympt unk	cxr pos micro mix sympt unk	59	45953	24.0	
							108	45953	48.0	
							64	116639	24.0	

			preventative therapy				147	116639	48.0	
							73	103373	48.0	
							35	69607	48.0	
					cxr pos micro unk sympt unk	cxr pos micro mix sympt unk	46	2877	48.0	
							64	9693	48.0	
							43	11366	48.0	
							45	12400	48.0	
Frimodt-Moller[56]	Asia	1961	Controlled trial - bacteriologically negative treatment	cumulative	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	11	86	12.0	Min-Inf
							18	86	24.0	
							25	86	36.0	
Styblo[28]	Europe	1967	Observational - prevalence survey follow-up	cumulative	cxr neg micro neg sympt neg	cxr unk micro pos sympt.mix	119	73000	12.0	
							156	73000	24.0	
							217	73000	36.0	
							241	73000	48.0	
Pamra[57]	Asia	1968	Controlled trial - bacteriologically negative treatment	cumulative	cxr pos micro neg sympt neg	cxr pos micro pos sympt pos	2	178	72.0	Min-Clin
						cxr pos micro pos sympt neg	55	178	72.0	Min-Sub
National Tuberculosis Insitute[27,29,38-44]	Asia	1968	Observational - population screen	cross sectional	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	23	329	18.0	Min-Inf
							36	271	60.0	Min-Inf
					cxr pos micro pos sympt unk	cxr pos micro neg sympt unk	86	269	18.0	Inf-Min
							70	178	36.0	Inf-Min

					cxr neg micro neg sympt unk	cxr pos micro pos sympt unk	44	31490	18.0	
							99	17936	60.0	
Aneja[58]	Asia	1977	Controlled trial - bacteriologically negative treatment	cross sectional	cxr pos micro neg sympt pos	cxr pos micro pos sympt unk	21	110	12.0	Min-Clin
Hong Kong Chest Service[59–62]	Asia	1981	Controlled trial - bacteriologically negative treatment	cumulative	cxr pos micro neg sympt mix	cxr pos micro pos sympt unk	40	176	3.0	Min-Inf
							49	176	6.0	
							61	176	12.0	
							67	176	18.0	
							69	176	24.0	
							70	176	30.0	
							71	176	36.0	
							71	176	60.0	
International Union Against Tuberculosis[32]	Europe	1982	Controlled trial - bacteriologically negative treatment	cumulative	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	97	6990	60.0	
Cowie[63]	Africa	1984	Controlled trial - bacteriologically negative treatment	cumulative	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	88	152	58.0	Min-Inf
Norregaard[64]	Europe	1985	Controlled trial - bacteriological	cumulative	cxr pos micro neg sympt mix	cxr pos micro pos sympt neg	6	28	60.0	Min-Sub

			ly negative treatment			cxr pos micro pos sympt pos	2	28	60.0	Min-Clin
Anastasatu[65]	Europe	1985	Controlled trial - bacteriologically negative treatment	cumulative	cxr pos micro neg sympt neg	cxr pos micro pos sympt unk	6	143	24.0	Min-Sub
Okada[30]	Asia	2004	Observational - prevalence survey follow-up	cross sectional	cxr pos micro neg sympt neg	cxr pos micro pos sympt unk	28	309	24.0	Min-Sub
				cumulative	cxr pos micro neg sympt neg	cxr pos micro pos sympt pos	23	309	24.0	Min-Clin
				cross sectional	cxr neg micro neg sympt neg	cxr pos micro pos sympt unk	32	21580	24.0	

Figure 3.5 shows the data that is used for the modelling in Chapter 4, with model states and transitions assigned as described in the methods and Figure 3.3 B-E. Nine of the potential 16 data collection type and transition combinations have data, with three transitions having no data to inform them with either data collection type (subclinical to minimal, subclinical to clinical, and clinical to subclinical). Where multiple study points are available, the cumulative data show a faster initial rise, followed by a slower plateau. The cross-sectional data show less of a pattern, although the proportion with disease described by the new state becomes lower over time, which would be expected with disease progressing or regressing further over time meaning fewer people are followed-up with disease in the new state.

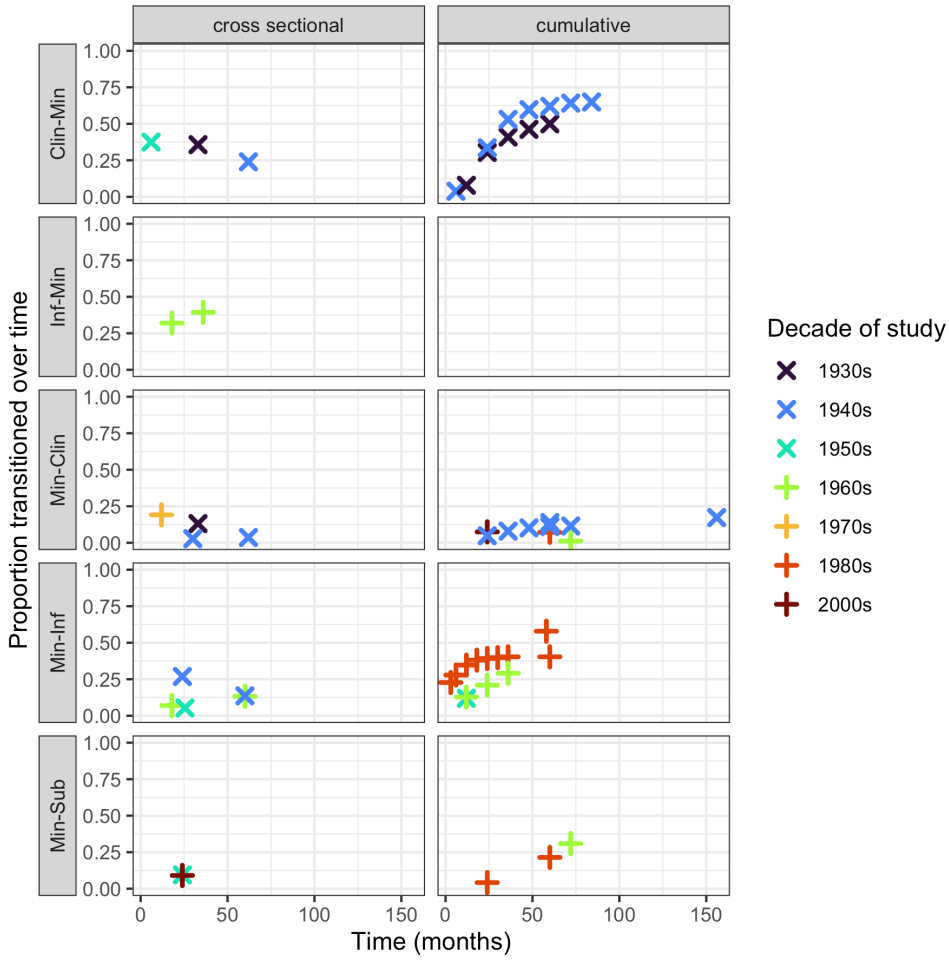


FIGURE 3.5: DATA USED IN THE MODELLING SEPARATED BY COLLECTION TYPE AND REPRESENTED MODEL TRANSITION, WITH COLOURS TO REPRESENT THE DECADE THE FINAL DATA WERE COLLECTED AND SHAPE TO REPRESENT IF THIS WAS BEFORE OR AFTER THE START OF MAINSTREAM ACCESS TO TREATMENT (WITH A CUT-OFF OF 1960). THE MODEL TRANSITIONS, SHOWN BY THE SPLITS ON THE LEFT HAND SIDE, ARE AS PRESENTED IN TABLE 3.2 WITH THE INITIAL STATE TO FINAL STATE, AS DESCRIBED IN SECTION 3.2.5. THE DATA COLLECTION TYPE, SHOWN BY THE SPLITS AT THE TOP, ARE AS PRESENTED IN TABLE 3.2 AND DISCUSSED IN SECTION 3.2.4.

As the data were collected over an interval of almost 100 years, and both the risk factors and treatments available have changed significantly over time, it was necessary to look at the data with an understanding of the difference in time of collection. Figure 3.5 splits the data into the

decade the study ended by colour, and whether this was before or after the widespread availability of effective treatments by shape (x for before 1960, + for 1960 onwards). The cumulative data for minimal to clinical had the greatest number of cohorts (5), with the widest spread of time (1940 - 2000) and despite these differences, the data all lie close to the same line. In three of the four other transition and study type combinations where there is data from both before and after the availability of treatment, old and new data share a point on the graph (cross-sectional minimal to subclinical and minimal to infectious, and cumulative minimal to infectious). As far as these data suggest, no difference between rates before and after treatment or over time can be concluded.

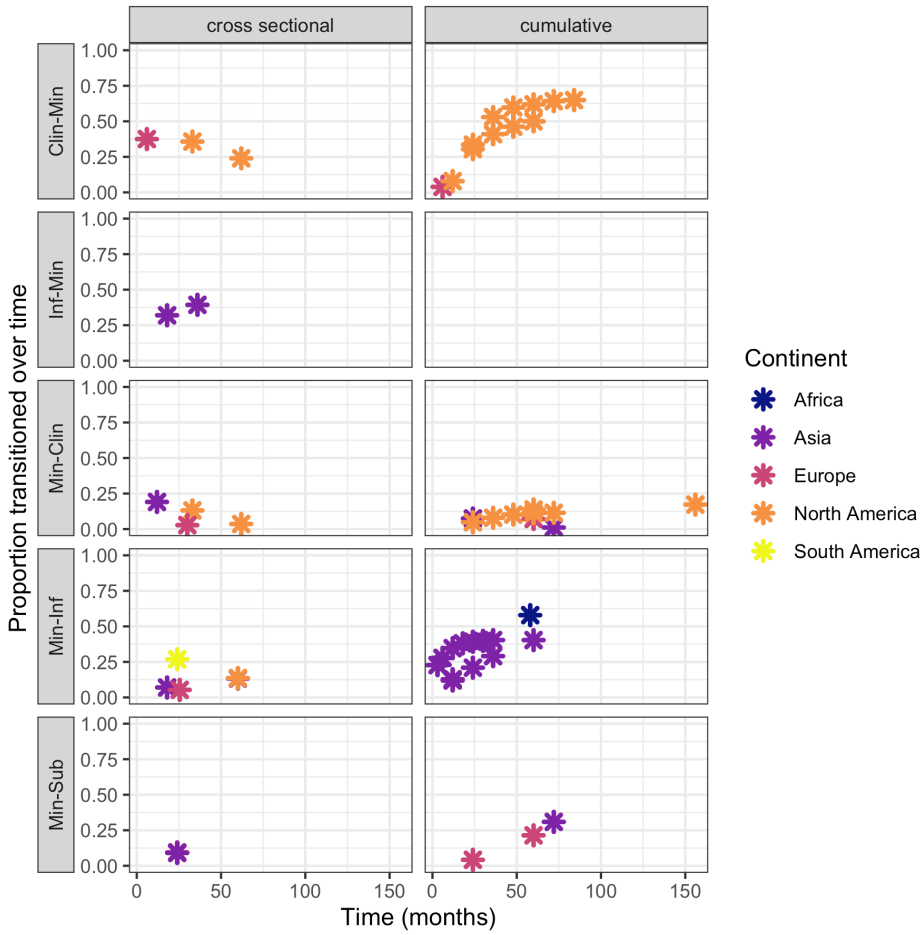


FIGURE 3.6: DATA USED IN THE MODELLING SEPARATED BY COLLECTION TYPE AND REPRESENTED MODEL TRANSITION, WITH COLOURS TO REPRESENT THE CONTINENT THE DATA WAS COLLECTED IN. THE MODEL TRANSITIONS, SHOWN BY THE SPLITS ON THE LEFT HAND SIDE, ARE AS PRESENTED IN TABLE 3.2 WITH THE INITIAL STATE TO FINAL STATE, AS DESCRIBED IN SECTION 3.2.5. THE DATA COLLECTION TYPE, SHOWN BY THE SPLITS AT THE TOP, ARE AS PRESENTED IN TABLE 3.2 AND DISCUSSED IN SECTION 3.2.4.

With the wide geographic variation in the data collection, I looked at whether there were any heterogeneities in the data based on continent of collection. Figure 3.6 splits the data by continent, data collection type, and model transition. There is insufficient evidence from this graph to suggest that there are geographical heterogeneities. Any difference between

geographical areas would show one area being higher (or lower) on the y-axis than other areas however, there is no strong pattern observable.

These data set the framework for the modelling presented in Chapter 4 and subsequent work on the progression from infection in Chapter 6.

3.4 Discussion

Main findings/Implications

This review has re-discovered the plethora of research conducted in the first half of the 20th century, and synthesised results from studies where testing and diagnosis were of similar quality to current standards. This data can now be used to help directly parameterise model structures looking to move away from the dichotomised infection and disease structure, to one that takes the varying spectrum of disease into account.

Comparison to other studies

Of the two previous historical reviews, this review was more similar to that conducted by Tiemersma *et al.* in that the search centred around pulmonary TB in adults.[12,13] However the focus beyond that of each study was different with Tiemersma *et al.* focusing on mortality and duration as opposed to disease progression.[13]. Of the 24 studies included here, only three were also included in the work by Tiemersma, and none of these were taken through to their qualitative analysis.[13] The key difference in study selection between the two studies was the interpretation of disease states. Whereas my review based extractions on radiology, bacteriology, and symptoms, Tiemersma *et al.* searched for mortality based on smear status of bacteriological positivity and interpreted the terms open and closed TB as smear positive and smear negative respectively.[13]

As mentioned in the results, there were three observational studies that concluded after the widespread introduction of treatments. Two of the studies followed people with or without evidence of disease on radiography but with negative bacteriology, thus maintaining standard of care.[29,31] The third study, by the National Tuberculosis Institute, was across an entire population in South India in the 1960s with no treatment regardless of test results.[28,30,39–45] Treatment for TB disease was available at this time, but it was not the standard of care in the area.[30] Although investigators made a limited quantity of isoniazid to local health clinics (to provide a month of treatment to those who had been diagnosed with bacteriologically positive disease) very few people took advantage of the medicine on offer and so the provision was stopped.[30] This is clearly unethical by current standards both in the limited nature of the treatment provided, which is not sufficient to cure disease, and in the fact that the treatment

was rescinded. The decision to include this study, despite the ethical issues, was taken with consensus decision that not using the data would make the burden placed on the study participants more onerous.[30]

Strengths and weaknesses

This review focussed on studies which reported disease based on the trio of radiography, bacteriology, and symptoms. This allowed for data collection that is directly translatable to current practice and understanding. Whereas other studies have assumed antiquated phrases of open and closed TB equate to smear positive and smear negative disease, the data collected here need no such assumptions.[13,14] However, there is no standardised definition for positive and negative when considering radiography, microbiology, or even symptoms. Whilst the test specific bias screen tried to mitigate this issue as far as possible, there are still differences between the studies, and the time difference between the first and last studies will have led to changes and improvements in testing and technology. There were also 190 studies excluded at full text for lack of bacteriological testing and a further 76 studies excluded due to the quality of reporting on testing. If a lower quality testing criteria had been applied, it is likely there would have been more studies available from which to extract data. However, accepting lower quality testing criteria would have lowered the quality of the data extracted. There would have been more assumptions made to describe the disease states, and thus more uncertainty in what the data were describing. This may have had limited to no impact on the model fitting results that follow in Chapter 4, but it would likely decrease progression parameters and increase regression parameters if, through misclassification, on average more presumed disease was not caused by *M.tb*. and therefore, the overall cohort would be less likely to progress to disease than one where everyone truly had TB.

The set-up of the review limited the studies available to extract data from. Firstly, due to the reviewers available, our languages were limited to English and German. Whilst there were still over 1500 titles eligible for full text screening in English and German, even without searching in other languages, 116 were excluded at full text in another language, and many more were excluded at title screening when the titles were not already translated into English or German. With more reviewers available, French, Dutch, and Spanish would be the next languages to have considered including at the original search, and at title and full text screenings. The second limitation of the review set-up was the strict twelve-month minimum follow-up required for inclusion. This was set up to ensure adequate follow-up of cohorts to be able to observe changes. However Table 3.2 and Figure 3.5 both show that 3 months would have been a sufficient minimum follow-up time. Only two studies were explicitly excluded for having less

than 12 months follow-up, but a shorter limit could have then increased the number of data sources by 8%.

There was very little information available on the infection and disease history of the cohorts. Whilst the perfect study to understand progression of infection and disease would follow people from birth with frequent repeated testing, this would be an unrealistic expectation for a study. Overall, we hoped to identify cohorts of people whose disease was known to have recently transitioned into one state, and then report the time at which a proportion had transitioned to a new state. Unfortunately, this was not reported. Instead data were collected on people with a known disease state and the time taken between registering in the study and progressing to a new state. Some studies had limits on how much time could have passed between initial detection of disease and enrolment in the study, but even then did not present the data of how the times between discovery and enrolment were distributed. Therefore the data presented are those of progression from a prevalent, rather than incident, disease state, a difference that will mean that the recorded time to progression will be an underestimate. However, with all the data presented, there do not seem to be large differences between progression from different studies, particularly when considering the cumulative data, which suggests that the impact may be limited.

The original data were collected up to 100 years ago and the world has changed significantly since then. Although no difference between rates could be concluded from the data collected and plotted, this does not mean that there is no difference, just that there is insufficient data to definitively describe one. Risk factors such as nutrition, smoking, and diabetes have changed over time but were not reported in the studies to make comparisons. It is likely that under-nutrition and smoking had a higher prevalence in the included populations than would be seen in many populations now, but that diabetes had a lower prevalence then than now. Lower prevalences of risk factors would slow progression rates, and higher prevalences would increase rates, but understanding the balance of two decreased risk factors and one increased, and how regression rates would similarly be affected is not possible without further work. A significant risk factor for TB is HIV which was not present in most of the populations included in this review and, for the few studies where it was, the number of people living with HIV, included in the cohort was minimal if there were any. Untreated HIV is likely to increase the risk of progression through disease however the extent of this increase and whether the underlying theoretical framework of disease is still applicable are unknown. Finally, following the introduction of treatments, came the development of drug resistant *M.tb*. Whilst drug resistant disease may progress and regress in the same way as drug susceptible disease, it is not possible to know, as none of the cohorts reported drug resistance.

Conclusions

This wide-reaching review has rediscovered previously lost data on the natural history of TB disease. Although the data have limitations, the span of the data in terms of time and geography, means that together a representative picture of the spectrum of TB disease can be drawn. While informative, more work is needed to convert these raw data into model parameters, which I will do in Chapter 4.

3.5 References

1. Onozaki, I. *et al.* National tuberculosis prevalence surveys in asia, 1990-2012: An overview of results and lessons learned. *Trop Med Int Health* **20**, 1128–1145 (2015).

2. Pai, M. *et al.* Tuberculosis. *Nat Rev Dis Primers* **2**, 16076 (2016).

3. Lin, P. L. & Flynn, J. L. The End of the Binary Era: Revisiting the Spectrum of Tuberculosis. *The Journal of Immunology* **201**, 2541–2548 (2018).

4. Frascella, B. *et al.* Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. *Clinical Infectious Diseases* **73**, e830–e841 (2021).

5. 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* **18**, e228–e238 (2018).

6. Waaler, H., Geser, A. & Andersen, S. The Use of Mathematical Models in the Study of the Epidemiology of Tuberculosis. *American Journal of Public Health and the Nations Health* **52**, 1002–1013 (1962).

7. Colijn, C., Cohen, T. & Murray, M. International Symposium on Mathematical and Computational Biology. in 123–148 (WORLD SCIENTIFIC, 2007). doi:10.1142/9789812708779_0008.

8. Fox, W., Ellard, G. A. & Mitchison, D. A. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease* **3**, S231–279 (1999).

9. Salazar-Austin, N., Dowdy, D. W., Chaisson, R. E. & Golub, J. E. Seventy years of tuberculosis prevention: Efficacy, effectiveness, toxicity, durability, and duration. *American Journal of Epidemiology* **188**, 2078–2085 (2019).

10. World Health Organization. Global tuberculosis report 2021. (World Health Organization, 2021).

11. Skavlem, J. H. Arrested Pulmonary Tuberculosis. *Diseases of the Chest* **4**, 8–11 (1938).

12. Marais, B. J. *et al.* The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era. 11 (2004).

13. Tiemersma, E. W., van der Werf, M. J., Borgdorff, M. W., Williams, B. G. & Nagelkerke, N. J. D. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PloS One* **6**, e17601 (2011).

14. Ragonnet, R., Trauer, J. M., Geard, N., Scott, N. & McBryde, E. S. Profiling mycobacterium tuberculosis transmission and the resulting disease burden in the five highest tuberculosis burden countries. *BMC Medicine* **17**, 208 (2019).

15. Szucs, E. Sputum Examination in So-called Closed Tuberculosis. *JAMA: The Journal of the American Medical Association* **86**, 946 (1926).

16. Sossen, B. *et al.* The Natural History of Untreated Pulmonary Tuberculosis in Adults: A Systematic Review and Meta-Analysis. doi:10.1101/2022.08.30.22279374.

17. *Covidence systematic review software, veritas health innovation, melbourne, australia. Available at www.covidence.org*.

18. National Tuberculosis Association. Diagnostic standards and classification of tuberculosis. (New York, N.Y., 1940).

19. Richards, A. S. *et al.* Quantifying progression and regression across the spectrum of pulmonary tuberculosis: A data synthesis study. *The Lancet Global Health* **0**, (2023).

20. Gothi, G. D. Natural History of Tuberculosis. *Indian Journal of Tuberculosis* **25**, (1977).

21. Barry, C. E. *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Reviews. Microbiology* **7**, 845–855 (2009).

22. Esmail, H., Dodd, P. J. & Houben, R. M. G. J. Tuberculosis transmission during the subclinical period: Could unrelated cough play a part? *The Lancet Respiratory Medicine* **6**, 244–246 (2018).

23. Houben, R. M. G. J. *et al.* Spotting the old foe — revisiting the case definition for TB. *The Lancet Respiratory Medicine* **7**, 199–201 (2019).

24. Houben, R. M. G. J., Esmail, H., Cobelens, F., Williams, C. M. L. & Coussens, A. K. Tuberculosis prevalence: beyond the tip of the iceberg. *The Lancet Respiratory Medicine* **10**, 537–539 (2022).

25. Richards, A. S. *et al. The natural history of TB disease-a synthesis of data to quantify progression and regression across the spectrum*. 2021.09.13.21263499 https://www.medrxiv.org/content/10.1101/2021.09.13.21263499v1 (2021).

26. Emery, J. C. *et al.* Estimating the contribution of subclinical tuberculosis disease to transmission an individual patient data analysis from prevalence surveys. doi:10.1101/2022.06.09.22276188.

27. Childress, W. G. Clinical evaluation, treatment, and follow-up of newly acquired tuberculous lesions. *New York State Journal of Medicine* **47**, 2560–4 (1947).

28. Gothi, G. D., Chakraborty, A. K., Krishnamurthy, V. V. & Banerjee, G. C. Prevalence and incidence of sputum negative active pulmonary tuberculosis and fate of pulmonary rradiological abnormalities found in rural population. *Indian Journal of Tuberculosis* **25**, 122–131 (1978).

29. Styblo, K., Dajkova, D., Kubik, A., Langerova, M. & Radkovsky, J. Epidemiological and clinical study of tuberculosis in the district of kolin, czechoslovakia. 56.

30. National Tuberculosis Institute, B. Tuberculosis in a rural population of South India: a five-year epidemiological study. *Bulletin of the World Health Organization* **51**, 473–88 (1974).

31. Okada, K. *et al.* Epidemiological impact of mass tuberculosis screening: A 2 year follow-up after a national prevalence survey. **NA**, (2012).

32. Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. *BMJ* **2**, 769–782 (1948).

33. Efficacy of various durations of isoniazid preventive therapy for tuberculosis: Five years of follow-up in the IUAT trial. *Bulletin of the World Health Organization* **60**, 555–564 (1982).

34. Beeuwkes, H., Hahn, R. G. & Putnam, P. A survey of persons exposed to tuberculosis in the household. *Am Rev Tuberc* **45**, 165–193 (1942).

35. Borgen L., Meyer SN. & Refsum E. Mass photofluorography, tuberculin testing, and BCG vaccination in the district of aker (norway) 1947-49. *Acta tuberculosea Scandinavica* **25**, 327–55 (1951).

36. Sikand, B. K., Narain, R. & Mathur, G. P. Incidence of TB as judged by re-surveys. A study of delhi police. *Indian Journal of Tuberculosis* **6**, 73–83 (1959).

37. GROTH-PETERSEN E., KNUDSEN J. & WILBEK E. Epidemiological basis of tuberculosis eradication in an advanced country. *Bulletin of the World Health Organization* **21**, 5–49 (1959).

38. Groth-Petersen, E., Gad, U. & Ostergaard, F. Mass chemoprophylaxis of tuberculosis. The acceptability and untoward side effects of isoniazid in a control study in Greenland. *The American Review of Respiratory Disease* **81**, 643–652 (1960).

39. Chakraborty, A. K. *et al.* Tuberculosis in a rural population of south india: Report on five surveys. *Indian Journal of Tuberculosis* **29**, 153–167 (1982).

40. Krishnamurthy, V. V., Nair, S. S., Gothi, G. D. & Chakraborty, A. K. Incidence of tuberculosis among newly infected population and in relation to the duration of infected status. *Indian Journal of Tuberculosis* **23**, (1976).

41. Krishnamurthy, V. V., Nair, S. S. & Gothi, G. D. A comparison of new cases (incidence cases) who had come from differeent epidemiological groups in the population. *Indian Journal of Tuberculosis* **25**, 144–146 (1978).

42. Gothi, G. D., Chakraborty, A. K. & Jayalakshmi, M. J. Incidence of sputum positive tuberculosis in different epidemiological grroups during five year follow up of a rural population in south india. *Indian Journal of Tuberculosis* **25**, 83–91 (1978).

43. Gothi, G. D., Chakraborty, A. K. & Banerjee, G. C. Interpretation of photofluorograms of active pulmonary tuberculosis patients found in epidemiological survey and their five year fate. *Indian Journal of Tuberculosis* **11**, 90–97 (1974).

44. Chakraborty, A. K. & Gothi, G. D. Relapses among naturally cureed cases of pulmmonary tuberculosis. *Indian Journal of Tuberculosis* **23**, 8–13 (1976).

45. Gothi, G. D., Nair, S. S., Chakraborty, A. K. & Ganapathy, K. T. Five year incidencee of tuberculosis and crude mortality in relation to non-specific tuberculin senssitivity. *Indian Journal of Tuberculosis* **23**, (1976).

46. Downes, J. The study of mortality among individuals with active pulmonary tuberculosis. **16**, 304–317 (1938).

47. Puffer, R. R., Stewart, H. C. & Gass, R. S. Tuberculosis according to age, sex, family history, and contact. *Am Rev Tuberc* **51**, 295–311 (1945).

48. Puelma, H. O. & Grebe, G. Analysis of one hundred cases of minimal pulmonary tuberculosis. *Dis Chest* **11**, 375–379 (1945).

49. BOBROWITZ ID., HURST A. & MARTIN M. Minimal tuberculosis; the prognosis and clinical significance of a sanatorium treated group. *American review of tuberculosis* **56**, 110–125 (1947).

50. BOBROWITZ ID. & HURST A. Minimal tuberculosis; problems in roentgenologic interpretation. *Radiology* **52**, 519–32 (1949).

51. Bosworth E. B. & Alling D. W. The after-history of pulmonary tuberculosis. I. Methods of evaluation. *American Review of Tuberculosis* **69**, 37–49 (1954).

52. Lincoln, N. S., Bosworth, E. B. & Alling, D. W. The after-history of pulmonary tuberculosis. III. Minimal tuberculosis. *Am Rev Tuberc* **70**, 15–31 (1954).

53. Alling DW., Lincoln NS. & Bosworth EB. The after-history of pulmonary tuberculosis. IV. Far advanced tuberculosis. *American review of tuberculosis* **70**, 995–1008 (1954).

54. Manser H. [Tuberculosis in aged and its course during sanatorium treatment]. *Schweizerische Zeitschrift fur Tuberkulose. Revue suisse de la tuberculose. Rivista svizzera della tubercolosi* **10**, 65–82 (1953).

55. Breu K. [Public health x-ray diagnosis of closed pulmonary tuberculosis later proved contagious]. *Beitrage zur Klinik der Tuberkulose und spezifischen Tuberkulose-Forschung* **111**, 437–44 (1954).

56. Tuberculosis Society of Scotland. A controlled trial of chemotherapy in pulmonary tuberculosis of doubtful activity. Report from the research committee of the tuberculosis society of scotland. *Tubercle* **39**, 129–137 (1958).

57. Frimodt-Moller, J. Results of treatment of non-bacillary tuberculosis in a domiciliary treatment programme, preliminary report. Proceedings of the 20th tuberculosis and chest diseases workers' conference. in 133 (1965).

58. Pamra, S. P. & Mathur, G. P. Effects of Chemoprophylaxis on Minimal Pulmonary Tuberculosis Lesions of Doubtful Activity. *Bull World Health Organ* **45**, 593–602 (1971).

59. Aneja, K. S., Gothi, G. D. & Samuel, G. E. R. Controlled study of the effect of spcific treatment on bacteriological status of 'suspect cases'. *Indian J Tuberc* **26**, 50–57 (1979).

60. Hong Kong Chest Service/Tuberculosis Research Centre, M. M. R. C. Sputum smear negative pulmonary tuberculosis controlled trial of 3-month and 2-month regimend of chmotherapy: First report. *The Lancet* **313**, 1361–1363 (1979).

61. Hong Kong Chest Service/Tuberculosis Research Centre, M. M. R. C. A study of the characteristics and course of sputum smear-negative pulmonary tuberculosis. *Tubercle* **62**, 155–167 (1981).

62. Hong Kong Chest Service/Tuberculosis Research Centre, M. M. R. C. A controlled trial of 2-month, 3-month, and 12-month regimens of chemotherapy for sputum smear-negative pulmonary tuberculosis: The results at 30 months. *Am Rev Respir Dis* **124**, 138–142 (1981).

63. Hong Kong Chest Service/Tuberculosis Research Centre, Madras/British Medical Research Council. A controlled trial of 2-month, 3-month, and 12-month regimens of chemotherapy for sputum-smear-negative pulmonary tuberculosis. *American Review of Respiratory Disease* **130**, 23–28 (1984).

64. Cowie, R. L. Diagnosis of sputum smear- and sputum culture-negative pulmonary tuberculosis. *SAMJ* **68**, 878 (1985).

65. Norregaard, J. Abacillary pulmonary tuberculosis. *Tubercle* **71**, 35–38 (1990).

66. Anastasatu, C., Berceea, O. & Corlan, E. Controlled clinical trial on smear negatice, x-ray positive new cases, with the view to establishing if and how to treat them. *Bull Int Union Tub* **60**, 108–109 (1985).

Chapter 4 Modelling the natural history of tuberculosis

4.1 Introduction

The systematic review presented in Chapter 3 sourced and collated data on the natural history of TB in untreated cohorts. Subsequently, I created a three-state compartmental model to describe the spectrum of disease and calibrated the model to the data collected in Chapter 3. This research paper has been uploaded as a pre-print to medRxiv and submitted to The Lancet Respiratory Medicine and is currently under review. I have reproduced the latest version under review. The supplementary materials are provided in Appendix B.

4.2 Cover Sheet

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

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

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

RESEARCH PAPER COVER SHEET

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

SECTION A – Student Details

Student ID Number	1700322	**Title**	Miss
First Name(s)	Alexandra S		
Surname/Family Name	Richards		
Thesis Title	Rediscovering the natural history of tuberculosis using modelling to combine historical and contemporary data		
Primary Supervisor	Rein Houben		

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

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

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

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	The Lancet Respiratory Medicine
Please list the paper's authors in the intended authorship order:	Alexandra S Richards, Bianca Sossen, Jon C Emery, Katherine C Horton, Torben Heinsohn, Beatrice Frascella, Federica Balzarini, Aurea Oradini-Alacreu, Brit Hacker, Anna Odone, Nicky McCreesh, Alison D Grant, Katharina Kranzer, Frank Cobelens, Hanif Esmail1, Rein MGJ Houben

Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	This paper uses the data from a previous systematic review and analyses it through a model. The initial idea was conceived by Rein Houben. I developed the aim of the manuscript, and the structure of the model with Rein Houben. I designed the fitting methodology with the help of Jon Emery and Katherine Horton. I coded the models, excluded the appropriate data from the systematic review and produced the figures and results. I drafted the manuscript and incorporated comments from all co-authors.

SECTION E

Student Signature	[redacted]
Date	19-09-2022

Supervisor Signature	Rein Houben [redacted]
Date	19-Sep-22

4.3 Title and Authors

The natural history of TB disease-a synthesis of data to quantify progression and regression across the spectrum

Authors: Alexandra S Richards[1,2], Bianca Sossen[3], Jon C Emery[1,2], Katherine C Horton[1,2], Torben Heinsohn[4,5,6], Beatrice Frascella[7], Federica Balzarini[7,8], Aurea Oradini-Alacreu[7], Brit Hacker[9], Anna Odone[10], Nicky McCreesh[1,2], Alison D Grant[11,12,13], Katharina Kranzer[14,15,16], Frank Cobelens[17], Hanif Esmail[18,19,20], Rein MGJ Houben[1,2]

Affiliations

1. TB Modelling Group, TB Centre, LSHTM, London, UK
2. Infectious Disease Epidemiology Department, LSHTM, London, UK
3. Department of Medicine, Faculty of Health Sciences, University of Cape Town, South Africa
4. Institute for Global Health, University College London
5. Helmholtz Centre for Infection Research
6. German Centre for Infection Research (DZIF), Braunschweig, Germany
7. School of Public Health, Vita-Salute San Raffaele University, Milan, Italy
8. Local Health Authority (ATS) of Bergamo, Bergamo, Italy
9. German Central Committee against tuberculosis (DZK e.V.), Berlin, Germany
10. Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia, Italy
11. TB Centre, LSHTM
12. Africa Health Research Institute, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, SA
13. School of Public Health, University of the Witwatersrand, South Africa
14. Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK
15. Biomedical Research and Training Institute, Harare, Zimbabwe
16. Division of Infectious Diseases and Tropical Medicine, University Hospital, LMU Munich, Munich, Germany
17. Department of Global Health and Amsterdam Institute for Global Health and Development, Amsterdam University Medical Centers, Amsterdam, Netherlands
18. Institute for Global Health, University College London, London, UK
19. Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, South Africa
20. MRC Clinical Trials Unit at University College London, UK

4.4 Abstract

Background: Prevalence surveys have found a substantial burden of subclinical (asymptomatic but infectious) TB, from which individuals can progress, regress or even persist in a chronic disease state. We aimed to quantify these pathways across the spectrum of TB disease.

Methods: We created a deterministic framework of TB disease with progression and regression between three states of pulmonary TB disease: minimal (non-infectious), subclinical, and clinical (symptomatic and infectious) disease. We estimated ranges for each parameter by considering all data from a systematic review in a Bayesian framework, enabling quantitative estimation of TB disease pathways.

Findings: Twenty-two studies contributed data from 5942 individuals. Results suggested that, after five years, 39.5%(95% uncertainty interval, UI, 31.4%-47.3%) of individuals with prevalent subclinical disease at baseline had recovered after regressing to minimal disease and 18.4%(95%UI, 13.7%-24%) had died from TB, leaving 14%(95%UI, 10.1%-18.6%) individuals still with infectious disease at five years, and the remainder with minimal disease at risk of further progression. Over the course of five years 50.2% (95%UI, 41.1%-59%) of the subclinical cohort never developed symptoms. For those with clinical disease at baseline, 45.9%(95%UI, 38.9%-52.1%) and 19.8%(95%UI, 15.1%-25.3%) had died or recovered from TB respectively, with the remainder in, or undulating between, the three disease states. The ten-year mortality of people with untreated prevalent infectious disease was 38%.

Interpretation: Our results show that for people with subclinical disease, classic clinical disease is neither inevitable nor an irreversible outcome. As such, reliance on symptom-based screening means a large proportion of people with infectious disease may never be detected.

Funding: TB Modelling and Analysis Consortium and European Research Council

4.5 Research in Context

Evidence before this study

In recent years the existence of a spectrum of TB disease has been re-accepted. The classic paradigm of disease is one active state of symptomatic presentation with bacteriologically positive sputum, now referred to as clinical disease. Within the spectrum, a subclinical phase (where people do not report symptoms but have bacteriologically positive sputum) has been widely accepted, due to prevalence surveys using chest radiography screening in addition to symptom screening. On average these prevalence surveys have found around 50% of people

with prevalent infectious TB had subclinical disease. There is also another state of minimal disease, or non-infectious disease, that is the earliest point on the disease spectrum after progression from infection. The likelihood or speed of natural progression, regression, or persistence of individuals across this spectrum remains unknown. As a consequence, the ability to accurately predict the impact of interventions has been limited. As individuals with bacteriologically-positive TB now receive treatment, contemporary data to inform the required transitions is highly limited. However, a large number of cohorts of patients were described in the pre-chemotherapy era. Until now, these data have not been synthesised to inform parameters to describe the natural history of TB disease.

Added value of this study

We synthesised data from historical and contemporary literature to explore the expected trajectories of individuals across the spectrum of TB disease. We considered a cohort of people with prevalent bacteriologically positive disease, with a 50/50 split of people with subclinical and clinical disease at baseline. We found that within five years, 29.6% of people recover from TB, defined as no chance of progressing to active disease without reinfection. However, we also find that 13.7% are still spending time infectious at the end of the five years. Our estimates for 10 year mortality and duration of symptoms before treatment aligned with the known and accepted values.

We also show that regression from subclinical disease results in a large reservoir of people with minimal disease, from which they can permanently recover, but can also progress again to subclinical disease. The undulating pathways that lead to regression and progression mean that 50.2% (41.1%-59%) of individuals with prevalent subclinical disease do not experience symptoms over the course of five years. This shows that clinical disease is neither a rapid, nor inevitable outcome of subclinical disease.

Implications of the available evidence

With these data-driven estimates of parameters, informed projections of the relative value of addressing minimal, subclinical, or clinical disease can now be provided. Given the known reservoir of prevalent subclinical disease and its contribution to transmission, efforts to diagnose and treat people with “earlier” stages of TB are likely to have a larger impact than strategies targeting clinical disease, particularly on individuals who never would have progressed to clinical disease.

4.6 Introduction

Despite effective treatment regimens being discovered in the 1950s, tuberculosis (TB) is still a major cause of morbidity and mortality globally. In 2019, there were an estimated 10 million people who fell ill with TB, and 1.4 million people died from TB.[1]

The current paradigm of TB disease assumes that there is a single state of active disease, with only progression to active disease from infection.[1] In reality people can move in both directions across a spectrum of disease.[2,3] After initial infection, individuals transitioning to pulmonary disease have been shown to progress through a state of minimal disease, where pathological changes due to *Mycobacterium tuberculosis* (*Mtb*) are visible on imaging techniques such as chest radiography (CXR) or computed tomography (CT), but individuals are not infectious (bacteriologically negative sputum).[4–6] Further progression leads to infectious disease (bacteriologically positive), within which there is a distinction between clinical and subclinical disease where individuals with subclinical disease do not report symptoms, but individuals with clinical disease report a prolonged cough or seek treatment due to their symptoms.[2,4,7] A recent review of national TB prevalence surveys found that around 50% of people with prevalent infectious disease have subclinical disease, and therefore will not be diagnosed by policies that rely on reported symptoms.[8]

While it is likely that not all individuals with minimal or subclinical disease will progress to clinical disease, the range or relative significance of alternative disease pathways is effectively unknown.[5,9,10] Efforts to quantify these pathways have remained limited by the absence of directly applicable parameter estimates.[11–13] A comprehensive review of literature has shown that many data sources exist, both historical and contemporary, which observed cohorts transitioning across the spectrum of disease.[14] However, no single study provides the overview of all trajectories across the different states, and with studies having varying durations, follow-up structures, and approaches to define and report disease states, the resulting heterogeneity complicates a simple comprehensive analysis.

Here we use a Bayesian framework to synthesise all available data to inform estimates of the natural rate of transition between minimal, subclinical, and clinical TB disease. We use these rates to simulate disease pathways in individuals, which we categorise to compare the frequency of different disease pathways in the population.

4.7 Methods

4.7.1 Data

The systematic review collected data that describe untreated cohorts at a minimum of two time points. Each time point, the cohort disease state was reported with a combination of CXR, bacteriology, and symptoms, with all studies required to directly report bacteriology or use standards set by the National Tuberculosis Association (NTA) that include bacteriology in the definitions.[6,14] The first time point described the state of a baseline group, and the second (and further) time points described the states of a subgroup after a recorded time. To enable synthesised analysis, two study types were included. In cumulative follow-up studies, individuals were closely followed and cumulatively recorded whether they had transitioned to a new state, either with a single, or multiple consecutive reporting points. After transitioning, individuals were excluded from follow-up. In cross-sectional follow-up studies, individuals were followed up at the single reported time point. Only their final state was recorded, without knowledge of any additional transitions that occurred before the study end. For inclusion in the analysis, a study needed to report on at least one cohort that transitioned between states, and included individuals needed to have, as a minimum, a CXR with signs interpreted as TB activity to fit in the minimal disease category. Detail on inclusion and exclusion criteria are included in the appendix (section S4).

Minimal disease was defined as bacteriologically negative, regardless of symptoms, based on observation from numerous studies that did not report differing progression rates and the poor specificity of symptoms in bacteriologically positive TB (see appendix section S3).[9,15] We adjusted the cohort size for people starting with minimal disease based on using tuberculin skin tests (TST) as a proxy for radiography changes that were truly caused by *Mtb* infection (see appendix, section S5.4). Whilst the systematic review collected data on whether a CXR was considered active or inactive, this distinction has not been carried forward here with a single grouping of minimal for both (see appendix, section S5.4).

For classification of outcomes, we assumed that when symptom status was only reported at enrolment, the symptom status persisted over the course of the study. If only positive symptom status was mentioned at the start but nothing about symptoms was mentioned at the end, even when symptoms had abated, this assumption could lead to a bias in data on progression towards clinical disease (increased rates of progression) as individuals whose disease progressed to subclinical rather than clinical would be combined with the group who progressed to clinical disease. Similarly, if only negative symptom status was mentioned at the start and progression was to symptom positive, but not reported, this could bias towards less

progression to clinical disease. To account for these potential biases, we performed a sensitivity analysis where any such data was instead used to inform progression to infectious disease, rather than rely on this assumption to determine between clinical and subclinical. Where symptom status was unknown at both time points, we classified people with bacteriologically positive disease as "infectious" as they could not be differentiated by symptoms to split between subclinical and clinical disease. If a paper referenced the NTA standards and used disease terminology of arrested, quiescent, or active from these standards, we have interpreted these to mean minimal, subclinical, and clinical respectively (see appendix sections S1 and S4).[6]

Both recovery from minimal disease and death from clinical disease were estimated through the calibration without data from the systematic review.[14] We assumed no knowledge on recovery, providing a uniform prior from 0 to 12 per year (representing no recovery to recovery of everyone within a month) as seen in Table 4.1. For death from clinical disease, the prior was taken from the estimated rate based on empirical data for mortality from "open" TB, which has a similar definition to clinical disease (see appendix section S2)[6,16,17]

Three further data points were included to inform the fit. Firstly, the median duration of infectious disease was included as a prior in the model assuming a normal distribution with a mean of 2 years and a standard deviation of 6 months (see appendix, section S5.3).[18] The other two data points relate to the distribution of disease at steady state. Firstly, the ratio of subclinical to clinical disease was included with a normally distributed prior with a mean of 1 and standard deviation of 0.25 to match a recent review of prevalence surveys (see appendix section S5.3).[19] While this is from populations with treatment, we do not have an equivalent source in absence of treatment. However, to compensate for this, we allowed wide priors for the model to settle on the best value given the data. For the purposes of fitting, we assume this prior value is what is found in a steady state situation. Secondly, the ratio of minimal to infectious disease was included with a normally distributed prior with a mean of 2.5 and wide standard deviation of 0.5 (see appendix section S5.3).[20] There is not a similar review to inform the ratio between minimal and infectious. We have used results from a post analysis of the 2016 Kenyan prevalence survey to provide an estimate for the prior, that is used with the same assumptions as the subclinical to clinical ratio.[53] Whilst this cannot be perfect, it again provides a prior range for the calibration to consider. This was paired with a wide uncertainty for the model to consider. In the survey, 10% of those screened with CXR were considered to have abnormalities indicative of TB, but on expert review, only 60% were truly considered abnormal. On Xpert testing, 90% of all originally screened as TB were negative, with 10% positive (which would give a 9:1 ratio). However, taking into account that only 60% were considered TB on expert review, that brings the ratio down to 5:1. Our own priors felt that this number was still too high,

so we halved the ratio (2.5:1) and used this as a midpoint for a wide prior. This estimate is also consistent with a repeated prevalence survey in Cambodia.[13] All these priors are outlined in Table 4.1. The equations used to calculate and fit to each of these priors are derived in appendix sections S5.2 and S5.3.

4.7.2 Data synthesis

To bring the data together we created a deterministic framework of TB disease, including the potential to move between the three disease states, as well as recovery from minimal disease and death from clinical TB disease (Figure 4.1, top row).

The transition rates were estimated by fitting to the data in a Bayesian framework. All data were considered simultaneously and a binomial distribution was chosen for the likelihood which allowed weighting by cohort size. Data points from cumulative studies were down-weighted so that the multiple data points of a cohort contributed as a single study (see appendix section S5.1).

We sampled the posterior values using a sequential Markov chain Monte-Carlo method (MCMC). An initial burn-in phase was used to find an optimal acceptance level, of between 25 and 35%, which was achieved by adapting the proposal distributions in both shape and scale. This was then discarded leaving chains with 10,000 iterations, which were visually inspected for convergence. The posterior parameter estimates came directly from the output of these chains, including the distribution, median, and 95% uncertainty intervals.

4.7.3 TB disease pathways

To quantify the different pathways through disease, we applied the parameter values from the Bayesian fitting to a cohort model that tracked individuals through their TB disease history. Once recovered or treated, individuals exited the model. As we were interested in the natural trajectory of an existing disease episode, we did not include re-infection. Each cohort tracked 10,000 people over 10 years and this was repeated 1,000 times with the parameters re-sampled at each repeat to capture uncertainty. We considered three cohort types; subclinical cohorts where all individuals initially had subclinical disease, clinical cohorts where all individuals initially had clinical disease, and mixed cohorts where half the individuals had subclinical disease and the other half had clinical disease.[8]

We ran the cohorts with and without the possibility of diagnosis and treatment. When included, treatment was implemented as a rate of 0.7 per year defining the chance of being diagnosed and successfully treated while symptomatic, which is a simple approximation of a 50% case

detection rate, when considering no competing risks, in a care system reliant on self-reported symptoms to initiate the care pathway.

We categorised the different pathways of disease observed over 12 month intervals. Where an individual received treatment, died, or recovered during those 12 months, they were classified as such. Individuals not classified as one of those outcomes could either have a static disease state or were classified as undulating (i.e. moving between two or more disease states). The cohort model reports disease state monthly, and if fewer than nine of the 12 months were spent in a single state, or an individual transitioned between states three or more times, the disease pathway was classified as undulating. Otherwise, nine months or more in a single disease state, and fewer than three transitions is classified as a static state, of the dominant state during that interval. See appendix section S6 for examples of these trajectories.

We report two durations of disease, one for infectious disease (subclinical and clinical), and one for all TB disease (minimal, subclinical, and clinical). The median duration of disease was calculated as the first point after the start of the simulation that fewer than 50% of the original cohort are present in one of the relevant states. We also recorded the number of months an individual spent with clinical disease before treatment or death, as well as throughout their disease episode, regardless of outcome.

Cumulative mortality from infectious TB disease in the absence of treatment was recorded at 10 years to allow comparison with existing estimates based on historical data.[21]

4.7.4 Sensitivity analyses

To test the robustness of the data synthesis results, we explored the impact of removing data provided from each study one at a time. In addition, the priors for mortality and the median duration of infectious disease were varied. For studies where symptoms were only provided in the start state of minimal, we re-ran the analysis with the transition for those studies to infectious rather than inferring a final state based on the initial symptoms. Sensitivities on the further analyses were also conducted; testing the parameter selection in the cohort model, introducing treatment at different case detection rates, and varying the thresholds for the definition of undulating disease.

All analyses were conducted in R version 4.0.3, using RStudio version 1.4.1103 and the Bayesian calibration was performed in LibBi version 1.4.5_3, using RBi version 0.10.3 and rbi.helpers version 0.3.2 as the interface.[22,23]

4.8 Results

4.8.1 Data synthesis

Twenty-two studies were included from the systematic review, providing 54 data points, describing 5942 people of whom 1034 transitioned between disease states. These cohorts were followed for intervals between 1923 and 2004, with studies conducted in North America (6), Europe (7), Asia (7), and one each from South America and Africa. In total there were 5 data points from minimal to subclinical, 14 data points from minimal to clinical, 15 data points from clinical to minimal, 18 data points from minimal to infectious, and 2 data points from infectious to minimal. Figure 4.1 shows the data points, including the relative weight of each data point, as indicated by the error bars. The best fit and uncertainty intervals to the data are shown by the lines and shaded area respectively in each plot. These data and the fitting are described in more detail in the appendix, sections S4 and S5.

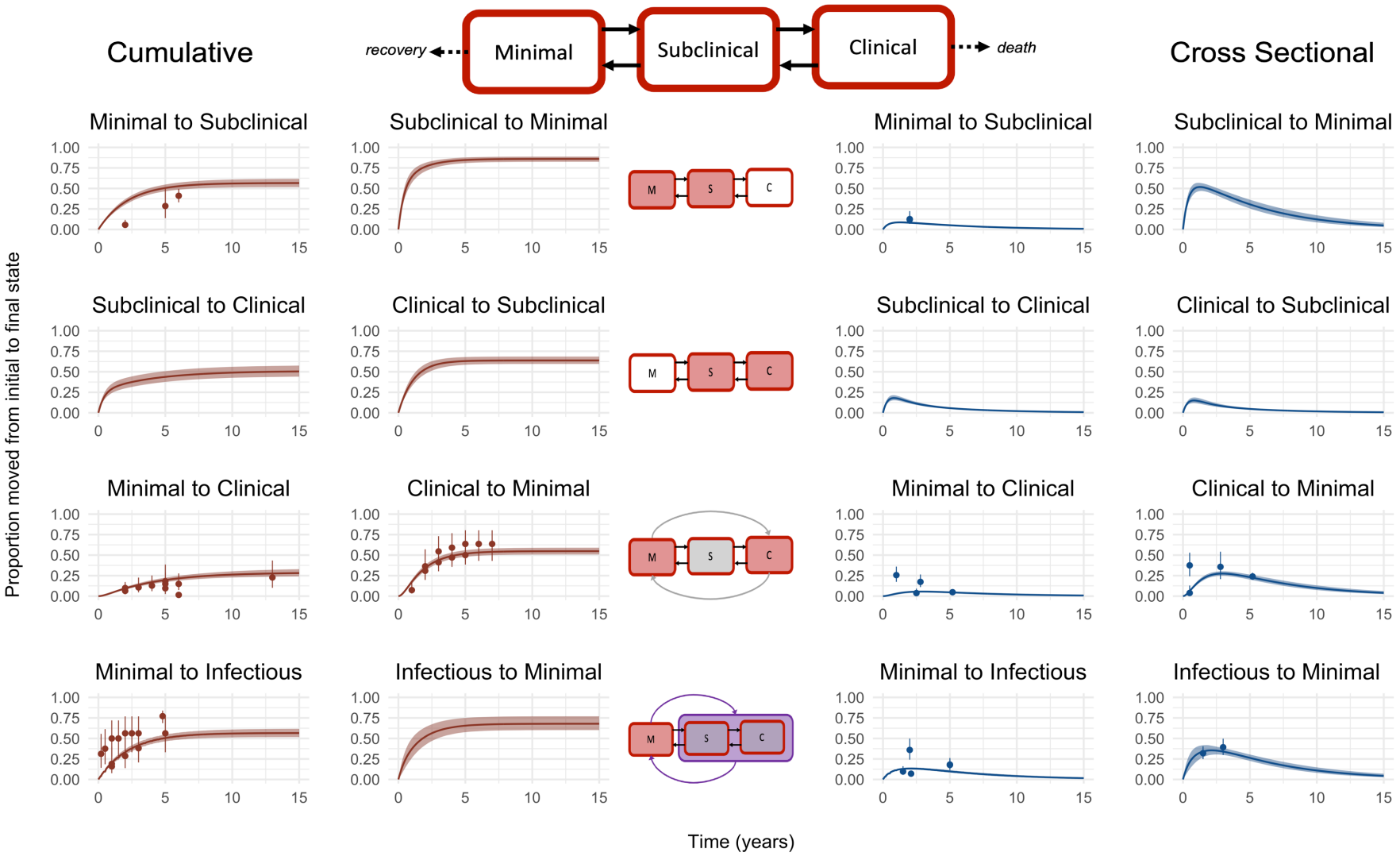


FIGURE 4.1: THE MODEL STRUCTURE AND THE RESULTS OF THE FITTING PROCESS COMPARED TO THE DATA PROVIDED. THE FIRST ROW SHOWS THE MODEL STRUCTURE, WITH THE SOLID LINES REPRESENTING PARAMETERS FITTED BY THE MODEL AND THE DOTTED LINE REPRESENTING FIXED PARAMETERS. EACH SMALL FIGURE SHOWS A UNIQUE COMBINATION OF DATA TRANSITION AND DATA TYPE. THE TWO LEFT HAND COLUMNS ARE THE FITS FOR THE CUMULATIVE DATA, AND THE TWO RIGHT HAND COLUMNS ARE THE FITS FOR THE CROSS-SECTIONAL DATA. THE ROWS SHOW PROGRESSION AND REGRESSION BETWEEN; 1) MINIMAL AND SUBCLINICAL, 2) SUBCLINICAL AND CLINICAL, 3) MINIMAL AND CLINICAL, 4) MINIMAL AND INFECTIOUS. THE MIDDLE COLUMN IS A VISUAL DESCRIPTION OF THE TRANSITION BEING FITTED ON EACH ROW. THE DOTS IN EACH GRAPH ARE THE POINT VALUES PROVIDED FROM EACH STUDY, WITH THE ERROR BARS REPRESENTING THE WEIGHTING OF THAT POINT VALUE AS PROVIDED IN THE FIT (SEE APPENDIX SECTION S5.1). THE SOLID LINE REPRESENTS THE MEDIAN TRAJECTORY OF THAT TRANSITION WITH THE CLOUD COVERING 95% OF THE SIMULATED TRAJECTORIES

Table 4.1 gives the median posterior parameter estimate for each model transition, with the 95% uncertainty interval. Uncertainty intervals for the parameters reflect the restricted parameter space when considering all the data simultaneously. Regression parameters were consistently higher than progression parameters.

TABLE 4.1: POSTERIOR PARAMETERS CALCULATED FROM THE FIT. THE PARAMETERS ARE PRESENTED AS ANNUAL RATES, ON THE TIMESCALE OF A YEAR, SO A PARAMETER VALUE OF ONE MEANS THAT, IN THE ABSENCE OF ANY OTHER COMPETING PARAMETERS, THE MEAN DURATION OF DISEASE IN THE INITIAL STATE IS ONE YEAR. PRIORS WITH A UNIFORM DISTRIBUTION ARE PRESENTED WITH THE MINIMUM AND MAXIMUM VALUE, PRIORS WITH A NORMAL DISTRIBUTION ARE PRESENTED WITH THE MEDIAN VALUE AND STANDARD DEVIATION

Parameter	Prior	Posterior estimate (95% uncertainty interval)	Source/comments
recovery from minimal	Uniform: 0-12/yr	0.19 (0.15, 0.24)	Model calibration
minimal to subclinical	Uniform: 0-12/yr	0.25 (0.22, 0.29)	Data synthesis/model calibration
subclinical to minimal	Uniform: 0-12/yr	1.52 (1.19, 1.94)	Data synthesis/model calibration
subclinical to clinical	Uniform: 0-12/yr	0.69 (0.54, 0.89)	Data synthesis/model calibration
clinical to subclinical	Uniform: 0-12/yr	0.57 (0.46, 0.71)	Data synthesis/model calibration
death from clinical	Normal: μ = 0.389/yr σ = 0.028	0.32 (0.27, 0.38)	Ragonnet et al/model calibration.[16]
duration of infectious disease	Normal: μ = 2 yrs σ = 0.5	0.99 (0.9, 1.11)	From NTI study/model calibration.[18] See appendix, section S5.3
prevalence ratio subclinical:clinical	Normal: μ = 1 yr σ = 0.25	1.3 (1, 1.68)	Frascella et al, Onozaki et al/model calibration.[19]
prevalence ratio minimal:infectious	Normal: μ = 2.5yrs σ = 0.5	3.93 (3.44, 4.52)	Mungai et al/model calibration.[20]

4.8.2 TB disease pathways

Figure 4.2 shows the relative proportions of each trajectory each year over the five years for simulated individuals with prevalent subclinical (Figure 4.2A) and clinical (Figure 4.2B) disease, as described in the methods. At five years, the proportion of people who die from TB is higher in the simulated cohort that starts with clinical disease (45.9%, 95% uncertainty interval, UI, 38.9%-52.1% vs 18.4%, 95%UI 13.7%-24%). The proportion of people in the minimal and recovered states is higher in the cohort starting with subclinical disease than it is in a cohort

starting with clinical disease. Two-thirds of individuals with subclinical TB had regressed to minimal or recovered after five years (67.4%, 95%UI, 53.9%-81.6%), compared to 40.3%, 95%UI, 31.6%-50.1% of individuals with clinical TB. At the end of five years, regardless of the initial state, approximately one in eight of the cohort remain in subclinical, clinical, or undulating in or out of those states (14%, 95%UI, 10.2%-18.5% for subclinical compared to 13.7%, 95%UI, 10.1%-18.2% for clinical).

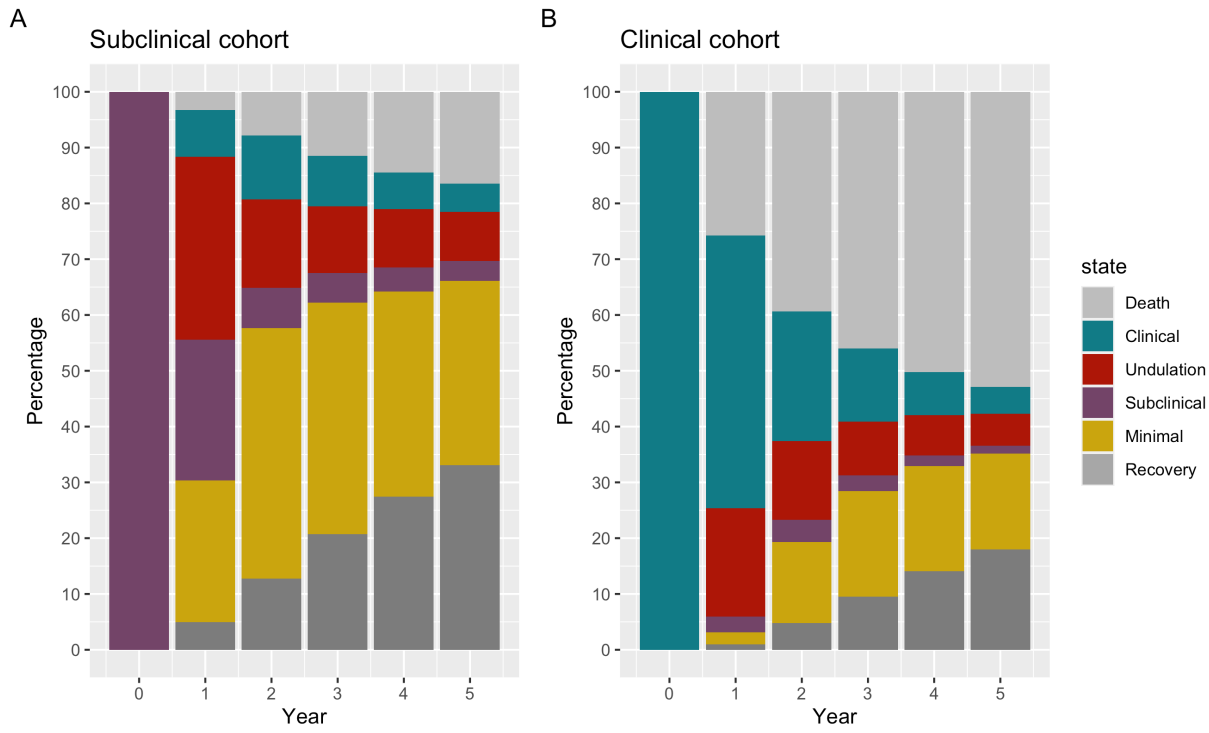


FIGURE 4.2: TRAJECTORIES OF DISEASE OVER TIME GIVEN DIFFERENT COHORT STARTS

Figure 4.3 shows the start and end states of a mixed subclinical and clinical cohort over five years. It highlights the patterns seen in Figure 4.2 with fewer people remaining with minimal disease at five years from the clinical half of the cohort, but similar numbers of people with clinical or subclinical disease from each half of the cohort. Death is twice as likely from the clinical half of cohort while recovery is twice as likely in the subclinical half of the cohort.

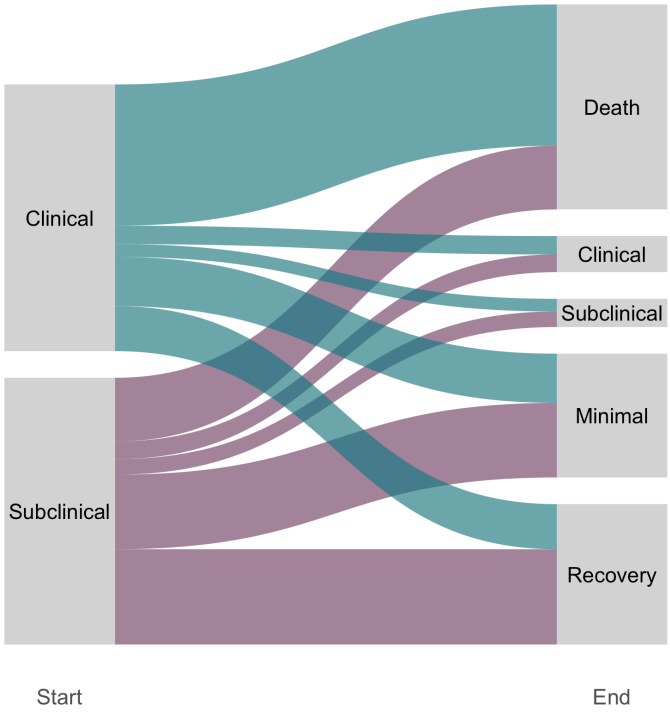


FIGURE 4.3: FINAL DISEASE STATE AFTER FIVE YEARS OF PEOPLE STARTING WITH SUBCLINICAL AND CLINICAL DISEASE

Looking at how many simulated individuals with subclinical disease at baseline never develop clinical disease, of those who completely recover within five years, 83.2% (95%UI, 78.9%-87.1%) never developed symptoms. This drops to 54.2% (95%UI, 45.8%-62.3%) in those with minimal disease at the end of five years, and further still to 25.9% (95%UI, 16%-38.7%) for those with subclinical disease at the end of five years. In total, in a cohort of individuals with subclinical disease at baseline, we estimate that 50.2% (95%UI, 41.1%-59%) would never develop symptoms.

In the absence of treatment, for a mixed cohort of half subclinical and half clinical at baseline, the median duration of infectious disease (subclinical and clinical) was 12 months (95%UI 10 - 15). If diagnosis and treatment were included, the median infectious period dropped to 8 months. However, we estimate the median duration of all TB disease, including the minimal state from which individuals can progress to infectious disease, to be 45 months without treatment and 35 with treatment. Illustrative figures can be found in the appendix, section S8.

In a simulated cohort with treatment available the duration of symptoms before death, regression to subclinical disease, or treatment varied between individuals from 1 months to 64 months, with a median of 6 months. For more on this distribution, see appendix section S7.

In a mixed cohort 37.7% (95%UI, 31.2%-45.2%) of the cohort died from TB within 10 years. A cohort of people with clinical disease at baseline had a higher proportion die from TB than one

with subclinical disease at baseline (51.2% (95%UI, 43.6%-58.2%) vs 24.2% (95%UI, 18.1%-31.9%)), as seen in Figure 4.3.

4.8.3 Sensitivity analyses

Where changes in median parameter values were observed in the sensitivity analyses, these did not transfer through to the other key output metrics. In-depth comparisons of the sensitivity analyses can be found in the appendix, section S9.

4.9 Discussion

4.9.1 Summary of findings

We synthesized available data from a systematic review of untreated cohorts of TB disease to parameterise progression and regression between minimal, subclinical, and clinical TB disease. With these parameters we quantified the pathways of individuals across the spectrum. Our results show that non-linear disease trajectories are common, and there is a high rate of natural regression from subclinical disease. This is demonstrated by the large proportion of individuals at five years who have regressed to minimal disease or fully recovered. Although the risk of death from clinical disease is high, of those who do not die, the majority have also regressed to minimal disease or have recovered over the course of five years. Of those who still have infectious disease after five years, regardless of starting point, over half undulate between states rather than remaining with a single state of disease long term. Where symptom screening is used to detect people with infectious disease, many people may not be offered timely treatment due to intermittent or absent symptoms, meaning either they progress to more severe disease or unknowingly contribute to further transmission of TB.

4.9.2 Interpretations

Of the four parameters estimated through fitting to the data, we found the regression rate from subclinical to minimal greatly exceed the other rates. While a high rate of recovery from subclinical to minimal could suggest that the majority of infectious TB disease would resolve without intervention, the relative sizes of the states as well as the high mortality rate from clinical disease likely counteract this trend, as can be seen in Figure 4.3. For example, given many more individuals become infected with *Mtb* than develop infectious disease, it is reasonable to assume that the population with minimal disease is much larger than those with clinical disease which could mean the absolute number of individuals with potential to progress towards infectious disease will exceed those regressing.[1] A rate of 1.52 per year would translate to a mean duration of 8 months with subclinical disease before regressing to minimal disease when not considering the transition to clinical disease, so although the rate is much higher than the other parameters, the duration it represents is not insignificant. The probability of

recurrence of disease in individuals who have regressed to minimal disease is also higher than the probability of fully recovering which creates a loop of undulation, where an individual progresses and regresses in and out of infectious disease. As such, the benefits of wider treatment, to people with subclinical disease or even with minimal disease, may be greater than has previously been assumed. For example, the rapid decline of prevalent infectious TB in China could be explained in part by the extensive treatment of individuals based on CXR alone in addition to those with a bacteriological diagnosis, which would have reduced the reservoir of minimal disease, as the combined rate of treatment and recovery exceeded that of progression from minimal.[24]

Each of the parameters are presented with a 95% uncertainty interval, some of which are narrower than others. Notably, the uncertainty interval around the value for subclinical to minimal is the largest of the main parameters. This may reflect the lack of data to directly fit this transition. However, when viewed in Figure 4.1, these wide intervals do not translate to overwhelmingly wide bands, illustrating how the simultaneous consideration of all data restrict the potential parameter values that provide a reasonable fit to all the data points. In practice, progression and regression will be more variable between individuals and between populations, driven by variations in e.g. HIV status, diabetes, malnutrition, and gender.[28] The cohorts represented in this analysis mostly comprised HIV-negative individuals, and the prevalence of other variables was unknown (e.g. malnutrition). While it is possible that these factors affect all or a subset of parameters, our results are based on a range of populations, times, and geography, and thereby provide an improvement over the limited data currently supporting estimates of TB progression and regression parameters.[16,29]

The parameters were estimated by fitting to the data collected from the systematic review alongside a prior of a two year infectious disease duration, and the ratios of subclinical to clinical disease and minimal to infectious disease at a steady state.[8,14,18–20] The posterior estimate for infectious disease duration is at the lower end of the expected range set in the prior, however the data that informed the prior estimate had variation cannot be represented with a point value.[18,21] We used previously reported values on TB mortality from smear positive disease as the prior for TB mortality from clinical disease.[16] Although our point estimate was slightly lower than the provided prior, the 10-year mortality for people with untreated prevalent infectious TB was comparable to the accepted value of 40%.[16,21] We also extracted duration of symptoms before treatment. Systematic reviews of self-reported symptom duration usually state between one and three months of symptoms prior to treatment, whereas we found a median of six months.[30] While our results find a longer duration, it is likely that self-reported symptoms are more under- than over-estimated.

We cannot directly compare our parameter estimates with current models, as no other models have split the disease spectrum into three states. Ku et al used the proportion subclinical in prevalence surveys to divide total duration of disease, and did not consider undulation between states, however, our estimate of a median of 6 months falls within the range of symptom duration reported.[31] The WHO technical appendix includes regression from "active" disease for those who self-cure or die before treatment as a single parameter with no further consideration of a spectrum of disease.[32] Salvatore et al have represented disease as progression and regression along a single continuum of disease burden, defined as a composite of bacteriology, pathology, and symptoms.[11] The potential rates of progression and regression were wide and overlapped with our data driven estimates.[11] A recent systematic review from Menzies et al on progression only considered a single disease state.[29] Some of these studies split the active disease state by bacteriological load (smear positive or smear negative, whilst still bacteriologically positive), however we have instead focused on bacteriological positivity alone, in line with the current reporting framework.[1]

We reported undulating disease based on a fixed threshold of nine months, which is a subjective choice. While a shift in threshold would change the proportion qualified as "undulating", the underlying movement between states will remain the same (see appendix, sections S6 and S9.2.3).

An important finding is the large proportion of people with subclinical disease who may never develop symptoms, i.e. clinical disease. Although our results suggest many regress towards minimal disease, or even recover completely, this does not mitigate the time these people spend with infectious subclinical disease or the non-infectious period where the *Mtb* infection remains active. As we show, in a population without treatment, almost half of those who had subclinical disease at baseline, still have TB disease five years later.

4.9.3 Limitations

Despite the extensive literature review, few data points could directly inform parameters. By including data on transitions between minimal and clinical, and between minimal and infectious, we were able to restrict the likely parameter space. For example, the cumulative minimal to infectious data provide a lower bound for the minimal to subclinical transition. While the chosen model structure will drive some of the results, limitations in the available data prohibit a more complicated model structure. In addition our three-state linear model structure is in line with historical and recent conceptualisations of the spectrum of TB disease.[2,3,6,8,13]

Both our data and simulations start from prevalent disease (minimal, subclinical, and clinical) without knowledge of previous disease trajectory and a single rate of transition for all. As such

the parameters represent a mix of both recent and more distal *Mtb* infections, where some individuals are rapidly progressing, as well as individuals who are undulating, or on their way to recovery. However this is a reflection of current prevalent TB states in a population, as found in prevalence surveys.[8,9] Prevalent disease is the immediate driver of TB morbidity, mortality, and transmission, and as such the population that TB policies look to address.

4.9.4 Conclusions

We estimate that only half of all people with subclinical TB disease will progress to clinical disease. As such we show a flaw in the assumption that targeting clinical disease will enable care for all individuals suffering from TB disease, or interrupt transmission from infectious disease. Our work also highlights an important question; where should the threshold be set for TB disease that requires treatment. While the current threshold of infectious disease can relatively easily be confirmed, minimal (i.e. bacteriologically-negative) disease is an important reservoir of potential future transmission in the population and has a substantial risk of progression to more advanced disease. To comprehensively interrupt current and future transmission, we need to expand our interventions to those which can detect and treat subclinical and even minimal disease if possible. These may have substantial individual and population benefits which with these parameter estimates, we can now more reliably quantify.[10,13,33]

4.10 Acknowledgements

ASR, JCE, KCH and RMGJH received funding from the European Research Council (ERC) under the European Union's Horizon 2020 programme (Starting Grant Action Number 757699). KCH is supported by the UK FCDO (Leaving no-one behind: transforming gendered pathways to health for TB). This research has been partially funded by UK aid from the UK government (to KCH); however the views expressed do not necessarily reflect the UK government's official policies.
HE, BS, received funding from the TB Modelling and Analysis Consortium (TB MAC)
HE received funding from the Medical Research Council (Grant ref: MR/V00476X/1)
NM is funded by the Wellcome Trust (218261/Z/19/Z)
FC received no funding for this work.
KK received no funding for this work.
FB, BF, AOA, AO received no funding for this work.

4.11 References

1 WHO. Global tuberculosis report 2020. 2020https://www.who.int/publications-detail-redirect/9789240013131 (accessed 10 Jun2021).

2 Barry CE, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J *et al.* The spectrum of latent tuberculosis: Rethinking the biology and intervention strategies. *Nat Rev Microbiol* 2009; **7**: 845–855.

3 GOTHI GD. NATURAL HISTORY OF TUBERCULOSIS. *Indian Journal of Tuberculosis* 1977; **25**.

4 Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG *et al.* Incipient and subclinical tuberculosis: A clinical review of early stages and progression of infection. *Clinical Microbiology Reviews* 2018; **31**: 24.

5 Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Coussens AK *et al.* Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18F]fluoro-d-glucose positron emission and computed tomography. *Nat Med* 2016; **22**: 1090–3.

6 Association NT. *Diagnostic standards and classification of tuberculosis.* 1940 ed. New York, N.Y., 1940http://hdl.handle.net/2027/coo.31924089435949.

7 Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC *et al.* Tuberculosis. *Nat Rev Dis Primers* 2016; **2**: 16076.

8 Frascella B, Richards AS, Sossen B, Emery JC, Odone A, Law I *et al.* Subclinical tuberculosis disease - a review and analysis of prevalence surveys to inform definitions, burden, associations and screening methodology. *Clin Infect Dis* 2020. doi:10.1093/cid/ciaa1402.

9 Okada K, Onozaki I, Yamada N, Yoshiyama T, Miura T, Saint S *et al.* Epidemiological impact of mass tuberculosis screening: A 2 year follow-up after a national prevalence survey. 2012; **NA**.

10 Houben RMGJ, Esmail H, Emery JC, Joslyn LR, McQuaid CF, Menzies NA *et al.* Spotting the old foe—revisiting the case definition for TB. *The Lancet Respiratory Medicine* 2019; **7**: 199–201.

11 Salvatore PP, Proano A, Kendall EA, Gilman RH, Dowdy DW. Linking individual natural history to population outcomes in tuberculosis. *J Infect Dis* 2017. doi:10.1093/infdis/jix555.

12 Arinaminpathy N, Dowdy D. Understanding the incremental value of novel diagnostic tests for tuberculosis. *Nature* 2015; **528**: S60–7.

13 Dowdy DW, Basu S, Andrews JR. Is passive diagnosis enough? *Am J Respir Crit Care Med* 2013; **187**: 543–551.

14 Sossen B. Systematic review. 2021.

15 Organisation. WH. Rapid communication on systematic screening for tuberculosis. World Health Organisation, 2020https://extranet.who.int/iris/restricted/bitstream/handle/10665/337372/9789240016552-eng.pdf (accessed 7 Dec2020).

16 Ragonnet R, Flegg JA, Brilleman SL, Tiemersma EW, Melsew YA, McBryde ES *et al.* Revisiting the natural history of pulmonary tuberculosis: A bayesian estimation of natural recovery and mortality rates. *Clin Infect Dis* 2020. doi:10.1093/cid/ciaa602.

17 Blahd M, Leslie EI, Rosenthal SR. Infectiousness of the 'closed case' in tuberculosis. *Am J Public Health Nations Health* 1946; **36**: 723–726.

18 National Tuberculosis Institute B. Tuberculosis in a rural population of south india: A five-year epidemiological study. *Bull World Health Organ* 1974; **51**: 473–88.

19 Onozaki I, Law I, Sismanidis C, Zignol M, Glaziou P, Floyd K. National tuberculosis prevalence surveys in asia, 1990-2012: An overview of results and lessons learned. *Trop Med Int Health* 2015; **20**: 1128–1145.

20 Mungai BN, Joekes E, Masini E, Obasi A, Manduku V, Mugi B *et al.* 'If not TB, what could it be?' Chest x-ray findings from the 2016 kenya tuberculosis prevalence survey. *Thorax* 2021; **76**: 607–614.

21 Tiemersma EW, Werf MJ van der, Borgdorff MW, Williams BG, Nagelkerke NJD. Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: A systematic review. *PLoS ONE* 2011; **6**: e17601.

22 Murray LM. Bayesian state-space modelling on high-performance hardware using LibBi. *arXiv:13063277 [stat]* 2013.http://arxiv.org/abs/1306.3277 (accessed 6 Aug2021).

23 R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria, 2020https://www.R-project.org/.

24 Wang L, Zhang H, Ruan Y, Chin DP, Xia Y, Cheng S *et al.* Tuberculosis prevalence in china, 1990-2010; a longitudinal analysis of national survey data. *Lancet* 2014; **383**: 2057–64.

25 Getahun H, Gunneberg C, Granich R, Nunn P. HIV infection-associated tuberculosis: The epidemiology and the response. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2010; **50 Suppl 3**: S201–7.

26 Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: A systematic review of 13 observational studies. *PLoS Medicine* 2008; **5**: e152.

27 Lonnroth K, Williams BG, Cegielski P, Dye C. A consistent log-linear relationship between tuberculosis incidence and body mass index. *International Journal of Epidemiology* 2010; **39**: 149–155.

28 Holmes CB, Hausler H, Nunn P. A review of sex differences in the epidemiology of tuberculosis. *The International Journal of Tuberculosis and Lung Disease* 1998; **2**: 96–104.

29 Menzies NA, Wolf E, Connors D, Bellerose M, Sbarra AN, Cohen T *et al.* Progression from latent infection to active disease in dynamic tuberculosis transmission models: A systematic review of the validity of modelling assumptions. *Lancet Infect Dis* 2018; **18**: e228–e238.

30 Storla DG, Yimer S, Bjune GA. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC public health* 2008; **8**: 15.

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

32 Glaziou P, Dodd PJ, Dean A, Floyd K. Methods used by WHO to estimate the global burden of TB disease. 2020.

33 Kendall EA, Shrestha S, Dowdy DW. The epidemiological importance of subclinical tuberculosis: A critical re-appraisal. *Am J Respir Crit Care Med* 2020. doi:10.1164/rccm.202006-2394PP.

Chapter 5 The impact of population level screening for TB disease

5.1 Introduction

Mass radiography screening to detect people with tuberculosis disease (TB) was used extensively in Europe and the USA in the early 20th century.[1,2] Over time, as prevalence reduced and mass screening became less cost effective, active case detection methods were dropped.[3–8] Most case detection now relies on the understanding that the majority of people with TB disease will present to a healthcare facility at some point on account of experiencing symptoms.[3–6] For those who do present to a healthcare facility, the current treatment plan, known as DOTS (Directly Observed Treatment, Short course), has been highly effective in treating people with symptomatic, bacteriologically positive TB disease.[3,4]However, recent prevalence surveys have highlighted that not all individuals with TB disease experience or are aware of symptoms.[9,10] Instead around 50% of people with bacteriologically positive disease did not report TB symptoms at screening, and in some areas this was found to be as high as 80%.[9,10]

The description of TB disease has evolved recently from the previous binary paradigm of latent infection and active disease (sometimes separated into smear negative and smear positive disease), to a wider spectrum that more fully encompasses the different stages and presentations of disease.[11–15] I split disease into three states; clinical, subclinical, and minimal.[13] Clinical disease aligns well with the previous definition of active disease, symptomatic and bacteriologically positive disease.[13] Subclinical disease is still bacteriologically positive, however the person is not aware of, experiencing, or reporting any symptoms.[13] Minimal disease is bacteriologically negative, but with radiological signs of disease in the lungs, visible by radiography or other imaging modality.[13] Prior to these disease states is the progression from new infection to disease.[13] Disease progression is not unidirectional, with regression possible, and further progression not inevitable.[12–14] This is further discussed in Chapter 6, where the choice of model structure is explained and data and methods of the parameterisation for the model are described.

The WHO currently only recommends systematic screening of the general population for areas where the prevalence of TB is at least 0.5% (500 cases per 100,000 population) due to high levels of undetected disease.[16] Within this remit, active screening methods in the community have been trialled with varying success.[17,18] Before wider roll-out of a programme, a range of questions need to be answered, including what diagnostic tools performs best, and how often to

screen. Here I focussed on developing a modelling framework in order to evaluate the performance of different screening algorithms and diagnostics against a range of epidemiological and resource metrics.

To assess the impact of screening programmes, I propose four key metrics, based on previous studies of TB case finding.[17–20] Metric one is the decrease in TB prevalence over time, which is defined as the reduction in all infectious disease (subclinical and clinical disease) in the population.[6] Metric two is how many people without TB disease are started on treatment (i.e. false-positive diagnoses) and metric three is how many people should have been treated for disease, but were missed as a result of false negative readings (i.e. false-negative diagnoses). Finally, metric four is how many positive confirmatory tests are returned for each person with infectious disease detected (the number needed to treat (NNT)).

This work aims to utilise newly developed knowledge on the spectrum of disease to help understand the wider impact of different screening tests.[13] It builds on the model structure initially developed in Chapter 4, and the extension to include progression from infection in Chapter 6, along with work to estimate the relative transmissibility of subclinical disease.[13,21,22]

5.2 Methods

5.2.1 Model structure, definitions, and set-up

The model structure uses the three-state model structure from Chapter 4 and the progression from infection from Chapter 6. By extending to a transmission model, two states and three transitions had to be added beyond those presented in Chapter 6. The full model structure can be seen in Figure 5.1A. Other than disease state, the population is assumed to be homogeneous.

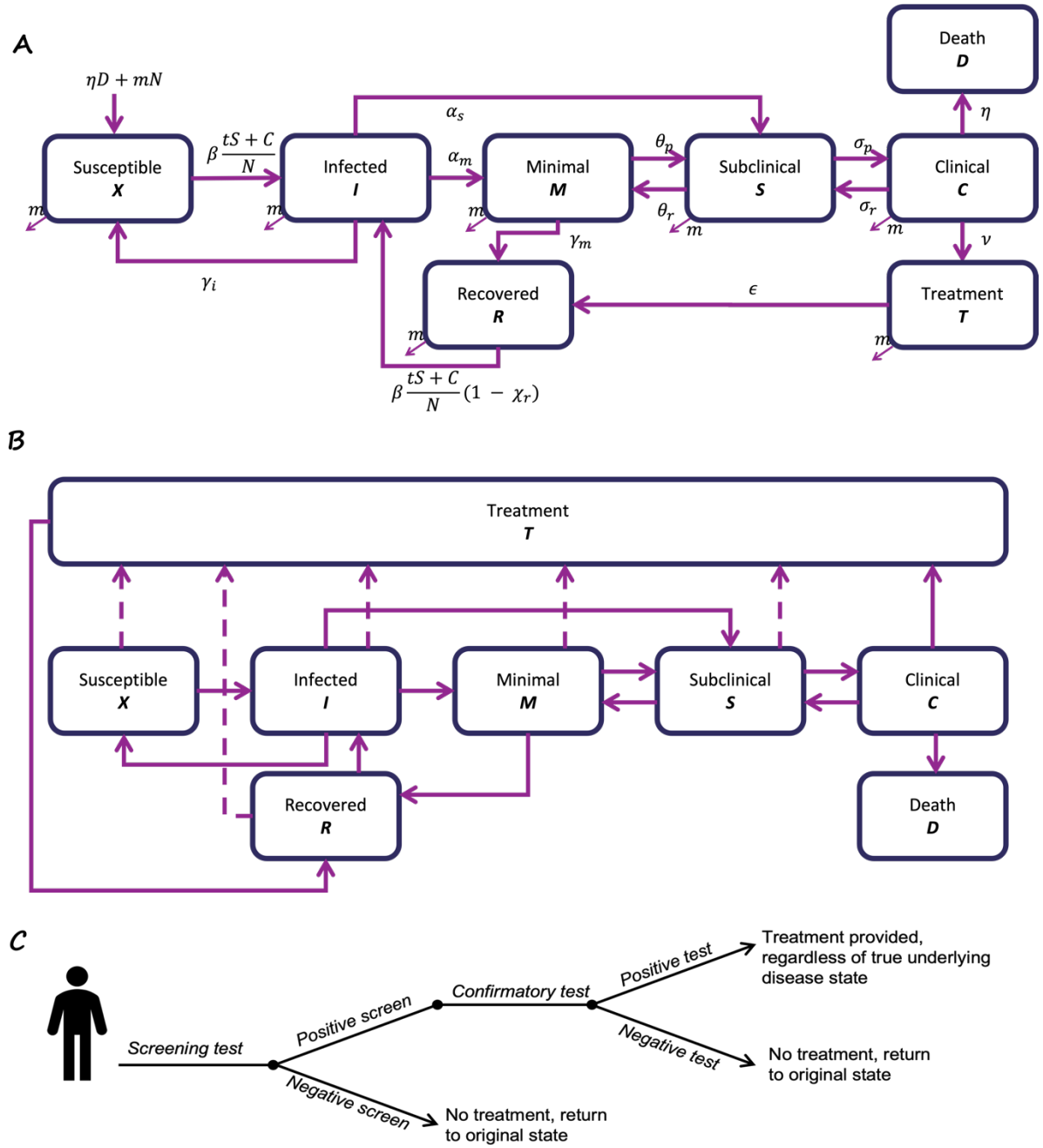


FIGURE 5.1: THE MODEL STRUCTURE USED IN THIS ANALYSIS OF DIFFERENT SCREENING TESTS. A SHOWS THE BASELINE MODEL, AND B SHOWS THE INTERVENTION MODEL, WITH DOTTED LINES SHOWING PATHS ONLY AVAILABLE DURING THE INTERVENTION. MINIMAL, SUBCLINICAL, AND CLINICAL ARE THE THREE DISEASE STATES DESCRIBED IN CHAPTER 4, WITH RECOVERY FROM MINIMAL AND DEATH FROM CLINICAL. PROGRESSION FROM INFECTION IS AS DESCRIBED IN CHAPTER 6, TO BOTH MINIMAL AND SUBCLINICAL, WITH RECOVERY. NEW INFECTIONS OCCUR THROUGH TRANSMISSION FROM PEOPLE WITH SUBCLINICAL AND CLINICAL DISEASE, WITH ADDITIONAL FACTORS OF RELATIVE TRANSMISSIBILITY, AND A CONTACT PARAMETER. TREATMENT IS A FIXED RATE FROM CLINICAL, WITH A FIXED RATE TO RECOVERY TO REPRESENT THE DURATION OF TREATMENT. PROGRESSION TO INFECTION OCCURS FROM SUSCEPTIBLE OR RECOVERED DEPENDENT ON THE PROPORTION OF PEOPLE WITH SUBCLINICAL AND CLINICAL DISEASE, AND THE RELATIVE TRANSMISSIBILITY OF SUBCLINICAL DISEASE, ALONG WITH A PROTECTION FROM REINFECTION IN THOSE WHO HAVE PREVIOUSLY RECOVERED. A LIFE EXPECTANCY, $\frac{1}{m}$ DEFINES THE RATE OF NATURAL DEATH FROM EVERY COMPARTMENT. FIGURE C SHOWS THE SCREENING ALGORITHM FOR EACH PERSON IN THE COMMUNITY, REGARDLESS OF TRUE UNDERLYING DISEASE STATE, IF A PERSON IS SCREENED NEGATIVE, THEY DO NOT PROGRESS TO A CONFIRMATORY TEST AND DO NOT RECEIVE TREATMENT. IF A PERSON SCREENS POSITIVE, THEY THEN PROGRESS TO A CONFIRMATORY TEST AND SUBSEQUENTLY AN INDIVIDUAL WITH A POSITIVE. CONFIRMATORY TEST WOULD

RECEIVE TREATMENT, AND AN INDIVIDUAL WITH A NEGATIVE CONFIRMATORY TEST WOULD NOT RECEIVE TREATMENT, REGARDLESS OF THE TRUE DISEASE STATE.

As in Chapter 4, there are three disease states. These are defined as:

- **Minimal (M) -** bacteriologically negative disease, with radiological signs of disease in the lungs, not considered to be infectious[13]
- **Subclinical (S)** - bacteriologically positive disease, not reporting symptoms of active TB disease, considered to be infectious[9,13]
- **Clinical (C)** - bacteriologically positive disease, with symptoms of active TB disease, considered to be infectious[13]

Transitions exist in both directions between minimal and subclinical, and subclinical and clinical. There are no direct pathways between minimal and clinical without passing through subclinical. **Recovery (R)** is possible from minimal, and **Death (D)** from disease is possible from clinical. To maintain a fixed population size, the births into **Susceptible (X)** are equal to the deaths. There is also a baseline passive screening to **Treatment (T)** pathway for detecting people with symptoms of disease. People who have successfully completed treatment, given an average duration, are then considered **Recovered (R)**.

The state of **Infection (I)**, is defined as a state where there is evidence of *M.tb* infection, e.g. through an immunological test, without any evidence of disease, bacteriological, radiological, or clinical. As described in Chapter 6, there are two progressions to disease from infection, infection to **Minimal (M)**, and infection to **Subclinical (S)**. Recovery from **Infection (I),** before any progression to disease, returns to **Susceptible (X)**.

Progression to infection is achieved through two pathways: transmission to the **Susceptible (X)** population, or transmission to the **Recovered (R)** population. The number of people infected at each time point is guided by the proportion of the population with Subclinical and Clinical disease, alongside a relative transmissibility of Subclinical disease and a single parameter to represent the product of the number of contacts per unit time and the transmission probability for each contact. For transmission to recovered, there is an additional parameter that represents the protection from infection conveyed by a previous infection, disease, and/or treatment.

The differential equations that describe this system are written as:

$$
\begin{aligned}
\dot{X} &= -\beta X\frac{(tS+C)}{N} + \gamma_i I + \eta C + mN - mX \\
\dot{I} &= -(\gamma_i + \alpha_m + \alpha_s)I + \beta\frac{(tS+C)}{N}\big(X + R(1-\chi_r)\big) - mI \\
\dot{M} &= -\big(\theta_p + \gamma_m\big)M + \alpha_m I + \theta_r S - mM \\
\dot{S} &= -\big(\theta_r + \sigma_p\big)S + \sigma_r C + \theta_p M - mS \\
\dot{C} &= -(\sigma_r + \nu + \eta)C + \sigma_p S - mC \\
\dot{D} &= +\eta C \\
\dot{T} &= -\epsilon T + \nu C - mT \\
\dot{R} &= -\beta\frac{tS+C}{N}(1-\chi_r)R + \gamma_m M + \epsilon T - mR
\end{aligned}
$$

The model was written in R, using these equations with the *deSolve* package.[23,24]

For the purposes of implementing population-wide screening interventions, the parameter sets needed to be such that the steady state infectious prevalence (sum of subclinical and clinical prevalences) was similar to, or greater than, the recommended disease prevalence for population-wide screening of 500 per 100,000.[6,16] As such, the population size in the model was fixed to 100,000. In order to find parameter sets that would reach this steady state with a fixed level of background screening, I used a history matching and emulation package in R, *hmer*.[25] Priors for each parameter used in Chapters 4 and 6 were the posterior ranges presented in those respective chapters. The average duration of treatment, ϵ was fixed at one year, and the life expectancy, $\frac{1}{m}$ was fixed at 80 years. The priors for the remaining parameters were:

- β: 0 – 12
- t: 0.62 - 6.19[22]
- χ_r: 0 – 1

Where parameters describe rates, these presented per year. The parameter priors and posteriors can be found in the results section and Table 5.2.

5.2.2 Screening algorithm and implementation

The purpose of this analysis was to investigate the results of multiple rounds of population-wide screening. This was implemented as 3 rounds of screening over five years, with years 1, 3, and 5 working through screening the population, and years 2, and 4 with no additional screening beyond the baseline passive case detection. Over the course of each screening year, one twelfth of the population were screened each month. To implement this, two parallel equation sets were used, one set for unscreened and one for screened. Using the *deSolve* event functionality, each month one twelfth of the population was move from the unscreened population to the screened population. This ensured no-one was screened twice. As the

unscreened population was decreasing at each time step, the proportion of the unscreened population screened increased (at time step 1, $\frac{1}{12}$ of the unscreened population were screened, at time step 2, $\frac{1}{11}$ of the unscreened population were screened, etc.). Screening outcome was based on the performance of two tests, the chosen screening test for the intervention (see next section) and an Xpert MTB/RIF confirmation. For each disease state, P(positive screen) × P(positive confirm) determined the proportion of people screened to receive treatment. As with treatment from clinical disease in Figure 5.1, these people were moved directly to the Treated (screened) state in the model from their respective disease state. All others were moved to the screened version of their own state. People already on treatment were not screened separately, just moved to the screened population at the same $\frac{1}{12}$ per month as all other states.

5.2.3 Screening tests

Beyond the baseline passive symptom screening of clinical disease, the screening tests compared were:

- Xpert MTB/RIF (alone as a screening test, and following all other screening tests)
- Digital radiography (X-ray)
- Symptom screening
- Blood biomarkers (CORTIS)
- C-reactive protein (CRP)
- 2 Theoretical tests

It is unlikely that these all perform the same across different states of disease but, in the absence of any other data, I have assumed that performance outside the three disease states (minimal, subclinical, and clinical) is identical. Where possible, I have found data to inform the performance of tests within each disease state. Table 5.1 shows the probability of returning a positive test at each disease state used in this analysis.

TABLE 5.1: THE PROBABILITY OF RETURNING A POSITIVE RESULT FROM EACH TEST GIVEN THE DISEASE STATE

Disease state	Xpert	Symptom	X-ray	CORTIS	CRP	Theoretical (all)	Theoretical (infectious)	CORTIS (26)
Susceptible (X)	0.01	0.19	0.17	0.1	0.10	0.30	0.30	0.25
Infected (I)	0.01	0.19	0.17	0.1	0.10	0.30	0.30	0.25
Minimal (M)	0.01	0.28	0.40	0.1	0.17	0.70	0.30	0.25
Subclinical (S)	0.69	0.15	0.95	0.3	0.27	0.90	0.70	0.70
Clinical (C)	0.91	0.85	0.95	0.9	0.44	0.99	0.99	0.99

Recovered (R)	0.01	0.19	0.25	0.1	0.13	0.50	0.30	0.25

The performance of Xpert is defined in bacteriologically positive disease against liquid culture as a gold standard, however it has different performances for individuals with "presumptive" TB and in general population screening.[16,26] As "presumptive" TB is designated for people with symptoms of disease, I have used this sensitivity for the probability of a positive result for clinical disease, and the general population sensitivity for the probability of a positive result for subclinical disease.[16,26] There are no reported sensitivities for Xpert in minimal disease, and as it is a state defined by being bacteriologically negative, I have assumed that the probability of a positive test in minimal disease is equal to the probability of a positive test for people without disease, in other words, the reported rate of a false positive, 1 − specificity.[16,26]

Some studies following the 2016 Kenyan prevalence survey provided information on the performance of CAD4TBv6 and presence of symptoms.[27,28] Scores of 55 or more from CAD4TB identified 95% of those with bacteriologically-confirmed pulmonary TB, so this has been used as the baseline performance of digital x-ray in TB disease.[27] This baseline has been applied to both subclinical and clinical disease.[27] The probability of a positive test for those without TB has been calculated by 1 − specificity from the same source.[27] These values match with the reported performance of chest radiography (any abnormality) set out by the WHO screening guidelines.[16] There are again no results for performance in minimal disease, and so I have assumed a drop in performance to approximately 40% of the sensitivity in clinical disease.

Symptoms for clinical disease are, by definition, present and a background rate of symptoms in those without TB has been inferred from another Kenyan prevalence survey post-analysis of people without TB.[28] The background rate found is analogous to other estimates of the frequency of respiratory symptoms.[29,30] Subclinical disease is defined as when an individual is not experiencing, not aware of, or not reporting symptoms, with clinical disease defined as when an individual is experiencing and reporting symptoms. However, this is all dependent on the symptom screen chosen, and for the purpose of this analysis, there is some overlap between subclinical and clinical disease on the chosen symptom screen where some people with typical clinical disease will not report symptoms on the screen, and some with typical subclinical disease will report symptoms on the screen. To account for this overlap between the two groups, I have reduced the probability of a positive result from clinical disease from 1 by as much as I increased the probability of a positive result from subclinical disease from 0. This difference is less than, but close to, the background level of symptoms experienced in the general population. There is evidence that people with minimal disease sometimes report experiencing symptoms.[13] The rate of symptoms in those with minimal is also estimated from

the Kenyan prevalence survey post-analysis, where those radiographic results with abnormalities suggestive of TB reported symptoms slightly more often than those without abnormalities suggestive of TB.[28]

The blood biomarker, RISK11 from the CORTIS trial, designed as a predictor for progression to disease, was also tested on people with prevalent disease, both symptomatic and asymptomatic.[31] For a given specificity, the sensitivity for both subclinical and clinical disease could be established from published ROC curves.[31] The sensitivity from the ROC curve gives the respective probabilities of a positive result in subclinical and clinical disease, and the specificity is assumed to give the probability of a negative result in those without disease (i.e. the probability of a positive result is 1 − specificity). As there is no explicit testing for minimal disease, I have estimated the probability of a positive result from RISK 11 in minimal disease based on the expected proportion of positive tests seen in those without TB.[31] Both the main analysis threshold of 60 and the sensitivity analysis threshold of 26 have been used as potential screening tests here, with the first considered an existing text, and the latter a theoretical test.

Most CRP reporting comes from HIV-positive cohorts.[32–35] A cross-sectional survey within communities in Kampala, Uganda, measured CRP levels alongside HIV and Xpert Ultra testing.[36] With an HIV prevalence of 10 -12% in those with TB, these CRP values are more representative for an HIV-negative model than one with 100% HIV prevalence.[36] Using a 10mg/dL cut-off, I have taken the probability of a positive result for clinical from the CRP level from all those with a positive Xpert results (more than trace), the probability of a positive result for minimal from all those with a trace positive Xpert, and the probability of a positive result for subclinical from the study average for both trace and more than trace.[36] The probability of a positive result for people without disease comes from the proportion of those with negative Xpert results who had a CRP reading of greater than 10mg/dL.[36]

The theoretical tests are based on the WHO Target Product Profiles (TPPs) for systematic screening of TB disease.[16] These profiles do not distinguish between disease as I have, so I have assumed that they perform to profile specification in clinical disease, and then created two profiles with reductions for subclinical and minimal disease with higher performances in both when the theoretical tool is able to detect all disease, compared to on that just detects infectious disease.

5.3 Results

5.3.1 Parameterisation

Initial rounds of *hmer* found that almost the entire prior parameter space was non-implausible,

as can be seen in Table 5.2 with the priors and the final parameter choices. However, not all parameter selections led to a steady state of 500 per 100,000 with random selections from the entire initial parameter space leading to steady state prevalences of anything from 0 to 6,000 per 100,000. Running *hmer* through multiple waves restricted the parameter space slightly and then the final parameter set was selected using Latin Hypercube Sampling across that restricted space to generate 10,000 parameter sets. These parameter sets were then reduced to parameter sets to sets that resulted in a steady state of between 475 and 525 per 100,000 in the interval of 4,000 to 4,500 months. This produced 1187 parameter sets to test the model system on. Figure 5.2 shows the distribution of these parameters within the non-implausible space, and any between parameter correlations.

TABLE 5.2: THE PARAMETERS IN THE MODEL, WITH THEIR PARAMETER NAMES, A PARAMETER DESCRIPTION, A PRIOR RANGE, AND A NON-IMPLAUSIBLE RANGE (POSTERIOR RANGE)

	Parameter description	Prior range	Non-implausible range
beta	contact parameter	0 - 12	6.24(3.41 - 10.04)
t	relative transmission from subclinical	0.62 - 6.18	3.44(1.68 - 5.58)
gamma_i	recovery from infection	0.60 - 1.79	1.16(0.68 - 1.69)
gamma_m	recovery from minimal	0.14 - 0.23	0.19(0.15 - 0.23)
alpha_m	progression from infection to minimal	0.03 - 0.15	0.08(0.04 - 0.14)
alpha_s	progression from infection to subclinical	0.01 - 0.06	0.04(0.02 - 0.06)
theta_p	progression from minimal to subclinical	0.20 - 0.28	0.24(0.2 - 0.28)
theta_r	regression from subclinical to minimal	1.21 - 2.02	1.61(1.26 - 1.98)
sigma_p	progression from subclinical to clinical	0.57 - 0.94	0.76(0.58 - 0.93)
sigma_r	regression from clinical to subclinical	0.46 - 0.73	0.59(0.47 - 0.72)
chi_r	protection from re-infection	0 - 1	0.68(0.13 - 0.93)
nu	baseline treatment from clinical disease	0.5 - 1	0.73(0.52 - 0.97)
eta	death from clinical disease	0.28 - 0.38	0.33(0.28 - 0.38)

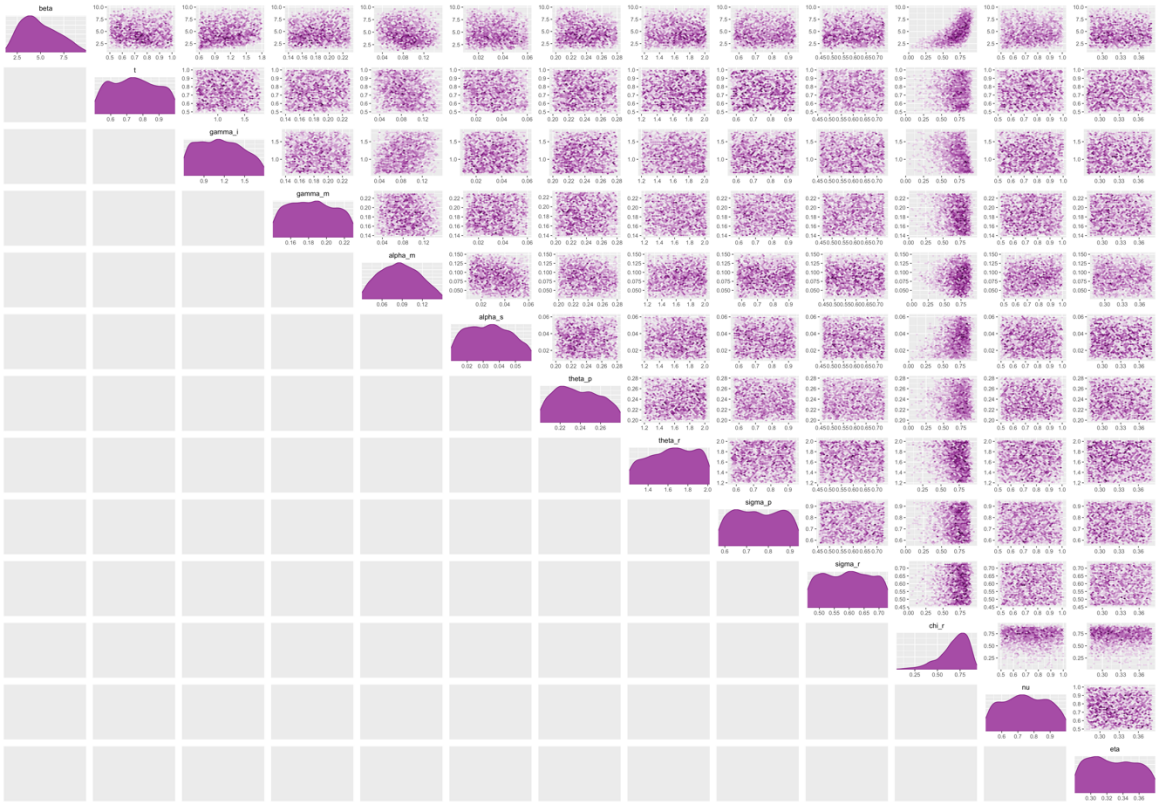


FIGURE 5.2: THE POSTERIOR RANGE OF EACH PARAMETER AND THE CORRELATIONS BETWEEN EACH PAIR OF PARAMETERS

5.3.2 Reduction in prevalence

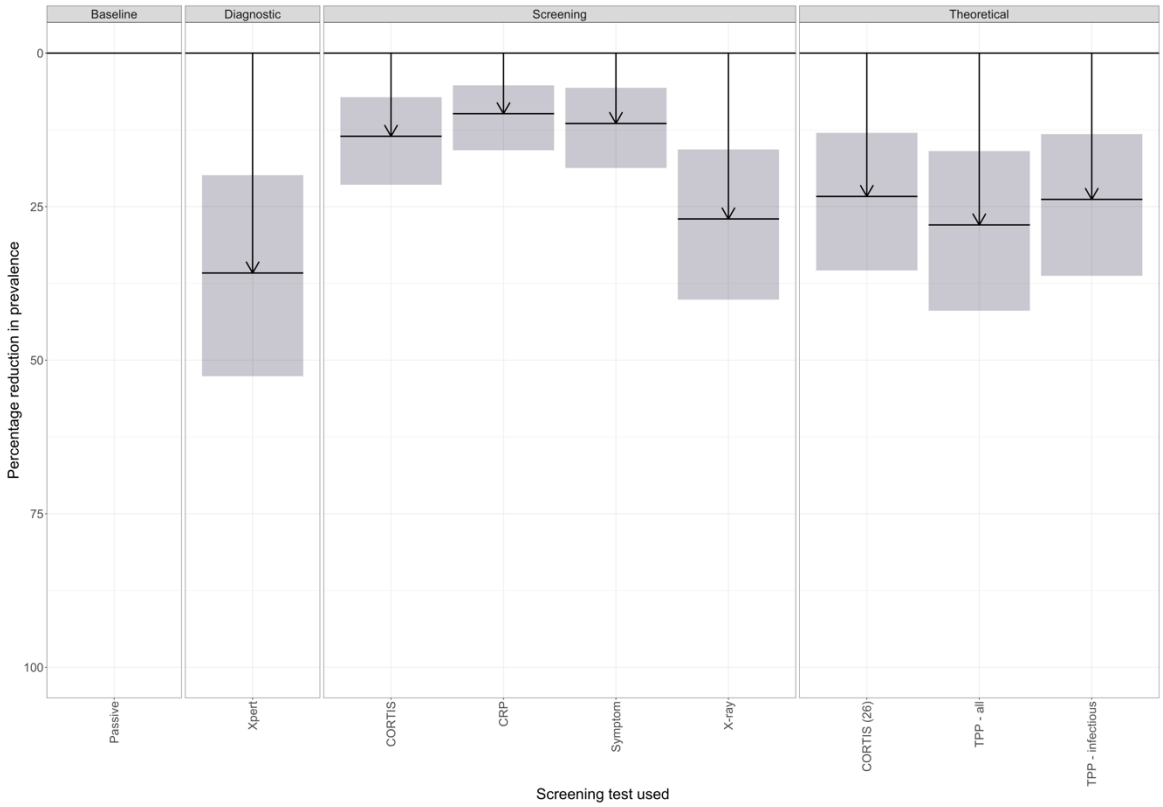


FIGURE 5.3: THE REDUCTION IN INFECTIOUS DISEASE PREVALENCE A YEAR AFTER THE END OF THE THIRD SCREENING INTERVENTION, FROM EACH SCREENING OPTION

Figure 5.3 shows the reduction in prevalence achieved using each screening test. From the baseline of passive case finding, the greatest reduction in infectious disease prevalence was seen when using Xpert testing for all (a reduction of 36% [20 - 53]). Of the tests that are currently available, x-ray screening followed by an Xpert confirmation performed almost as well as Xpert alone (27% [16 - 40]).

The three theoretical tests performed similarly at reducing infectious disease prevalence. The alternative threshold for the CORTIS biomarker reduced the prevalence by 23% [13 - 35], which is similar to the reduction from the TPP inspired infectious test (24% [13 - 36]). The TPP for all disease, with a higher sensitivity for minimal disease, reduced the most of the three theoretical tests, with a reduction between that of x-ray and Xpert (28% [16 - 42]).

These infectious disease prevalences were all measured a year after the end of the final intervention, but required sustained interventions to prevent a return to higher infectious disease prevalences.

5.3.3 False positive diagnoses

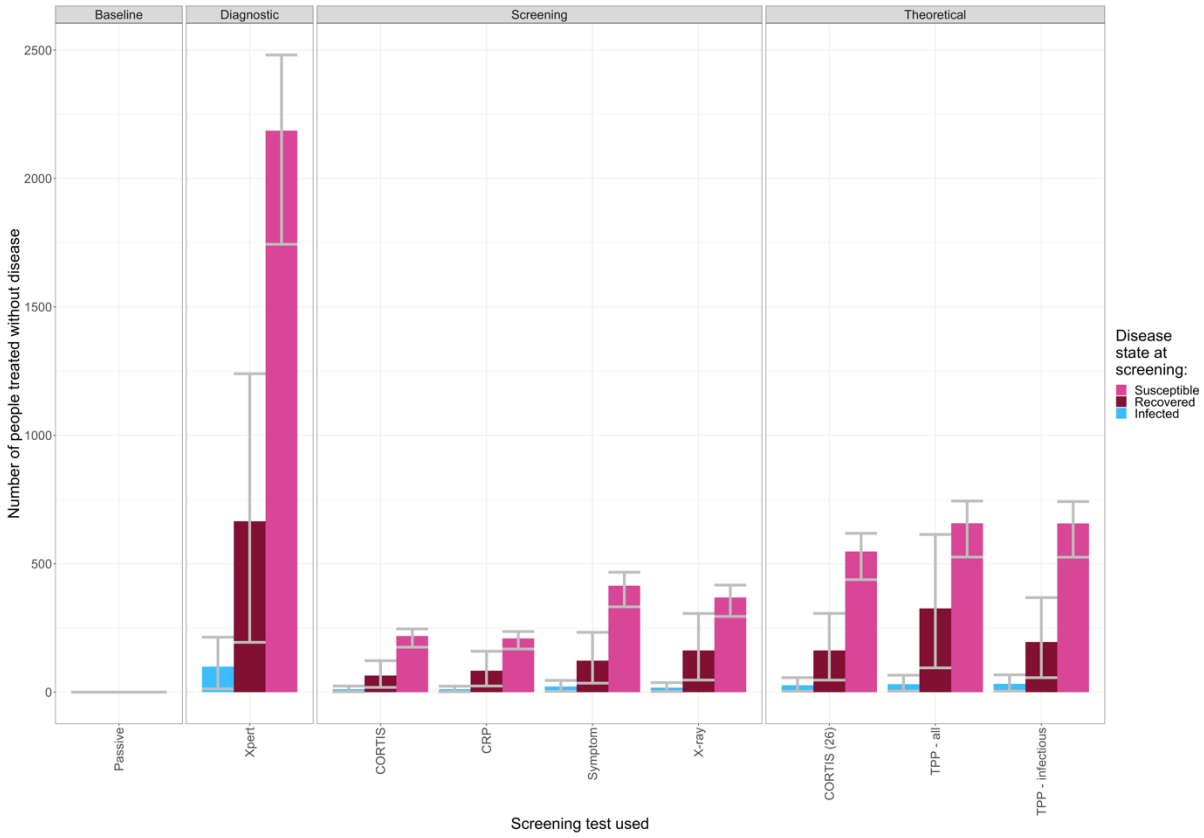


FIGURE 5.4: THE NUMBER OF PEOPLE FALSELY TREATED FOR TB OVER THE COURSE OF THREE SCREENING INTERVENTIONS, SPLIT BY THE NON-DISEASE STATE IN WHICH THEY WERE SCREENED.

The number of people without disease who would receive treatment can be split into three groups: people who have recovered from disease, either with or without treatment; those who

are infected without disease; and those who are susceptible to infection. As the largest group by size, the greatest number of false diagnoses occur in people susceptible to infection. For screening with CRP, X-ray, and the all-disease TPP the confidence intervals of those treated whilst susceptible to infection overlaps with the number treated of those who have recovered from disease. Figure 5.4 shows how the number of people without disease who would be treated varies across each screening test.

Against a population of 100,000 with an infectious disease prevalence of approximately 0.5%, screening the entire population with Xpert could lead to 2186 [1744 - 2481] people who are susceptible to infection receiving treatment, alongside 99 [13 - 214] people with a TB infection and 665 [194 - 1240] people who had previously recovered from disease. Screening with CRP, followed by an Xpert confirmatory test mistreats the fewest number of people with 209 [168 - 236] people who are susceptible to infection receiving treatment, 11 [2 - 23] people with a TB infection receiving treatment and 84 [24 - 159] people who had previously recovered from disease receiving treatment.

5.3.4 False negative diagnoses

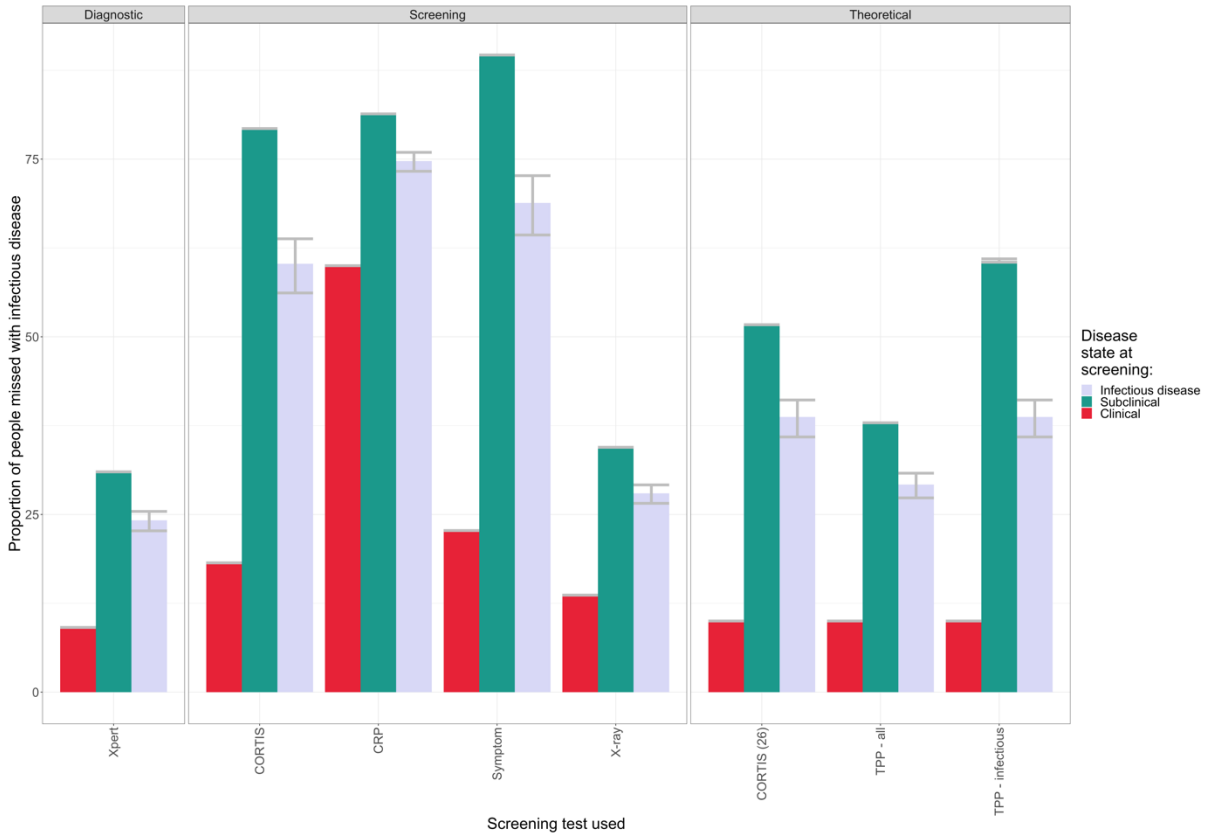


FIGURE 5.5: THE PROPORTION OF PEOPLE IN EACH INFECTIOUS DISEASE STATE INCORRECTLY SCREENED AS NOT HAVING DISEASE

Xpert is designed to detect bacteriologically positive disease, so when considering disease missed by screening, I am only considering subclinical and clinical disease missed by the screening algorithm, (i.e. infectious disease only). A false negative can be achieved through two mechanisms, people with clinical or subclinical disease who either test negative at the screening test, or who return a positive screening test followed by a negative confirmatory test. Figure 5.5 shows the percentage of people with infectious disease missed for treatment by the screening and confirmatory test combination.

Xpert screening alone misses the fewest people with infectious disease at the time of screening (24% [23 - 25]). CRP followed by Xpert screening misses the most overall (75% [73 - 76]).

5.3.5 Number needed to treat

TABLE 5.3: THE NUMBER OF PEOPLE WHO RECEIVED TREATMENT FOR EACH PERSON WITH INFECTIOUS DISEASE CORRECTLY IDENTIFIED

Screening				**Diagnostic**	**Theoretical**		
Symptom	X-ray	CORTIS	CRP	Xpert	TPP - all	TPP - infectious	CORTIS (26)
151 (89-933)	71 (41-490)	63 (38-394)	99 (59-661)	366 (216-2431)	133 (74-953)	127 (77-837)	106 (64-696)

I calculated the number positive confirmatory tests, which equates to the number of people treated, per true diagnosis of infectious disease. Table 5.3 shows this for each screening test. When everyone is tested with Xpert, with no prior screening test, the number needed to treat is the highest. For the remainder of the screening tests, CORTIS leads to the fewest treated (63 [38 – 394] per true positive diagnosis), whereas symptom screening requires the highest number of confirmatory tests (151 [89 -933]).

5.4 Discussion

This work presents a framework for analysing the impact that screening tests could have in a population. Of particular importance in this framework is the representation of the spectrum of disease. Test performance within the spectrum of disease can vary; symptom screening works well for clinical disease but poorly for subclinical disease, and Xpert performs better in smear-positive disease than smear-negative disease.[14,16,26,36,37] Therefore, a framework that enables these variations to be incorporated in finer detail has the potential to better predict the impact of different tests. Including the minimal disease state introduces a reservoir of disease that is

not adequately detected by the screening or confirmatory tests compared here and helps understand the impact of missing these people with early disease.

Due to the uncertainty in the test performances, the output from this model, with current inputs cannot be conclusive. The broad results will likely hold true; screening an entire population with Xpert and following up with complete treatment for all who test positive will have a large impact in reducing prevalence. In practice, this has already been recorded within the ACT3 trial.[18] It is also likely that, of the currently available tests, mass radiography will have the highest impact. However, beyond that, the uncertainty in the point values for test performance prevents any stronger conclusions being drawn from the modelling results.

Sensitivity and specificity are both phrases that are widely used to discuss the performance of tests, however, this assumes a dichotomy of disease or not disease. My work has focussed on understanding how disease is not one uniform state, and that then renders a dichotomous definition of performance incompatible with the model structure. With multiple disease states, a single sensitivity and specificity value is difficult to interpret because it is against a single disease state. For example, a sensitivity defined for clinical disease, would not necessarily count subclinical disease as a true positive. Instead, a sensitivity for each disease state is required, and then a specificity for non-disease. I have referred to these instead as the probability of returning a positive test in each disease (or non-disease) state. While the current focus of TB interventions remains on interrupting transmission, my model can quantify the potential benefit of preventing transmission through inadvertent treatment of minimal disease. I have seen that a theoretical test that detects all disease states would reduce prevalence and transmission by more than one that detects only infectious disease states, by preventing more people with minimal disease progressing to subclinical and clinical disease. If these positive diagnoses from people with minimal disease were included as true, the performance metrics, in particular the number of confirmatory tests needed for a true positive result, would improve. However, in order to keep the analysis in line with the focus on infectious TB disease, I excluded minimal disease from the analysis of true positive diagnoses.

The strengths of this study lie in the comparability and flexibility of the analyses. For a fixed population, different screening test profiles can be analysed within an otherwise identical simulation to find a definitive and objective impact of each test. These impacts have been measured as three intensive screening interventions, in the space of five years. A future extension of this work could be into a tool that can be used by anyone to quickly find the potential impact of the combination of all proposed screening and confirmatory tools along with frequency of testing to highlight the optimal interventions for a given setting.

The main weakness to this study is the data on test performances. These are reported in line with the infection/disease binary model of TB, and so how tests perform with different stages of disease are not measured. Performance of Xpert in those with "presumptive" TB compared to a general population is very different (0.91 vs 0.69). This is likely a matter of thresholds; people with "presumptive" TB are unwell, and have higher levels of *M.tb* in their sputum, whereas there are people in the general population who would test positive on a gold standard test but with lower levels of *M.tb* they are less likely to be detected by Xpert. Thus, sensitivities are lower in populations with lower thresholds of disease, despite the test still performing the same. A single, spontaneous sputum culture is accepted as the gold standard for clinical and subclinical disease, however, there is no gold standard for minimal disease. Therefore, there is no standard against which to test the performance of screening tests in minimal disease. In order to better predict the true impact of an intervention, a clear understanding of the performance of that intervention in each disease state is required.

There are numerous options for extending this work. To align with different locations and TB burdens, the baseline infectious disease prevalence could be changed at the start. Within the screening algorithm itself, the rate of screening could be increased (or decreased) and the frequency similarly. There have been screening interventions that reach entire populations in weeks or months, and so reducing the time taken to screen the entire population could feasibly be reduced to 6 months, or less. Similarly, it would be possible to increase the frequency of the testing or introduce further screenings after the third. Both the choice of frequency and duration aligned this work with the ACT3 trial, to compare with real world results, but as a predictor for the efficacy of screening tests, variability in both would be useful.[18] Different testing algorithms could also be considered, for example having a two-test process, where a positive result from either screening test resulted in a confirmatory test. These two tests could be optimised for different points on the spectrum of disease to maximise the the number of people with TB disease referred for further testing. However, without a confirmatory test that can detect minimal disease, the impact of screening for minimal disease would be limited. Therefore, the algorithm could be updated to remove the screening test or alternatively theoretical performances for confirmatory tests could be used. This model could be used to help optimise any future performance recommendations.

I have compared these tests by the reduction in prevalence that they confer (through the 3 screens in five years). An alternative metric could be the reduction in incidence over time. However, to use this would require a solidification of definitions. Currently incidence is defined as progression to bacteriology positive disease. However, with undulation between minimal and subclinical disease possible within this model, it would be possible to progress to subclinical

disease multiple times. A more reasonable definition for this case may then be a first progression to subclinical disease, however a re-progression from an extended period with minimal disease may need to be considered as a new progression. A second consideration would be the first progression into any disease (then including minimal). With a consistent measurement, any reduction should be easy to compare over time, however it may not pair well with quoted estimates of incidence in real-world settings.

It was already known that population-wide screening can have a positive impact on the burden of TB, and this work has not changed that.[18] I have created a framework to predict the size of this impact for different screening tests, in different populations. However, in order for this, or any framework to produce accurate predictions, an understanding of the performance of screening tests across the spectrum is required. In turn, this will require the development of a gold standard test for minimal disease.

5.5 References

1. Comstock, G. W. Tuberculosis studies in Muscogee County, Georgia; community-wide tuberculosis research. *Public Health Reports (Washington, D.C.: 1896)* **64**, 259–263 (1949).

2. National Tuberculosis Institute, B. Tuberculosis in a rural population of South India: a five-year epidemiological study. *Bulletin of the World Health Organization* **51**, 473–88 (1974).

3. Dye, C., Garnett, G. P., Sleeman, K. & Williams, B. G. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *The Lancet* **352**, 1886–1891 (1998).

4. World Health Organization Communicable Diseases. *What is DOTS? : A guide to understanding the WHO-recommended TB control strategy known as DOTS*. WHO/CDS/CPC/TB/99.270 (1999).

5. Banerji, D. & Andersen, S. A Sociological Study of Awareness of Symptoms among Persons with Pulmonary Tuberculosis. *Bulletin of the World Health Organization* **29**, 665–683 (1963).

6. World Health Organization. *Global tuberculosis report 2021*. (World Health Organization, 2021).

7. World Health Organization. *WHO expert committee on tuberculosis - eighth report*. https://apps.who.int/iris/bitstream/handle/10665/40606/WHO_TRS_290.pdf?sequence=1&isAllowed=y (1964).

8. World Health Organization. *WHO expert committee on tuberculosis - ninth report*. https://apps.who.int/iris/bitstream/handle/10665/41095/WHO_TRS_552_eng.pdf?sequence=1&isAllowed=y (1974).

9. Frascella, B. *et al.* Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. *Clinical Infectious Diseases* **73**, e830–e841 (2021).

10. Onozaki, I. *et al.* National tuberculosis prevalence surveys in asia, 1990-2012: An overview of results and lessons learned. *Trop Med Int Health* **20**, 1128–1145 (2015).

11. 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* **18**, e228–e238 (2018).

12. Lin, P. L. & Flynn, J. L. The End of the Binary Era: Revisiting the Spectrum of Tuberculosis. *The Journal of Immunology* **201**, 2541–2548 (2018).

13. Richards, A. S. *et al. The natural history of TB disease-a synthesis of data to quantify progression and regression across the spectrum*. 2021.09.13.21263499 https://www.medrxiv.org/content/10.1101/2021.09.13.21263499v1 (2021).

14. Pai, M. *et al.* Tuberculosis. *Nat Rev Dis Primers* **2**, 16076 (2016).

15. Esmail, H., Macpherson, L., Coussens, A. K. & Houben, R. M. G. J. Mind the gap managing tuberculosis across the disease spectrum. *eBioMedicine* **0**, (2022).

16. World Health Organization. *WHO consolidated guidelines on tuberculosis Module 2: Screening Systematic screening for tuberculosis disease*. (World Health Organization, 2021).

17. Burke, R. M. *et al.* Community-based active case-finding interventions for tuberculosis: A systematic review. *The Lancet Public Health* **6**, e283–e299 (2021).

18. Marks, G. B. *et al.* Community-wide Screening for Tuberculosis in a High-Prevalence Setting. *The New England Journal of Medicine* **381**, 1347–1357 (2019).

19. Kranzer, K. *et al.* The benefits to communities and individuals of screening for active tuberculosis disease: a systematic review. *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease* **17**, 432–446 (2013).

20. Houben, R. M. G. J. *et al.* Spotting the old foe — revisiting the case definition for TB. *The Lancet Respiratory Medicine* **7**, 199–201 (2019).

21. Horton, K. C., Richards, A. S., Emery, J. C., Esmail, H. & Houben, R. M. G. J. Re-evaluating progression and pathways following Mycobacteria tuberculosis infection within the spectrum of tuberculosis disease. doi:10.1101/2022.12.14.22283422.

22. Emery, J. C. *et al.* Estimating the contribution of subclinical tuberculosis disease to transmission an individual patient data analysis from prevalence surveys. doi:10.1101/2022.06.09.22276188.

23. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2020).

24. Karline Soetaert, Thomas Petzoldt & R. Woodrow Setzer. Solving differential equations in R: Package deSolve. *Journal of Statistical Software* **33**, 1–25 (2010).

25. Iskauskas, A. *Hmer: History matching and emulation package*. (2022).

26. Zifodya, J. S. *et al.* Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. *The Cochrane Database of Systematic Reviews* **2**, CD009593 (2021).

27. Mungai, B. *et al. Accuracy of computer-aided chest X-ray screening in the Kenya National Tuberculosis Prevalence Survey*. http://medrxiv.org/lookup/doi/10.1101/2021.10.21.21265321 (2021) doi:10.1101/2021.10.21.21265321.

28. Mungai, B. N. *et al.* 'If not TB, what could it be?' Chest X-ray findings from the 2016 Kenya Tuberculosis Prevalence Survey. *Thorax* **76**, 607–614 (2021).

29. Esmail, H., Dodd, P. J. & Houben, R. M. G. J. Tuberculosis transmission during the subclinical period: Could unrelated cough play a part? *The Lancet Respiratory Medicine* **6**, 244–246 (2018).

30. Monto, A. S. & Sullivan, K. M. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiology and Infection* **110**, 145–160 (1993).

31. Scriba, T. J. *et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): A randomised controlled trial. *The Lancet Infectious Diseases* **0**, (2021).

32. Yoon, C., Davis, J. L. & Cattamanchi, A. C-reactive protein and tuberculosis screening: a new trick for an old dog? *Int J Tuberc Lung Dis* **17**, 1656 (2013).

33. Yoon, C. *et al.* Diagnostic accuracy of C-reactive protein for active pulmonary tuberculosis: a meta-analysis. *The International Journal of Tuberculosis and Lung Disease* **21**, 1013–1019 (2017).

34. Yoon, C. *et al.* Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *The Lancet Infectious Diseases* **17**, 1285–1292 (2017).

35. Yoon, C. *et al.* Yield and efficiency of novel intensified tuberculosis case-finding algorithms for people living with HIV. *American Journal of Respiratory and Critical Care Medicine* **199**, 643–650 (2018).

36. Kendall, E. A. *et al.* The spectrum of tuberculosis disease in an urban ugandan community and its health facilities. *Clinical Infectious Diseases* **72**, e1035–e1043 (2021).

37. Leeflang, M. M. G., Rutjes, A. W. S., Reitsma, J. B., Hooft, L. & Bossuyt, P. M. M. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ : Canadian Medical Association Journal* **185**, E537–E544 (2013).

Chapter 6 Extending the natural history beyond disease

6.1 Background

In Chapter 4, I introduced a new model structure and parameterisation that discretises the spectrum of TB disease, previously widely referred to as "active" disease only. As demonstrated in Chapter 5, this model structure alone can provide tools for understanding processes within disease, such as effects of screening, treatment, and onwards transmission potential. However, it does not explicitly cover progression of *Mtb* infection which means it is an incomplete model of the natural history of TB.

Expanding the model to include progression to disease from initial infection is important for other analyses or explorations of the effects of preventative therapies or for pre- or post-exposure vaccines. Parameterisations of infection to disease are widely used in modelling, however in these parameterisations disease is assumed to be a mostly homogeneous state, with a single threshold. My new model structure, from Chapter 4, presents multiple thresholds of disease, and so the progression from infection is likely to be different for progress to each threshold. Progression from infection to minimal disease is different to progression from infection to clinical disease, in terms of severity, and likely in terms of speed. Therefore, there is a need to investigate how best to represent progression from infection within my framework of the spectrum of disease.

The systematic review reported in Chapter 3 not only found data on progression through disease, but also from early infection through to disease. Although this was not used in the analysis of Chapter 4, it can be used to understand the full natural history of infection and disease.

6.1.1 Prior work

Katherine Horton has worked through the data from the systematic review to determine what, if any data collected was suitable for informing the progression from new infection to established disease. Three studies were deemed to provide sufficient data; works by Daniels in 1944, Madsen in 1942, and the National Tuberculosis Institute's study in Bangalore in 1974.[1–3].

As the data were collected at fixed intervals, adjustments were made to both the expected time frame of infection prior to discovery and the time from infection to disease. For time since TST conversion, we assumed that the true time of infection between the last time point (negative TST) and the new time point (positive TST) was uniformly distributed between the two reporting points. For the true time of progression from infection to disease, we used data from

Poulsen to inform a gamma distribution that likely described the rate of progression over the interval.[4] We aligned the gamma distribution to start at the time point sampled for TST conversion and then truncated it to sample the time of progression to disease within the time interval between the last negative disease screen and the first positive disease screen.

Together with Jon Emery and Katherine Horton, I compared potential model structures for transitioning from initial infection to disease, as an extension of the three-state disease model introduced in Chapter 4. These potential model structures can be seen in Figure 6.1. Model A shows a structure with a single progression from infection to disease. This is unlike many TB models where there are two progressions from infection to disease. Model B adds an intermediary state between infection and disease, which could be similar to the slow latent compartment included in many models however, like model A, only has one progression to disease, and this time is with an additional state to further slow progression. Model C combines these two models with progression to disease directly from infection to minimal, and also through the intermediary state, more similar to commonly used models. However, most models consider fast progression from infection to infectious disease, subclinical or clinical in my model structure, and this is what is added for models D and E. Model D adds progression from infection to subclinical disease to the model C structure, with three possible paths for progression from infection, whereas model E adds the progression to infection, but drops the intermediary state, leaving two paths for progression from infection: to minimal and to subclinical. While many models would use a slow latent state that this model structure doesn't, the minimal disease state has the potential to counteract the effect of losing the slow latent compartment by slowing the time for many from infection to subclinical disease.

With a cursory fit, just considering the progression to disease and comparing the fits visually, models D and E performed similarly. With fewer parameters, and insufficient data to properly inform the "intermediate" state, model E was chosen for the main analysis. Post-hoc analysis of the DIC for the full fit for each model was similar (680 — 690), with B performing the worst, A and C performing similarly, and D and E performing similarly. Much of this similarity is driven by the stability in the disease parameters and the differences are almost entirely due to the choice for progression to disease which is why a visual comparison could determine the same "best" model despite seemingly small differences in the DIC.

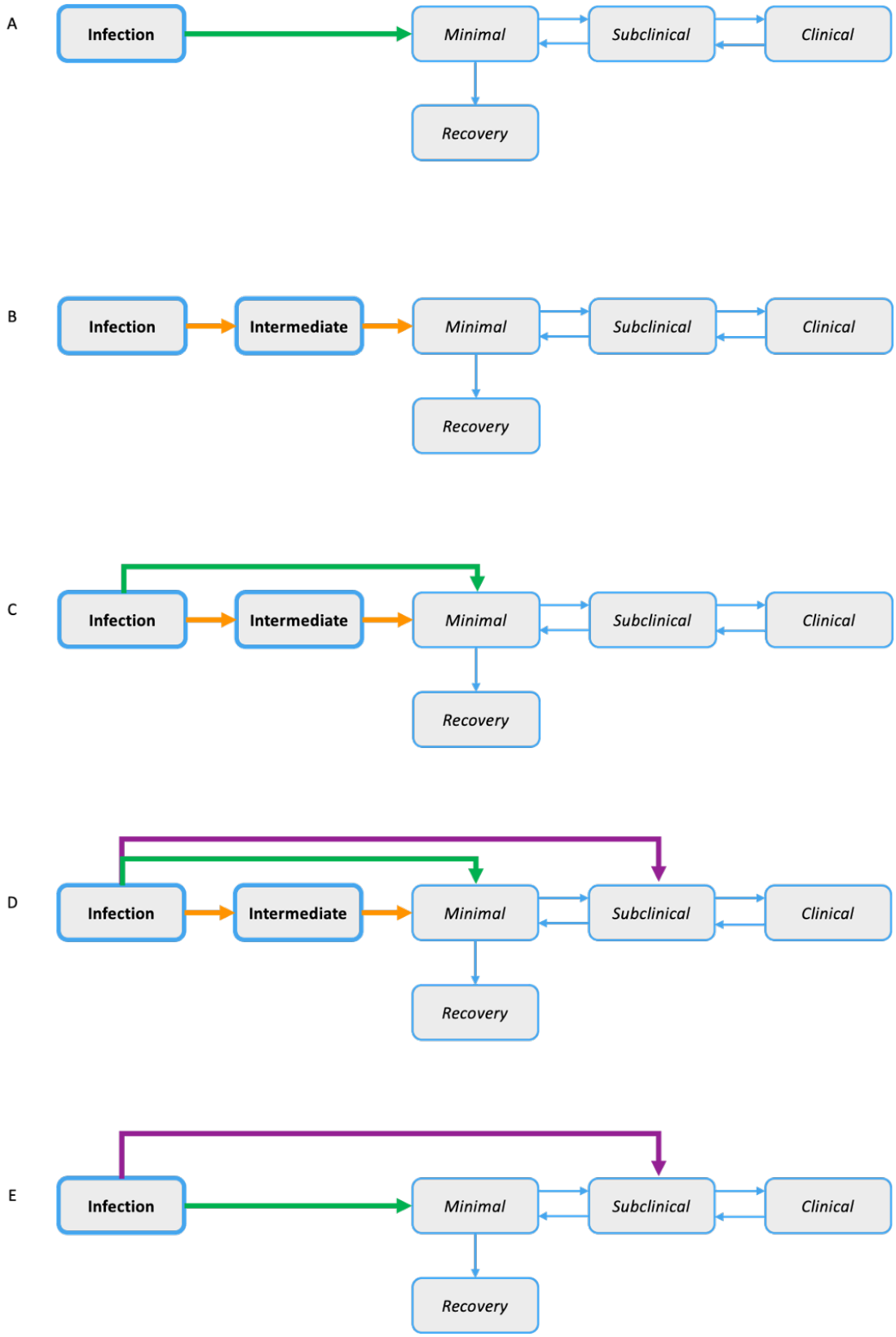


FIGURE 6.1: THE DIFFERENT PATHWAYS CONSIDERED FROM INFECTION TO DISEASE. THE SIMPLEST (A) PROGRESSES DIRECTLY FROM INFECTION TO MINIMAL DISEASE. B ADDS IN AN INTERMEDIARY STATE BETWEEN INFECTION AND MINIMAL DISEASE CREATING A DELAYED PROGRESSION. C COMBINES BOTH A AND B, WITH THE RAPID PROGRESSION TO MINIMAL DISEASE AND THE DELAYED PROGRESSION THROUGH THE INTERMEDIARY STATE. D IS AN EXTENSION ON C WITH A FURTHER RAPID PROGRESSION DIRECTLY TO SUBCLINICAL DISEASE. E DROPS THE INTERMEDIATE STATE AND JUST HAS TWO OPTIONS FOR PROGRESSION FROM INFECTION TO MINIMAL DISEASE AND SUBCLINICAL DISEASE.

6.1.2 My work

Here I present the work taken to join the two halves of the model and fit them into one coherent natural history model. The aim is to maintain consistency in the processing of data to ensure that, as in Chapter 4, all data are considered together simultaneously and in the same way. While the prior work sorted the data and performed a cursory fit of the infection to disease structures to determine a suitable progression structure to use, it neither formally calibrated the model to the data, nor used a coherent methodology with that used in Chapter 4 to be able to join the two halves of the model.

6.2 Methods

The chosen model structure, a five-state compartmental model as seen in Figure 6.1E, contains nine parameters. Of these nine parameters, there are four rates of movement between disease states, two rates of progression from infection, two recovery rate parameters, and one mortality rate parameter. The disease parameters are the same as those used in Chapter 4: minimal to subclinical, subclinical to minimal, subclinical to clinical, and clinical to subclinical. The two new parameters for progression from infection are infection to minimal and infection to subclinical. The two recovery parameters, one from minimal, one from infection, remove individuals from the model, requiring a new infection to re-enter. Finally, the TB death parameter, as in Chapter 4, is only from clinical disease. The structure is not stratified by any risk factors, assuming a homogeneous population. There are also no new infections or deaths due to anything other than TB.

Parameter values were estimated by fitting all the data to the model simultaneously within a Bayesian framework. Assuming no prior knowledge of the parameter values, every parameter other than TB mortality was given a uniform prior. Although for the fitting in Chapter 4, the prior range was given as between 0 and 12 per year, all of the parameters found were below 2.[5] Therefore, for all parameters the uniform prior was restricted between 0 and 3 per year, representative of a median duration of at least 4 months within each state. TB mortality was the exception, as there is an existing estimate for the rate from clinical disease based on historical data.[5–7] This was given a normally distributed prior with mean 0.39 and standard deviation 0.03.[7]

6.2.1 Extra fitting targets

The following section describes the process used to be able to fit three extra targets: the duration of disease, the ratio of subclinical to clinical disease, and the ratio of minimal to infectious disease. Each of these is fitted to a target estimated from available contemporary data as explained below. I have assumed that each of the ratios are observed in TB systems in

equilibrium and so, to speed the process of fitting these targets, rather than performing simulations until equilibrium, I have reduced the problem by deriving an equation for each target that relies only on the parameters. Each derivation is included below.

Duration of disease

As with Chapter 4 I have included fitting targets from outside the systematic review data to help ensure the results represented known qualities of TB disease. The first is that of the duration of disease, with two others on the ratios of subclinical to clinical and minimal to infectious at equilibrium.

The duration of disease estimate originates with data from the National Tuberculosis Institute that has been interpreted by Tiemersma et al.[3,6] Tiemersma et al use an assumption of exponential duration of disease to quote an average duration of disease as three years and calculate this from the number of incident cases occurring between each survey. They define a parameter δ, the inverse of disease duration, and state that the incident cases in an interval T, can be calculated as $\frac{1-exp(\delta T)}{\delta T}$.[6] Between two surveys, 1.5 years apart, a δ of 0.3 fits the cumulative distribution for the number of observed cases and thus $\frac{1}{\delta} = 3.33$ years is the average duration given by the data.[6] However, missed cases mean that is likely an over-estimate and so they state that 3 years is a good estimate for the average duration of disease.[6] This average, from how it has been calculated, is the mean, so, given the same distribution assumption, the median duration can be given by $\frac{ln(2)}{\delta}$. Taking $\delta = 0.33$ means that the mean duration is 3 years, giving a median duration of approximately 2 years.

These data come from incident cases between each survey, meaning that I have assumed that this value should be fitted against a simulated cohort of people starting with subclinical disease. To find this median duration of infectious disease, I then want to find that at two years, 50% of the cohort that started with subclinical disease, still has subclinical or clinical disease or has returned to subclinical or clinical disease having temporarily recovered to minimal disease. The time at which each iteration of the fitting process would reach the 50% point, can be calculated by finding the prevalence of infectious disease at two years and, from that, calculating the time at which the prevalence would have been 50%. An exponential function can be written as $p = p_0 e^{(rt)}$ where p is prevalence, t is time, and r is a constant. We know that at $t = 0$, $p = 1$ so $p_0 = 1$. If I simplify e^r to b, the equation simplifies to $p = b^t$. To find the value of the constant b, we set $p = 0.5$ and $t = t_{med}$ and find: $b = \left(\frac{1}{2}\right)^{\frac{1}{t_{med}}}$.

We can then substitute this back into the original equation to find the calculation for t_{med} given the prevalence, p at $t = 2$:

$$\begin{aligned} p &= 1\frac{2}{t_{med}} \\ ln(p) &= ln\left(\frac{1}{2}\right)\frac{2}{t_{med}} \\ t_{med} &= -2\frac{ln(2)}{ln(p)} \end{aligned}$$

This means, from fitting at a single time point, we can estimate the median duration of disease using the assumption of an exponential distribution of duration.

Prevalence Ratios

A recent review of prevalence surveys has found that people with subclinical disease make up about half of all the people with infectious disease.[8,9] There is not a similar review to inform the ratio between minimal and infectious. I have used results from a post analysis of the 2016 Kenyan prevalence survey to provide an estimate for the prior, that is used with the same assumptions as the subclinical to clinical ratio.[10] Whilst this cannot be perfect, it again provides a prior range for the calibration to consider. This was paired with a wide uncertainty for the model to consider. In the survey, 10% of those screened with CXR were considered to have abnormalities indicative of TB, but on expert review, only 60% were truly considered abnormal. On Xpert testing, 90% of all originally screened as TB were negative, with 10% positive (which would give a 9:1 ratio). However, taking into account that only 60% were considered TB on expert review, that brings the ratio down to 5:1. On discussion with co-authors and people who have been working in TB research for over 10 years, we felt that this number was still too high, so we halved the ratio (2.5:1) and used this as a midpoint for a wide prior. This lower estimate is also consistent with a repeated prevalence survey in Cambodia.[11]

I have assumed that these ratios would be observed for a system in a steady state. The steady state can be described by a set of differential equations, and from that an appropriate expression for each of the ratios at a steady state can be created.

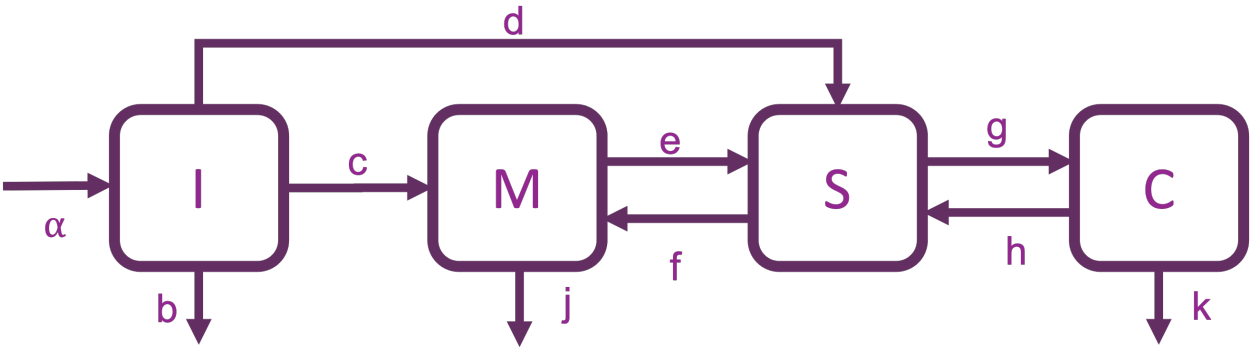


FIGURE 6.2: A DIAGRAM OF THE MODEL STRUCTURE, LABELLED FOR THE PURPOSE OF DERIVING THE EQUILIBRIUM EQUATIONS

With labelled transitions, the model structure can be seen in Figure 6.2 and the equations that describe this system are:

$$\begin{aligned}
\dot{I} &= \alpha - (c + d + b)I \\
\dot{M} &= cI - (e + j)M + fS \\
\dot{S} &= dI + eM - (g + f)S + hC \\
\dot{C} &= gS - (h + k)C
\end{aligned}$$

Subclinical to Clinical

The ratio of Subclinical to Clinical can be calculated from the equation from $\dot{C}$ where there is no change in C over time

$$\begin{aligned}
0 &= gS - (h + k)C \\
gS &= (h + k)C \\
\frac{S}{C} &= \frac{h + k}{g}
\end{aligned}$$

I have used this equation in the fitting with a normally distributed prior, with a mean of 1 and a standard deviation of 0.25.[5,9]

Minimal to Infectious

And then from the equation for $\dot{M}$:

$$\begin{aligned}
0 &= cI - (e + j)M + fS \\
cI &= (e + j)M - fS \\
I &= \frac{(e + j)M - fS}{c}
\end{aligned}$$

From the equation for $\dot{S}$:

$$
\begin{aligned}
0 &= dI + eM - (g+f)S + hC \\
dI &= (g+f)S - hC - eM \\
I &= \frac{(g+f)S - hC - eM}{d}
\end{aligned}
$$

Joining them together and writing C in terms of S:

$$
\begin{aligned}
\frac{(e+j)M - fS}{c} &= \frac{(g+f)S - hC - eM}{d} \\
d(e+j)M - dfS &= c(g+f)S - chC - ceM \\
d(e+j)M + ceM &= dfS + c(g+f)S - chC \\
(d(e+j) + ec)M &= dfS + c(g+f)S - \frac{hgc}{h+k}S \\
M &= \frac{\big(df + c(g+f)\big)(h+k) - hgc}{(h+k)(d(e+j)+ec)}S
\end{aligned}
$$

So then, for $\frac{M}{S+C}$:

$$
\begin{aligned}
\frac{M}{S+C} &= \frac{\frac{\big(df + c(g+f)\big)(h+k) - hgc}{(h+k)(d(e+j)+ec)}S}{\left(1 + \frac{g}{h+k}\right)S} \\
\frac{M}{S+C} &= \frac{\frac{\big(df + c(g+f)\big)(h+k) - hgc}{(h+k)(d(e+j)+ec)}}{\frac{h+k+g}{h+k}} \\
\frac{M}{S+C} &= \frac{\big(df + c(g+f)\big)(h+k) - hgc}{(h+k)(d(e+j)+ec)}\frac{h+k}{h+k+g} \\
\frac{M}{S+C} &= \frac{\big(df + c(g+f)\big)(h+k) - hgc}{(d(e+j)+ec))(h+k+g)}
\end{aligned}
$$

I have used this equation in the fitting with a normally distributed prior, with a mean of 2.5 and a standard deviation of 0.5.[5,9]

I can check that this matches with the disease only model by setting $d = 0$:

$$
\begin{aligned}
\frac{M}{S+C} &= \frac{(c)(g+f)(h+k) - hgc}{ec(h+k+g)} \\
\frac{M}{S+C} &= \frac{(g+f)(h+k) - hg}{e(h+k+g)}
\end{aligned}
$$

Which matches the equation from the disease only model seen in Chapter 4 (shown in Appendix B).

Extensions and reductions of these equations to match the alternative model structures in Figure 6.1 all reduce down, or scale up, appropriately when changing the parameters involved.

6.2.2 Likelihood weightings

The data presented in Chapter 4, alongside the data in Table 6.1 were joined together to create the full data set. The overall likelihood calculation is comprised of likelihoods for each data point, and the individual likelihoods for each data point use a binomial distribution which allow for weighting to account for the different cohort sizes. Therefore, data from a larger cohort has more importance in the fitting.

Some studies report on a single cohort multiple times. This is true of all three studies providing data on progression from infection. To prevent the repeated reporting outweighing single cohorts, the data are down-weighted by the number of repeats by artificially shrinking the cohort, whilst maintaining the proportion of people who have progressed at each time point.

6.2.3 Fitting process

The posterior values for the parameter were sampled using a sequential MCMC method. An initial burn-in of 10,000 runs was then adapted to find an optimal acceptance rate of between 25% and 35% by varying both shape and scale. At this desired acceptance rate, the final 150,000 iterations were kept and from this, the median parameters and 95% uncertainty intervals were reported.

6.2.4 Further analysis

To observe the future trajectories of infection based on the parameters found, I created a stochastic cohort model where the risk of progression was calculated independently for each individual at each monthly time step by selecting a random number from a uniform distribution between 0 and 1.

For a cohort of 1,000 people, all starting with infection, I ran 10,000 repeats of this model, following up for 120 months. At each repeat, a single value for each parameter was chosen from the respective posterior distribution, and each person used the same parameter set. I then summarised each of these repeats with the number of people in each state at the end of the simulation. Results from this simulation are presented as the median value of each state over all 10,000 repeats.

The Bayesian fit was performed in LibBi using RBi and rbi.helpers as an interface to R and RStudio which was used for all subsequent analyses.[12–14]

6.3 Results

The results for the posterior estimate of each parameter, in comparison to the provided prior, are shown in Table 6.1. The resulting trajectories to match the data are shown in Figure 6.3.

TABLE 6.1: THE MODEL PARAMETERS, THE SYMBOLS USED TO REPRESENT THEM IN FIGURE 6.2, THE PRIOR PROVIDED TO THE CALIBRATION, AND THE POSTERIOR ESTIMATE. THE FINAL COLUMN SHOWS THE POSTERIOR ESTIMATES FROM THE DISEASE-ONLY MODEL IN CHAPTER 4.

Parameter	Symbol	Prior (per year)	Posterior (95% CI)	Disease only model
Recovery from Infection	b	Uniform: 0 — 3	1.23 (0.69, 1.99)	–
Infection to Minimal	c	Uniform: 0 — 3	0.08 (0.04, 0.17)	–
Infection to Subclinical	d	Uniform: 0 — 3	0.03 (0.01, 0.07)	–
Recovery from Minimal	j	Uniform: 0 — 3	0.18 (0.14, 0.23)	0.2 (0.15, 0.25)
Minimal to Subclinical	e	Uniform: 0 — 3	0.24 (0.2, 0.28)	0.26 (0.22, 0.3)
Subclinical to Minimal	f	Uniform: 0 — 3	1.58 (1.22, 2.05)	1.51 (1.18, 1.95)
Subclinical to Clinical	g	Uniform: 0 — 3	0.72 (0.56, 0.95)	0.69 (0.54, 0.88)
Clinical to Subclinical	h	Uniform: 0 — 3	0.57 (0.46, 0.73)	0.58 (0.45, 0.72)
Mortality	k	Normal: mean = 0.389, sd = 0.028[7]	0.33 (0.28, 0.38)	0.32 (0.26, 0.37)
Subclinical to Clinical	$\frac{S}{C}$	Normal: mean = 1, sd = 0.25[9]	1.26 (0.96, 1.6)	1.3 (1.01, 1.65)
Minimal to Infectious	$\frac{M}{S+C}$	Normal: mean = 2.5, sd = 0.5[10]	3.39 (2.78, 4.04)	3.9 (3.36, 4.5)
Duration		Normal: mean = 2, sd = 0.5[6]	1.17 (1.05, 1.32)	0.99 (0.9, 1.12)

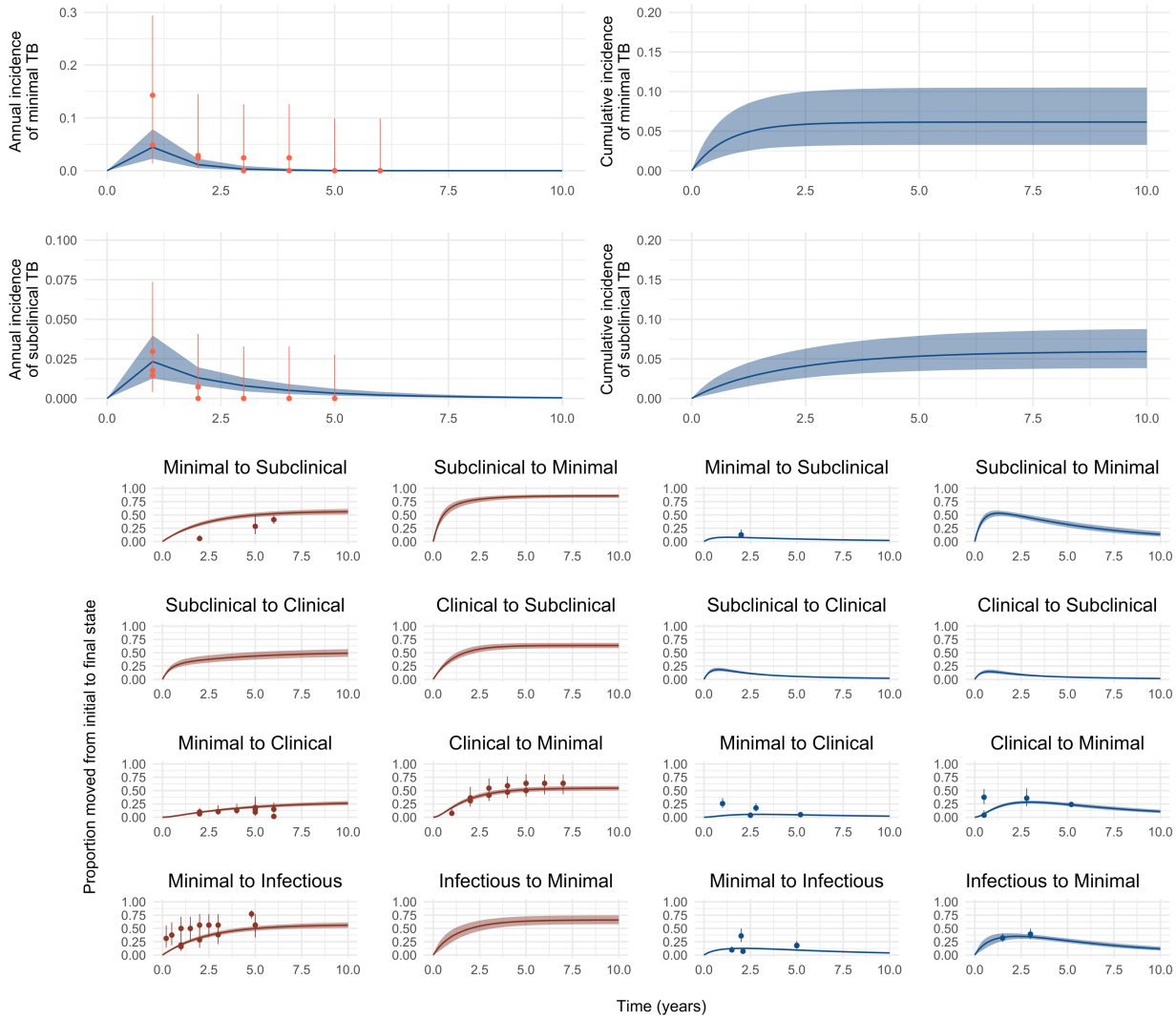


FIGURE 6.3: THE RESULTS OF THE FITTING AGAINST THE NEW DATA ON INFECTION TO DISEASE (INFECTION TO MINIMAL AND INFECTION TO SUBCLINICAL) AND THE ORIGINAL DATA FROM THE DISEASE ONLY FIT. THE TOP TWO GRAPHS SHOW THE INCIDENCE OF MINIMAL TB FROM INFECTION, THE LEFT SHOWS THE CHANGE IN ANNUAL INCIDENCE THAT WAS FITTED TO THE DATA FROM INITIAL INFECTION, AND THE RIGHT SHOWS THAT AS A CUMULATIVE INCIDENCE. THE SECOND ROW SHOWS THE SAME BUT FOR INCIDENCE INTO SUBCLINICAL DISEASE. THE BOTTOM FOUR ROWS SHOW THE CUMULATIVE INCIDENCE (RED) AND CROSS-SECTIONAL PREVALENCE (BLUE) FOR EACH OF THE EIGHT TRANSITIONS BETWEEN THE THREE DISEASE STATES.

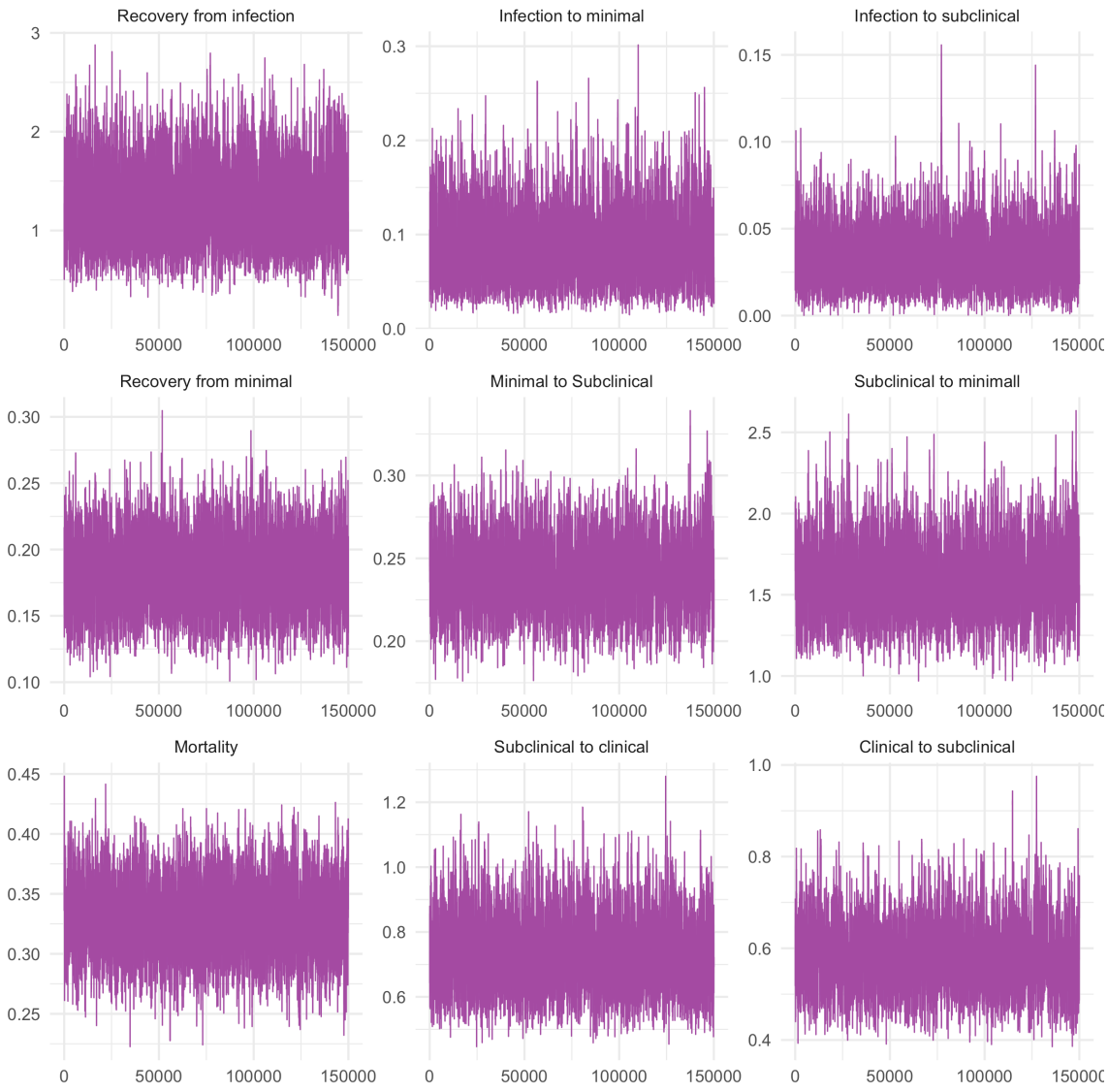


FIGURE 6.4: THE TRACE PLOTS SHOWING THE OUTCOME OF THE CALIBRATION FOR EACH OF THE PARAMETERS IN THE FULL INFECTION TO DISEASE MODEL

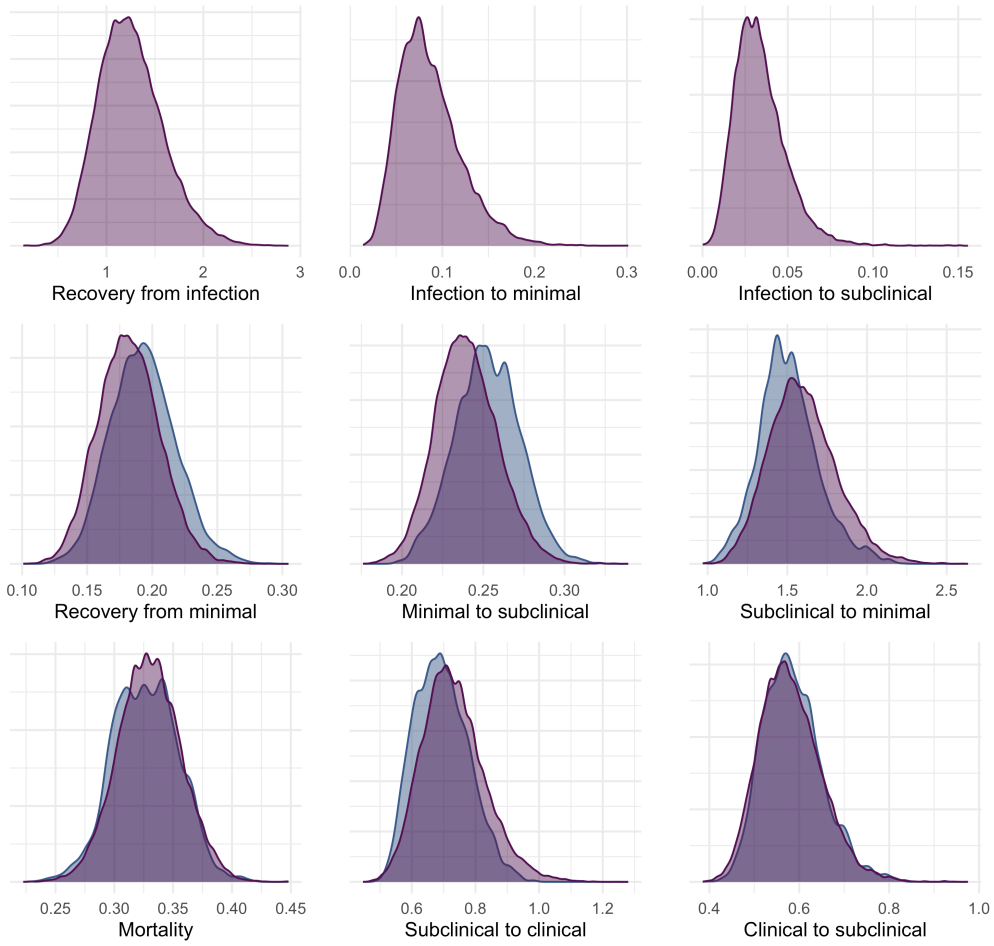


FIGURE 6.5: THE POSTERIORS FOR EACH OF THE PARAMETERS IN THE FULL INFECTION TO DISEASE MODEL. THE FIRST ROW SHOWS ALL PARAMETERS STARTING AT INFECTION. THE SECOND AND THIRD ROWS SHOW THE PARAMETERS THAT ARE IN BOTH THE DISEASE ONLY MODEL (CHAPTER 4) AND THE INFECTION TO DISEASE MODEL. POSTERIORS FOR THE INFECTION TO DISEASE MODEL ARE SHOWN IN PURPLE, WITH THE POSTERIORS FOR THE DISEASE ONLY MODEL IN BLUE.

To demonstrate the trajectories these parameters describe, I have created a Sankey plot. Figure 6.6 shows the progression of a model cohort with *M.tb* infection over the course of 10 years, with an intermediary follow-up point at two years. The height of each box is proportional to the number of people from the initial cohort with that stage of infection or disease. The colours of the paths represent the final disease state of each trajectory, and the widths represent the proportion of people following that path. The paths depicting recovery from infection are transparent to allow for clearer interpretation of the remaining pathways.

Recovery from infection without progression to disease is the most common outcome over the course of 10 years in the model (91% (95% UI 82% — 95.9%)). Most of this progression occurs in the first two years ((93.4% (95% UI 80.2 — 98.8%)) of those who recover). After 10 years, almost no individuals remain infected, (0% (95% UI 0 — 0.1%)) with everyone either having progressed to disease or recovered.

For those who progress beyond infection, recovery from disease is the most common outcome in the model (64.0% (95% UI (49.3 — 77.1%))). Individuals with subclinical or clinical disease at two years have a low chance of still being infectious 8 years later (5.0% (95% UI (0.0 — 21.7%))). For 1.6% (95% UI 0.6 — 3.6%) of all simulated individuals infected by *M.tb*, infection and the subsequent disease will persist for over 10 years.

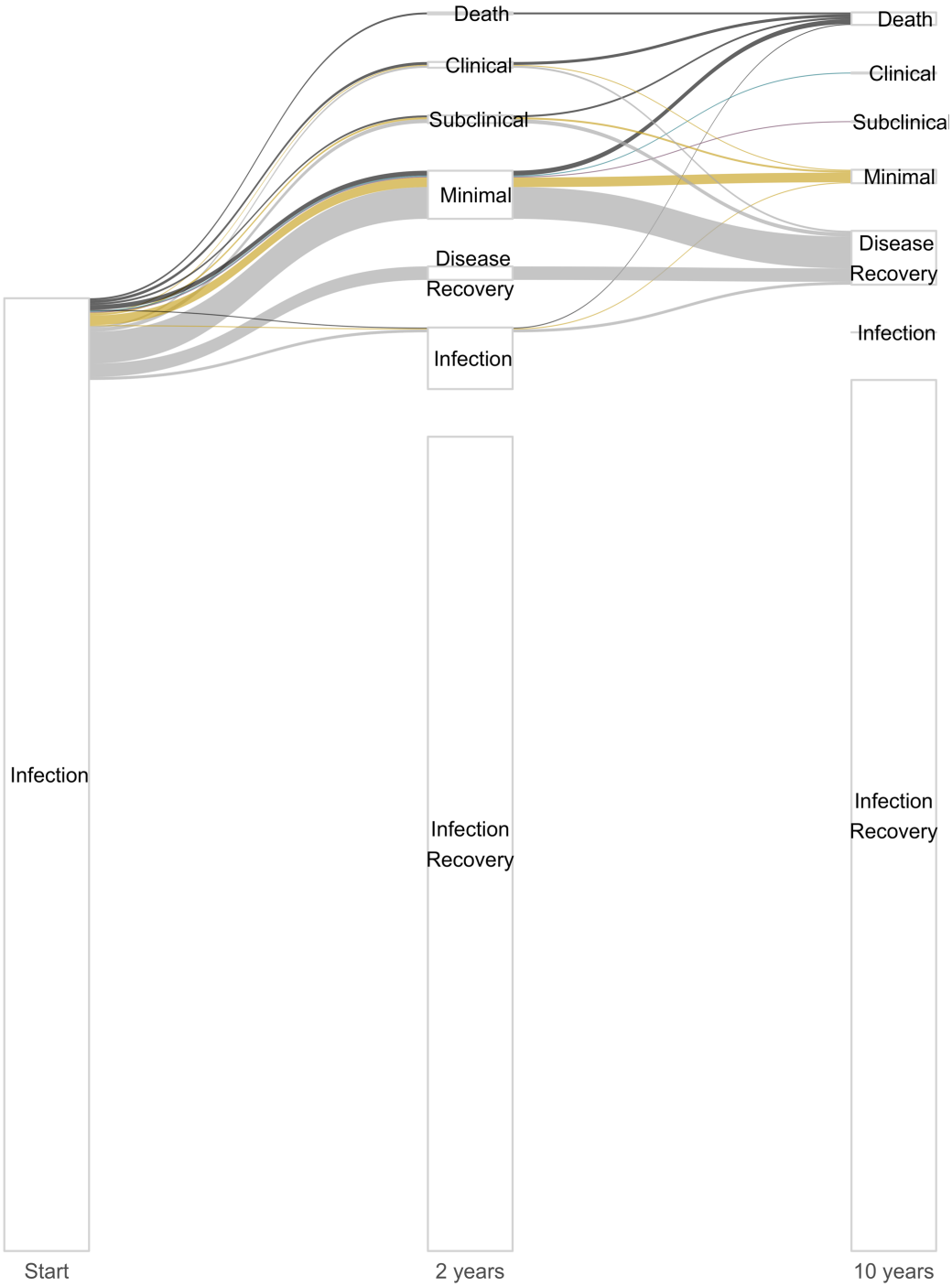


FIGURE 6.6: THE TRAJECTORY OF INFECTION OVER TEN YEARS, WITH AN INTERMEDIARY POINT AT TWO YEARS. THE HEIGHT OF EACH BAR SHOWS THE NUMBER OF INDIVIDUALS WITH THAT STATE AT THAT TIME, THE COLOUR OF THE PATHS REPRESENTS THE FINAL STATE OFF THE PATH, AND THE WIDTH OF EACH PATH IS PROPORTIONAL TO THE NUMBER OF PEOPLE FOLLOWING THAT TRAJECTORY.

6.4 Discussion

This work provides a coherent, data-driven natural history for the entirety of TB from first infection to disease to, ultimately, death or recovery. The search for data was extensive, and so this is likely the best parameterisation possible based on historical data.

There are only minimal differences between the disease only fit parameters from Chapter 4 and the matching disease parameters presented here, which confirms that these fits are consistent. The data to inform progression from infection only directly affect progression from minimal to subclinical, however, as all the disease states are closely linked, an incompatible value for minimal to subclinical would have knock on effects on all the rest of the parameters.

Model structures that best represent data from the Sutherland BCG trials (not used here due to the inconsistent reporting of infection timelines) have a variable rate of progression from infection to disease.[15,16] Although this model structure only uses one infection state, as opposed to two latent states for fast and slow progression, the resulting progressions to subclinical and minimal disease can, respectively, be considered fast and slow progression.[16] Progression directly from infection to subclinical doesn't necessarily represent a different mechanism in disease, just a faster progression through the stages that equivalently can be represented through a direct progression.

There is a very high rate of recovery from initial infection with this parameterisation. This has implications for the long-held assumption that infection is lifelong. It would mean that estimates that 25% of the world being currently latently infected, while reflective of the proportion exposed, is likely substantially overestimating the proportion currently with a viable infection.[17] An annual rate of recovery of 1.23 (0.69, 1.99) would mean that the mean *M.tb* infection lasts about 10 months discounting progression to disease. As seen in Figure 6.6, after two years there are still a small number of people with infection who have not either progressed or recovered (6% (95% UI 0 — 30%)), but by ten years, almost all have progressed or recovered (100% (95% UI 99.9 — 100%)). The long term reactivation of dormant infection, as seen in a molecular study of a father and son who developed an identical strain of disease 33 years apart, would seem to be an abnormal event with the model structure.[18] There is no good documentation or data on long term reactivation of *Mtb* infection and so the parameterisation is only based on short-term data. Data from improved tests that can accurately detect infection and potential for progression would be required to inform any long-term reactivation. The general estimate is that between 5 and 10% of people infected with *Mtb* progress to disease, which is exactly the result seen here.[19]

The data describing the parameters between states were sourced from studies with cohorts without treatment, as described in Chapter 3.[5,20]However, the data used to inform the ratios between subclinical and clinical, and minimal and infectious, are from populations where treatment is widely available.[8–10] To my knowledge, no equivalent data on prevalence ratios are available from the pre-treatment populations. It is likely that treatment, particularly if it relies on passive case detection, will skew the ratio of subclinical to clinical to a lower proportion of clinical. Despite leaving a wide prior, the calibration found that without treatment this proportion skewed towards subclinical disease, which was unexpected. Minimal disease is not a state with a gold standard test, and so the estimated proportion of people with minimal disease in the data, is actually the proportion of people where an expert reader has decided the radiographic film is likely showing TB. This is not precise, as I have previously alluded to, by reducing the cohort size of minimal disease in Chapter 4 by 25% to account for those likely misdiagnosed. It would be similarly likely that minimal disease here is also overestimated. However, the calibration of the model also skews towards a higher proportion of people with minimal disease.

The data collected and used did not adequately stratify by risk factors such as age, gender, or nutrition. The subsequent parameterisation therefore contains a mixture of these, and other, risk factors. Stratification by these risk factors would need to be informed from extra data sources as any stratification of the data used would likely render the parameters non-identifiable. Two important stratifications for TB models are drug resistance and HIV status and neither of these were present in the data to have the possibility of inclusion for stratifications.[16,21] Whilst the model and parameterisation are a good fit for the HIV negative population data that were collected, this will not transfer to an HIV positive population, and will again need more data to inform either new parameterisations for each stratification of HIV status, or data to estimate how the parameters will likely change for each stratification.

The extension of this model from three disease states to a full progression from initial infection, is based on a new combination of data, while still matching prior approaches and estimates of progression. Although not possible to include all relevant risk factors, the parameterisation provides a baseline model for the natural history of *Mtb* infection and TB disease using a small but strong selection of data and is ready to be extended as needed. The results have also opened new points of discussion on the nature and longevity of *M.tb* infection without progression to disease.

6.5 References

1. Daniels, M. Primary tuberculosis infection in nurses (summary of the prophit survey). *Lancet* **244**, 165–70 (1944).

2. Madsen, T. Studies on the epidemiology of tuberculosis in denmark. *Acta Tuberculosea Scand (suppl)* **Supp 6**, 1–176 (1942).

3. National Tuberculosis Institute, B. Tuberculosis in a rural population of South India: a five-year epidemiological study. *Bulletin of the World Health Organization* **51**, 473–88 (1974).

4. Poulsen, A. Some clinical features of tuberculosis. 1. Incubation period. *Acta Tuberc Scand* **24**, 311–346 (1954).

5. Richards, A. S. *et al. The natural history of TB disease-a synthesis of data to quantify progression and regression across the spectrum*. 2021.09.13.21263499 https://www.medrxiv.org/content/10.1101/2021.09.13.21263499v1 (2021).

6. Tiemersma, E. W., van der Werf, M. J., Borgdorff, M. W., Williams, B. G. & Nagelkerke, N. J. D. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PloS One* **6**, e17601 (2011).

7. Ragonnet, R. *et al.* Revisiting the Natural History of Pulmonary Tuberculosis: A Bayesian Estimation of Natural Recovery and Mortality Rates. *Clinical Infectious Diseases* **73**, e88–e96 (2021).

8. Onozaki, I. *et al.* National tuberculosis prevalence surveys in asia, 1990-2012: An overview of results and lessons learned. *Trop Med Int Health* **20**, 1128–1145 (2015).

9. Frascella, B. *et al.* Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. *Clinical Infectious Diseases* **73**, e830–e841 (2021).

10. Mungai, B. *et al. Accuracy of computer-aided chest X-ray screening in the Kenya National Tuberculosis Prevalence Survey*. http://medrxiv.org/lookup/doi/10.1101/2021.10.21.21265321 (2021) doi:10.1101/2021.10.21.21265321.

11. Okada, K. *et al.* Epidemiological impact of mass tuberculosis screening: A 2 year follow-up after a national prevalence survey. **NA**, (2012).

12. Murray, L. M. Bayesian state-space modelling on high-performance hardware using LibBi. *arXiv:1306.3277 [stat]* (2013).

13. RStudio Team. *RStudio: Integrated development environment for r*. (RStudio, PBC., 2020).

14. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2020).

15. Sutherland, I. *The ten-year incidence of clinical tuberculosis following 'conversion' in 2,550 individuals aged 14 to 19 at time of conversion*. (1968).

16. 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* **18**, e228–e238 (2018).

17. Houben, R. M. G. J. & Dodd, P. J. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS medicine* **13**, e1002152 (2016).

18. Lillebaek, T. *et al.* Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection. *The Journal of Infectious Diseases* **185**, 401–404 (2002).

19. Sutherland, I., Svandova, E. & Radhakrishna, S. The development of clinical tuberculosis following infection with tubercle bacilli. *Tubercle* **63**, 255–68 (1982).

20. Sossen, B. *et al.* The Natural History of Untreated Pulmonary Tuberculosis in Adults: A Systematic Review and Meta-Analysis. doi:10.1101/2022.08.30.22279374.

21. Sharp, P. M. & Hahn, B. H. The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 2487–2494 (2010).

Chapter 7 Discussion

In this thesis, I set out to understand the natural history of pulmonary TB through utilising the literature published before the development of anti-tuberculosis drugs.

In this chapter, I will summarise the findings from each of the research chapters and assess how these findings achieved the aims and objectives stated in Chapter 1. I will then discuss the strengths and limitations of my work, and how these should be taken into account when interpreting the results. Following this, I will look where my results can be used in future work, in both research and policy, and how this differs from what is currently used. I will also discuss gaps my work still leaves open to be investigated. Finally, I will discuss my overall conclusions from this thesis.

7.1 Summary of findings

This thesis sought to address two aims:

- Create a model of pulmonary TB in adults from infection to death that incorporates the spectrum of disease with data-driven parameter estimates
- Describe how an updated model structure could affect our understanding of best practices for finding and treating people with disease

These aims were addressed through five objectives, over four chapters:

- Conduct a systematic review of pre-treatment literature to find data on progression of pulmonary TB in adult cohorts (Chapter 3)
- Develop a natural history model of the spectrum of TB disease and calibrate the parameters to data from the systematic review (Chapter 4)
- Extend the natural history model to include progression from infection and calibrate to additional data from the systematic review (Chapter 6)
- Simulate an untreated population through disease to quantify the progression and regression pathways taken (Chapter 4)
- Compare the effectiveness of different screening programmes to detect infectious disease on simulated treated populations (Chapter 5)

Here I will summarise the key findings from the thesis and then discuss each chapter in more detail. This discussion will centre on how I addressed each objective and then compare the results to other published works and gaps in the literature.

7.1.1 Key findings

- Pulmonary TB disease encompasses a wide range of presentations from small pathological changes in the lungs, to asymptomatic but infectious disease, to disease with a range of symptoms (Chapter 3)
- Disease progression may not be linear; symptoms and infectiousness may abate, and radiological abnormalities regress (Chapter 3)
- The median duration of infectious disease without treatment until resolution or death, based on data from the pre-treatment era, is likely much shorter than previously estimated, at about 12 months (Chapter 4)
- Symptoms are not guaranteed in TB disease, with half of all people with subclinical disease not expected to develop symptoms, and undulation between disease states can mean that neither symptoms nor bacteriological results can be considered static (Chapter 4)
- Of the currently available tools for active case finding, radiological screening followed by confirmatory Xpert in communities is estimated to have the highest impact on reducing transmission and prevalence, by virtue of detecting the highest proportion of people with subclinical disease (Chapter 5)
- TB infection may not be lifelong, with most people clearing an infection quickly without progressing to disease, and the majority of any progression to disease occurring within the first two years from initial infection (Chapter 6)

7.1.2 Systematic review of historical literature

Chapter 3 addressed one objective towards aim 1:

> Conduct a systematic review of pre-treatment literature to find data on progression of pulmonary tuberculosis in adult cohorts

Treatments for TB are, in most instances, effective and provide much better outcomes than no treatment. For this reason, any trials or studies where people are found to have bacteriologically-positive disease provide the necessary treatment to those individuals. This means that there are no recent observational studies of the progression of untreated disease. In Chapter 3, I presented a systematic review of the literature from the first half of the 20th century and, where possible, later studies of observed, untreated TB progression.

There was a plethora of research in the first half of the 20th century with an aim to better understand the progression through infection and disease, with the review finding almost 1,700 relevant titles, and over 200 studies that passed the question-specific exclusion criteria. With a

focus on pulmonary disease in adults, and a reliance on bacteriological testing and some form of imaging to report disease, much of the research does not meet rigorous trial and diagnostic standards expected now, with almost half the studies being excluded due to risk of bias. Another 40% of all relevant studies were excluded due to lack of coherent reporting sufficient to extract data. This left 24 studies to be included in the review. These studies had well defined testing and follow-up protocols and thus the data extracted from them could be considered sufficiently high quality.

Most modelling frameworks of disease assume TB disease to be a homogeneous state, sometimes split by smear status, with no stratification within disease. However, the data collected from the 24 studies showed that up to two thirds of a group with bacteriologically-positive disease could regress to a bacteriologically-negative state of disease, as defined by pathology discovered through radiography, over the course of five years, and up to 50% of a cohort will progress in the opposite direction. The parameter for transmission in models is not a measured value, and instead is a variable that is adjusted during calibration to ensure that data points are matched. However, with regression from bacteriologically-positive to bacteriologically-negative disease, not all people classified as having "active" pulmonary disease would necessarily be infectious, and it is also likely that there is a spectrum of infectiousness. This is then different from the classic paradigm of a homogeneous "active" disease state with equal infectiousness throughout. The review also gave the opportunity to collect data on symptom presentation which, whilst less well reported than either bacteriology or radiology, can help provide insight on different presentations of disease. Symptoms are important to consider when assessing the potential impact of screening programmes, as passive case detection relies on people seeking healthcare for symptoms, and those that do not have symptoms may still have infectious disease.

This is not the first review of historical literature for the purposes of understanding aspects of the natural history of disease, but it is the only one with a focus on the progression through pulmonary disease in adults.[1,2] While Marais *et al.* focussed on the distinct question of intra-thoracic disease in children, Tiemersma *et al.* also focussed on pulmonary disease in adults.[1,2] However, beyond the population in question, Tiemersma *et al.* aimed to understand the duration of disease and the difference in mortality rates between smear-positive and smear-negative disease.[2] In comparison, I have collated data on the progression through the spectrum of disease. I have also avoided categorisation into smear-positive and smear-negative disease. Whilst these terms have been widely used in recent work and classifications, in pre-chemotherapy studies, individuals with bacteriologically positive disease were not categorised by smear status. When Tiemersma *et al.* conducted their initial analysis, the WHO differentiated

bacteriological positivity by smear status (until 2013) and data collection at the time highlighted this differentiation.[2–4] Thus, the purpose of the work by Tiemersma *et al.* was to understand the differing risk of mortality based on smear status, and so inferring smear status based on the data was the only way to achieve this.[2] However, the current approach in data collection and by the WHO is to consider all bacteriologically-positive disease as similar, regardless of smear status, due to Xpert replacing smear as the tool for diagnosis.[4] For me to create a distinction between smear status from my collected data, based on assumptions that are not explicitly stated in the text, would reduce the strength of the data and would provide minimal additional benefit.[2,4]

7.1.3 Modelling the natural history of TB disease

Chapter 4 addressed two objectives, the first towards aim 1 and the second towards aim 2:

> Develop a natural history model of the spectrum of TB disease and calibrate the parameters to data from the systematic review
>
> Simulate an untreated population through disease to quantify the progression and regression pathways taken

Prevalence surveys have found a high proportion of people with bacteriologically-positive disease do not present with TB symptoms, generally defined as one or more of a cough, a fever, night sweats, haemoptysis, and weight loss.[5,6] However, most models of TB natural history do not differentiate bacteriologically positive disease by symptom status.[7] There have been a lack of data available to directly parameterise a model structure that discretises the disease spectrum. Models are fitted to prevalence and notification data, or accepted use natural history parameters from a few select sources. However, my work in Chapter 3 has sourced data that describe progression and regression through the spectrum, discretised by the tests that were used within the studies. In this chapter, I used these data to inform the parameterisation of a three-state disease model.

My model structure splits "active" disease into clinical, subclinical, and minimal disease states. **Clinical disease** is bacteriologically positive, and symptomatic. **Subclinical disease** is bacteriologically positive but asymptomatic and can be detected through radiographic screening due to lesions present on the lungs. **Minimal disease** is bacteriologically negative, however, like subclinical disease, can be detected through radiographic screening. It reflects pathology that is not currently infectious. There are progression parameters from minimal to subclinical and subclinical to clinical, and regression parameters from clinical to subclinical and subclinical to minimal, making a three state, linear structure. In addition, there is a parameter for natural recovery from minimal disease and for TB death from clinical disease.

In addition to the data collated in the systematic review (Chapter 3), I used four external data sources as priors to help inform the fit: mortality rate, duration of infectious disease, ratio of prevalent subclinical to clinical disease, and ratio of prevalent minimal to infectious disease. The mortality rate was informed by the systematic review by Tiemersma *et al.* on the duration and fatality of TB where they conflated the historical terms of open and closed disease with smear-positive and smear-negative disease.[2,8] I used parameters from the subsequent analysis by Ragonnet *et al*. where they transformed the data from the review by Tiemersma *et al.* into a parameter for modelling, taking into account contemporary mortality rates.[2,8] Although presented as data and parameters from smear-positive and smear-negative disease, I believe the interpretation of open TB can more closely represent symptomatic disease and closed TB as either asymptomatic or bacteriologically-negative disease. There was no consistent definition of "open" TB, with some using symptoms as the defining factor, and others using bacteriology.[9–11] Of the studies included in the review, some explicitly state that the mode of testing is through smear (of patients at a dispensary).[11–13] Other studies have no mention of how, or if, they test bacteriologically, but follow diagnostic criteria that originally were based solely on observations from radiography, and latterly had symptomatic presentations added.[11–13] Even later studies where both culture and smear were used, they were treated as complimentary tests with little analysis separating the smear-positive cases from all bacillary cases.[14] For these reasons, I have used the mortality rate from "smear-positive" disease to inform the mortality rate for clinical disease.[2,8] The estimate for the duration of infectious disease also originates from the Tiemersma *et al.* review, derived separately from two studies found during the review.[2,15,16] The ratios between disease states were based on an assumption of disease reaching a steady state, and then were informed by a review of prevalence surveys (for the subclinical to clinical ratio), and a post-hoc analysis of a single prevalence survey (for minimal to infectious).[6,17]

I combined all these data and created a mechanism to calibrate the model structure to all the data simultaneously within LibBi, a Bayesian inference system for state space models.[18] I specified uniform priors for between state transition model parameters of 0 to 12 per year (representing no movement to a mean duration of one month in a state) and the same for recovery from minimal. For death from clinical, the duration, and each of the ratios I assumed a normal distribution around the value given, either with standard deviations to fit the reported confidence intervals (death and subclinical to clinical ratio) or with wide standard deviations of 0.5 (duration and minimal to infectious ratio.[2,6,8,15,17]

With this fit, I found estimates for each parameter and for each of the extra data points. The parameter found between minimal and subclinical was the slowest of the four parameters

between states, while subclinical to minimal was the fastest. The median duration of infectious disease was found to be about 12 months, which is half the value suggested by the prior.[2] The ratios of subclinical to clinical and minimal to infectious disease were larger than the prior, with almost twice the prior expected proportion of people with minimal disease.

With the parameter ranges estimated, I could then simulate individual progressions through disease to understand the heterogeneity in trajectories in the absence of treatment. For prevalent subclinical disease, over the course of 5 years, the most common outcome was natural recovery. Half of all people with subclinical disease would never be expected to experience symptoms.

Although no other natural history models are parameterised in the same way, or with the same structure, there are still possible comparisons. Ku *et al.* used reported prevalence and notifications in different countries to estimate the duration of disease.[19] They estimated the mean total duration of disease before notification, and hence treatment, could range from 9 months in Ethiopia to 36 months in Lao.[19] My estimate for median total duration without treatment is at the lower end of this range at 12 months. However, maintaining the assumption of an exponential distribution on disease duration, the mean duration is 18 months. Whilst the model structure Ku *et al.* have used does contain both an asymptomatic and symptomatic state (subclinical and clinical equivalents) and the potential for "self-cure" from each state, there is no possibility for undulation between subclinical and clinical states and they only consider disease detectable through bacteriological testing.[19] Without regression of disease that allows for subsequent progression, duration of disease is likely underestimated and progression and recovery rates overestimated.

Ryckman *et al.* have also created a model for bacteriologically-detectable disease.[20] This model contains four compartments based on the combinations of smear and symptom status (each either positive or negative).[20] Rather than the approach from Ku *et al.* where self-cure could occur from either symptomatic or asymptomatic disease, Ryckman *et al.* specify that spontaneous resolution only occurs from asymptomatic, smear-negative (bacteriologically-positive) disease.[19,20] The latter is closer to my approach, although I have included the extra compartment of minimal disease before full resolution. The mean duration of disease ranged from 4.3 months to 15.5 months dependant on initial disease state, with a population mean of 7.7 months.[20] This is much shorter than the duration found in my model, however these durations are in the presence of treatment, whereas my main analysis has no treatment included, just relying on the natural history data collected.[20]

7.1.4 The impact of population level screening for TB disease

Chapter 5 addressed one objective towards aim 2:

> Compare the effectiveness of different screening programmes to detect infectious disease on simulated treated populations

Community based active case finding was widely used in the first half of the 1900s with both community and occupational mass radiography. In 1964 the WHO expert committee recommended that mass radiography should only be a secondary level of case finding, prioritising those who seek treatment for symptoms first, and by 1974, the same committee officially abandoned the policy of mass radiography.[21,22] Almost 50 years later, the global TB burden is not reducing fast enough, with the End TB 2020 milestones missed, and so with improvements in x-ray technology along with wider interest in alternative screening tests, the question has returned to understanding the benefits that mass screening could bring.[4,23–27] By combining my parameter estimates for disease, and transmission potential of subclinical disease from Emery *et al.*, I created a tool to compare the performance of different screening tests with respect to their use as a mass screening tool within a community.

I proposed a two-step process: screening an entire population with a single test and then confirming any screen positive tests with Xpert. I assumed that process took a year to screen the population, with no missed or double screenings, and was then repeated twice more. These three screenings took place over a five-year period, with active screening in years one, three and five, and only passive case detection in years two and four.

To initialise the model, the population was assumed to be in a steady state. The number of required progressions to minimal disease to maintain that steady state was compared with the number of people with subclinical and clinical disease, and their relative infectiousness to create a relationship between number of people with infectious disease and the expected progressions to minimal disease, for when the expected increase in treatment reduces the number of people with infectious disease.

I compared eight different screening tests against a baseline of passive case detection: symptoms, digital radiography, C-reactive protein (CRP), blood biomarkers at two levels (CORTIS), Xpert, and two hypothetical tests based on the WHO target product profiles for a new screening test.[26–29] Where Xpert was used as a screening test, there was no subsequent confirmatory testing. The data available on these tests for their performance within each disease state were limited. Minimal disease is currently only detectable through radiography or other imaging techniques, and there is no gold standard against which to measure the

performance of screening tests. Therefore, the performance in each of these tests in minimal disease was estimated. Where possible, estimates of sensitivities for minimal (and subclinical, where necessary) were extrapolated from the available data. Otherwise, the false positive rate for tests in people without TB was assumed to be the true positive rate for those with minimal disease, which should provide a reasonable lower bound for test performance.

I modelled the potential impact of these screening tests, each starting from the same baseline, with the same screening algorithm. The impact was measured on five metrics: the reduction in prevalence, the reduction in transmission, number of false positive diagnoses, people with subclinical or clinical disease who were missed for treatment, and the number of confirmatory tests required for detection of subclinical or clinical disease. On combining of all these metrics, community x-ray screening followed by confirmatory Xpert finds the highest number of true positives, and the fewest false positive diagnoses. However, the theoretical screening tests based on the WHO TPPs provided a higher coverage of treatment to those with infectious disease, and thus a greater reduction in prevalence. If such a test could be created and distributed cheaply, and monetary support provided for the required Xpert confirmatory tests, the reduction in the transmission and prevalence of disease could be large.

The results of this modelling are limited due to the choice of sensitivities for each screening test. The choices were the best given the available data, however, the data available were limited. Sensitivities are reported, albeit in different settings, as the proportion of disease that is detected, mostly assuming a single homogeneous state of "disease". Occasionally sensitivities are reported for smear-negative and smear-positive disease separately.[30] However, with the move away from treating disease as a homogeneous state and towards reflecting a spectrum in both our understanding and policy recommendations, an understanding that these tests perform differently across the spectrum due to different thresholds of disease is necessary. For clinical and subclinical disease, the gold standard of liquid culture is usable, however, a gold standard test for detecting minimal disease would need to be created to help fully capture the spectrum.

This analysis can be repeated for any selection of tests, prevalence, and relative infectiousness. The model outputs were largely comparable with the results of the ACT3 trial.[23] Therefore, with refinements to the model inputs and setting specific assumptions, this model structure can become a flexible tool for assessing the expected impact of a targeted active screening intervention in a new setting. This should help provide the necessary evaluation on whether an intervention could be successful in reducing prevalence of disease, and whether any other

interventions would be more successful, either providing greater reductions or providing the same reduction more efficiently.

7.1.5 Infection to disease

Chapter 6 addressed one objective towards aim 1:

> Extend the natural history model to include progression from infection and calibrate to other data from the systematic review

In Chapter 4, I parameterised a discretised form of the spectrum of disease for the purpose of mathematical modelling. Whilst the disease-only model structure can be utilised for answering some questions (such as the work in Chapter 5), it is not a full model of TB natural history. With the changes made to the disease structure it would have been inadvisable to assume that the three disease states could simply replace the one state that models normally use in a TB model structure, and that progression to disease would remain the same.

Using data originally identified through the systematic review process in Chapter 3, I extended the model structure from Chapter 4 to incorporate progression from infection. Commonly, TB models split progression from infection to disease into two pathways, with either states or progression rates referred to as fast and slow latent.[7] Of the model structures I tried for this work, the best-fitting model structure to the data had an analogous structure to that of fast and slow latent. In my chosen model structure, following infection, fast progression bypasses the minimal disease state, straight into subclinical disease. Slow progression progresses from infection to minimal disease.

Fitting the data to this structure, I found that the most common trajectory from infection in the model was recovery from infection, at an annual rate of 1.2. Progression to minimal was next at 0.1 per year, and progression direct to subclinical disease was the least common with a rate of 0.02 per year. Estimating the number of people remaining with an infection that has neither progressed nor recovered over time was modelled with an exponential decay. With these parameters I find that within 2 years, fewer than 10% of all people will have an infection that has neither progressed nor cleared. By 5 years this proportion reduces to as little as 0.1%.

In summary, in Chapter 6, I have presented a coherent model from initial infection to disease that both fits with my original estimates of disease progression, and also with widely accepted estimates for proportions of people who progress from infection. Whilst the conclusion of high recovery from infection rather than a latent disease state differs significantly to previous modelling assumptions, there are previous postulations on recovery of infection for almost all those who do not progress within the first two years of infection.

7.2 Strengths, Limitations, and Alternative Perspectives

7.2.1 Strengths

7.2.1.1 Grounding in data

Every part of this thesis is grounded in data, and I strived to ensure that these data were of high quality with as little bias in selection or processing as possible. For the systematic review (Chapter 3) I helped develop a rigorous protocol, that was published and adhered to, meaning that many of the studies were excluded because they were not explicit enough in their testing and reporting of disease states. This left me with a smaller selection of data than ideal, but all of those data were reported to a similarly high standard and of comparable standard to that used today. The selection of data from the review that I used for the modelling did not provide direct data points for each transition within the model. However, there were data describing both progression and regression, and through combining these data with data from outside the review, a restricted parameter space could be found for each transition.

The data I added from outside the review in Chapter 4 came from, where possible, previously conducted systematic reviews.[2,6,8,15] Whilst a systematic review does not guarantee the quality of data, the selected reviews included a large number of studies which helps to reduce the impact of bias within each individual study. Despite the estimate for the duration of disease coming from a wide reaching review, the calculation was based on just two studies, both analysed separately.[2,15,16] As such, I assumed a wider prior on the duration of disease than I did for the ratio of subclinical to clinical disease, to represent the uncertainty in the estimate.[2,5,6,15,16] Similarly, the estimate for the ratio of minimal to infectious disease was from a single study, so was given a wider prior in the fit to represent the uncertainty.[17,28]

When comparing different screening tests in Chapter 5, data are more scarce on performance. However, where available, the performances of each test have been used as quoted as the result of trials or studies where they were used.

7.2.1.2 Generalisability

The data used throughout this thesis come from a wide time range (1923 to current). A concern with older data is that they are no longer comparable with contemporary data. Although I cannot say for sure that this is not an issue, comparisons of the same data types collected in the systematic review do not show a distinct difference between old and new data, with the earliest and latest data points showing similar progression rates. This may be explained by the fact that many of the structural drivers of TB that are present now, such as malnutrition, poor housing, and poverty, have likely been present in all the data used. A similar issue could be present in

reverse, with the ratios between subclinical and clinical disease seen now not being applicable in the early 1900s. To my knowledge, no equivalent data on prevalence ratios are available from pre-treatment populations however, studies in high prevalence communities in Alaska showed similar ratios in the 1950s just as treatment was becoming available.[31,32] Along with being temporally diverse, the data from the systematic review are geographically diverse. Most of the data came from North America and Europe, but there were also studies included from Asia, Africa, and South America.

The screening algorithm, presented in Chapter 5, whilst completely theoretical, brings potential generalisability and wide applicability. In Chapter 5, I used the baseline parameters for natural history from Chapter 4 and best estimates for test performances. Further research into test performances in different stages of infection and disease could help improve the accuracy of results from the algorithm. Similarly, the population distribution and prevalence can be adjusted to match any given population. If parameter adjustments become available that accurately transform the natural history parameters to ones that accurately describe progression of disease in people living with HIV or in drug-resistant forms of TB, then those parameters can also be updated.

7.2.2 Limitations

7.2.2.1 Age and availability of data

Despite all the strengths of the data, it also comes with limitations. Much of the data are old. As stated previously, data matched by transition type do not show large differences in progression over time. However, beyond the applicability of the data, there are other issues with the sources of the data.

The reporting of the data was different to now and the style of writing has changed. The abstract first style of publishing did not exist. Many studies did offer a summary at the end that would be similar to an abstract but these were not available with the title when performing an initial database search for the systematic review (Chapter 3). This meant that the title screening was based on sometimes as little as two words. Therefore, there may be some studies that were not included because the title seemed irrelevant, despite relevant content within the work.

For studies that did pass the title screening, many were not freely available on the internet. Almost 10% of the studies deemed relevant were not available through the multiple internet and library sources searched. Even with access to a number of personal research collections of studies of TB, it was not possible to find all studies. Beyond these searches, it becomes harder to

source studies conducted in the early 1900s, not least because it is no longer possible to contact the researchers themselves for copies of their publications.

For the more than 200 studies that made it past all these obstacles, it came back again to the writing and reporting styles. Over half were excluded on the risk of bias, with 70% of those excluded based on the reporting of the testing regimens. Often this was due to a lack of any clear statement about testing was accomplished or how reported disease states were linked to those tests. If it had been possible to contact the study authors to confirm their testing regimens, more of the studies could have been included. Similarly, for the studies where there were insufficient data to be extracted, contacting the study authors for clarity may have made more data available.

The final studies that were included provided invaluable data on the progression and regression of disease. However, some of the transitions in my model had no data directly describing them. This does not mean that the transitions never happen, as qualitative reports suggest that they do, just that they were not recorded. The lack of record means that there is no confirmation of how often these transitions happen. In particular, there were very few studies that carefully reported data on symptoms, and so differentiating between clinical and subclinical disease became challenging. There were data describing movement between minimal and clinical disease in both directions and, using the assumption that the only progression between minimal and clinical is through subclinical, it was possible to constrain all the parameters without data describing each. The limited data for those transitions is reflected in the wider confidence intervals around the posterior estimates. Ultimately, the data found were sufficient, but any more data, or the ability to discuss results with the original study authors could have made a big difference in the data included in the review and may have allowed for smaller confidence intervals in the estimates.

Data on screening test performance, although contemporary, were also not sufficiently available, especially for minimal and subclinical disease. Minimal disease is not currently a well recognised category of disease to test for, and some tests (CRP specifically) were trialled on cohorts of HIV-positive people almost exclusively, and so I cannot be confident that they apply equally to a simulated HIV-negative cohort. This is a much easier limitation to counter in the future, by advocating for wider testing of potential screening tests beyond people with culture-positive disease, providing suitable reference standards can be found.

7.2.2.2 Symptoms

This thesis has focussed on the concept of subclinical and clinical TB disease, differentiated by symptoms. However, symptoms are defined differently in different situations. Some prevalence

surveys define a positive symptom screen as more than two weeks of a cough, some as cough of any duration, and some as different combinations of the five symptoms.[6] It can be assumed that passive case detection in healthcare uses similarly diverse definitions. There are also reported issues of people either not being aware of their own symptoms or not reporting them. In addition, there are people who are always "symptomatic" through another cause, who then do not notice a change in their symptom status.[34]

This brings to the fore the question: what does subclinical disease actually mean? Through the systematic review in Chapter 3, symptom status was recorded as positive or negative based on the study-specific definitions of symptoms, which varied between studies. Some older studies, particularly sanatoria studies which tracked only a few individuals, kept records of multiple temperature measurements each day. More recent studies defined symptom positivity based on either a subsection of the WHO symptom screening list, or, more often, reporting of more than two or three weeks of a cough.[6,33,35] This brings uncertainty into the interpretation of my results. My intention is for subclinical to be defined as no reported symptoms, and clinical as any reported symptoms. However, as a maximum, subclinical will mean less than three consecutive weeks of cough at the point of testing, and will likely have a true threshold closer to no symptoms.

An adult experiences about two instances of respiratory illness each year.[36] With an estimated median duration of infectious TB lasting a year, it could be expected that, even if individuals otherwise have subclinical disease, they will experience some form of symptoms with the unrelated respiratory infection. As the symptoms of many respiratory infections are similar to those of TB, someone with "subclinical" TB but symptoms of another respiratory illness could seek care for those symptoms and have their TB detected. Although these symptoms would not be caused by TB, the detected disease would be considered clinical due to the diagnosis pathway. This would mean that the possibility of detecting subclinical disease is actually higher than the true baseline of subclinical disease if a respiratory infection with similar symptoms initiates TB confirmatory testing. On the other hand, the empirical data used to create the parameter estimates may have over-stated the number of people with clinical disease due to an unrelated respiratory infection causing similar symptoms.

7.2.2.3 HIV and drug-resistance

The data for the natural history parameterisation originate, mostly, from the pre-treatment era and before HIV was formally recognised.[37,38] As such, drug-resistance and co-infection with HIV were not dealt with in the underlying data for the natural history parameters. However with 16% of TB deaths in people living with HIV, and 4% of new infections and 20% of previously

treated infections having resistance to at least one drug, these issues are not insignificant.[4,39] Generally drug-resistant disease is assumed to progress similarly to drug-susceptible disease and differentiations between the two in models allow for different treatment regimens to be applied.[40] I did not collect any data to confirm whether this is a reasonable assumption. On the other hand, stratifications for HIV status do change the underlying natural history in TB models, and thus different parameters are required.[41] However there will not be the same data available to estimate the parameters required for the untreated natural history of TB stratified by HIV status, and so adjustments will have to be based on external calibration targets.

On the other hand, the data for screening test performance were sourced entirely from studies after drug-resistance and HIV were recognised.[17,26,27,42–44] The performance of CRP in particular was derived almost entirely from populations with HIV with a high prevalence of TB.[27,42–45] Without studies investigating the performance of screening tests in communities without HIV, it is not possible to know how the tests would perform differently, however the expectation is that they would. This means that although selected rigorously from the available data, the results of Chapter 5 may over-, or under-estimate the true impact a test could have. For CRP, it is likely that the sensitivity should be lower than estimated, whereas for the other tests any effect is less clear.[42,45] However there were insufficient data to help improve the estimates.

7.2.3 Alternative perspectives

7.2.3.1 Disease model structure

I opted to parameterise a three-state, linear model structure. The data collected in the systematic review support this choice, as can be seen from the ultimate parameterisation of the model. It also matches with previous conceptualisations of the progression of disease.[30,46–49] However, the data to support this structure could also support other model structures. Here I will describe some other model structures and why I ultimately chose the structure I did.

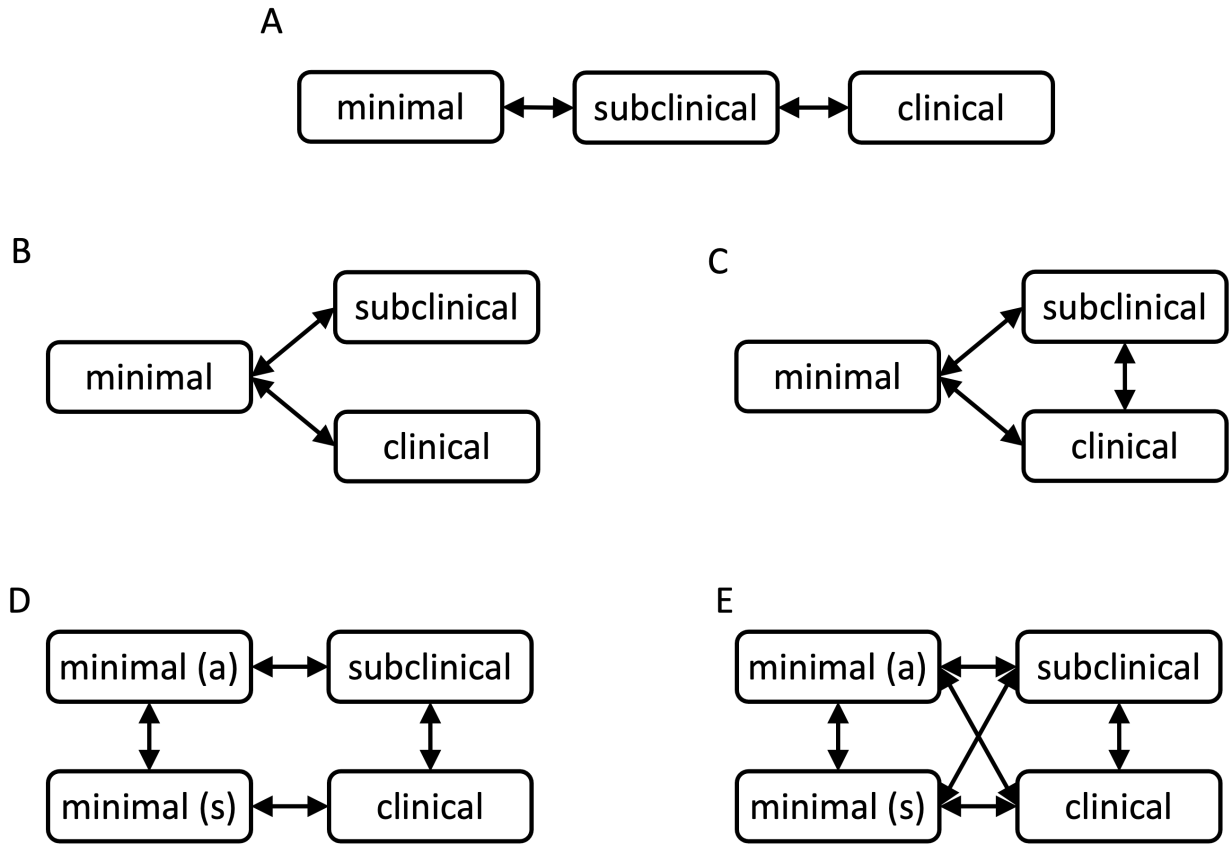


FIGURE 7.1: ALTERNATIVE STRUCTURES FOR THE DISEASE ONLY MODELLING WORK IN CHAPTER 4

There were limited data on symptoms found in the review (Chapter 3). In terms of the three disease states parameterised in Chapter 4 (and in Figure 7.1A), the only data explicitly mentioning a subclinical state showed a progression from minimal to subclinical. There were no data from subclinical to minimal, nor between subclinical and clinical. This could instead suggest a model structure that from minimal disease branches in one direction to subclinical disease and another for clinical disease. This would add no further transitions to the model, but they would instead be minimal-subclinical and minimal-clinical transitions, shown in **Figure 7.1B**. For this to be the case, subclinical and clinical disease would have to be distinct manifestations of disease, and the only way to develop symptoms from subclinical disease would be to recover from subclinical to minimal and subsequently progress. As I did for the main analysis, the parameter of mortality could come from the Ragonnet *et al.* analysis, but a question would exist whether death was possible from subclinical disease.[2,8] The equivalent parameter for smear-negative disease could be used for mortality from subclinical disease if it were deemed necessary but, with such a small value, it could also be considered negligible.[8] This analysis would be no more complex in terms of parameter numbers than the main analysis, but it would make dealing with the subgroup of people with bacteriologically-positive disease in a cohort presenting a mixture of symptoms difficult.

The data for cohorts where symptoms were present in only some of the people with bacteriologically-positive disease, or where symptoms were not mentioned at all, were termed infectious disease in the modelling. The split between subclinical to clinical in these cohorts was randomised for the purposes of fitting into the model structure. That both people with and without symptoms were considered to have similar enough disease to be analysed in one group could suggest that those working with these people felt movement between the two states was common, or at least not worthy of explicit separation. There would then need to be the possibility to transition between the subclinical and clinical states to allow for this, such as the structure in **Figure 7.1C**. An alternative way to visualise this structure is model A but with an extra set of paths directly between minimal and clinical. Again, this structure could be described by the data; however, it would leave the transitions between subclinical and clinical even more free from data than they were in the original fit. This is caused by two extra transitions, but no extra data, and the data that describe the progression from minimal to clinical (and minimal to infectious after the randomised split) only describing the transition between minimal and clinical (or minimal and subclinical) with no data explicitly informing the move between subclinical and clinical.

Other heterogeneities that could be included in the model structure include splitting the minimal compartment based on the presence of symptoms or the perceived activity of lesions seen in radiography. These extra data are not well reported for all studies, with the majority of reported minimal disease not reporting symptoms, and one third reporting mixed activity on lesions.[50] Therefore, any model structure that needed to include this difference would either require removing data due to insufficient clarity of data, or would have to make assumptions on what the true presentation was. Either of these heterogeneities within minimal could be structured as in **Figure 7.1D** or **Figure** 7.1**E**. Whilst D would be considered the simpler version of the model structure, there would still be insufficient data to inform even more transitions. There were no data collected between asymptomatic and symptomatic (or inactive and active) minimal states, and there would still be no direct data between subclinical and clinical. With limited data, other assumptions need to be put in place to artificially limit the potential solution space. Ryckman *et al.* found this with a similar structure, although splitting between smear and symptom status rather than bacteriological and symptom status.[20] Given the data available, and the assumptions required to make the data fit that structure, the situation becomes one of adjusting the data to match prior assumptions on model structure, whereas model structure should be led by the data available. The data limitations of Figure 7.1D, are only exacerbated in Figure 7.1E. It would only really be possible if there were new data that explicitly described the new transitions.

It is always possible to make a model structure more complex but what is more useful is finding the right balance between simplicity, representing the data available, and providing answers to the questions being asked. As stated by Hans Waaler in his work on early TB models:[51]

> *To be of practical value … a model should be sufficiently detailed to be realistic and at the same time be so constructed technically that it lends itself to rapid solutions by computer*

Whilst a single disease state is more simplistic, and there are well accepted parameterisations and fitting methodologies for the structure, it does not help to explain well-known heterogeneities seen in TB disease. The modelling approach of splitting bacteriologically-positive disease into smear-positive and smear-negative disease allowed for understanding the expected differences seen based on different testing methodologies. However, with a move away from smear differentials and towards understanding the effects of passive case finding and transmission from people who do not have clinical disease, an understanding of subclinical disease is required.[4] Given each of the complexities of Figure 7.1B – E, and the minimum required within the model structure, I believe that Figure 7.1A provides an optimal balance between simplicity, representation of the data, and potential to answer the questions being asked.

7.2.3.2 Infection to disease

The different model progression structures for infection to disease were discussed in Chapter 6. As with the disease structure for Chapter 4, the simplest structure that fitted the data was chosen. However, the result of the fit, in particular the speed at which people "recover" from infection, subverts the widely accepted understanding of persistent "latent" TB infection. For a long time, it was accepted that one third of the world was living with a latent infection that could, at any time, progress to disease. More recent modelling by Houben and Dodd estimated that proportion at closer to one quarter, but still all with a risk of progression to disease.[52] However, Behr *et al.* argue that our understanding of infection is incomplete, particularly around the meaning of a positive TST or IGRA test.[53,54] They suggest that the majority of people who progress from infection will do so within the first two years, and the majority of the remainder will instead recover from their infection having never progressed, leaving a much smaller group of people who could later progress to disease than a TST or IGRA survey would suggest.[53,54] Modelling work by Emery *et al.* also proposes, from autopsy and TST reversion studies, that infection is not lifelong, with a lower bound of 73% of people recovering from their infection within their lifetime.[55]

The most direct evidence against complete recovery of all infections is that of the Danish father and son who had identical strains of TB, 33 years apart.[56] The parameter estimate for recovery from infection found in Chapter 6 is 1.17 per year. This translates to a mean duration of infection of 9 months when all transitions out of the state are considered. This would agree with the hypothesis that most people are no longer infected after two years and instead just have an immunological memory that returns a positive TST or IGRA.[53–55] Due to the nature of an exponential decay, with the parameters found, whilst an occurrence of someone remaining infected without progressing for 33 years is highly unlikely, it is not a mathematical impossibility, particularly given the number of people who have ever been infected by *M.tb*.. Thus, such events are not excluded by my parameterisation. However, there are data that show that late reactivation is not as uncommon as the parameterisation would suggest. Other studies, where in individuals with immunosuppression, through receiving tumour necrosis factor (TNF) inhibition, solid organ or stem cell transplants, or through being infected with HIV, have observed the rate of TB disease development in those with positive TST results.[54] Once collated, these studies showed that between 1% and 11% of these individuals progressed to active disease.[54] Whilst the summary definition of active disease is vague, it would suggest that not all infections that do not progress are truly cleared.

An alternative possible explanation using the parameterisation from my model, is that the Danish son progressed to minimal disease, potentially with short bouts of subclinical (and even clinical) disease over the years, but only experienced sufficient symptoms to seek healthcare 33 years later. The same may be true of those who later became immunosuppressed. However, the data used to create these parameters did not extend beyond six years after initial infection. Therefore, it cannot be stated with certainty that what has been treated as a recovery from infection within the model is truly a recovery.

In countries with reducing prevalence, the average age at the time of disease diagnosis increases.[57] For example, in the UK, whilst the most common age group for onset of TB disease is 15-44 years, most who develop TB are people born outside of the UK in areas with higher TB prevalence.[57] When just considering the population born in the UK, there is a much higher risk in those over 75 years old.[57] This is highly unlikely to be because TB is actively circulating in only the older population, and potentially that they are harbouring an old infection from childhood, when TB prevalence in the UK was higher.[57]

Instead of recovery clearing all infections, an alternative explanation, however unlikely, could be that the "recovery" calculated in the model fit is actually a mix of true recovery and dormancy with a very low level of pathology or metabolic activity. This low level of pathology

would not have been detected in my pre-chemotherapy data but would be classified as minimal disease within my conceptual structure. A more sensitive chest screening test such as PET/CT could potentially detect this state and help improve understanding.[58] This could then be used as a gold-standard for TB pathology, which other screening tests could be compared against.

7.3 Recent modelling developments

Since I started working on this thesis, there have been other developments in modelling the spectrum of TB disease. Here I will discuss four papers and preprints that incorporate subclinical disease.

Ku *et al.* addressed the issue of duration of disease in both asymptomatic and symptomatic phases.[19] They note that prevalence to notification ratios that are used to characterise duration of disease differentiate between asymptomatic phases of bacteriologically positive disease prior to symptomatic episodes.[19] They created three model structures to describe potential progressions through disease.[19] Each model features forward-only progression from asymptomatic disease, to symptomatic disease (split by care seeking or smear status in models 2 and 3) to notification.[19] Self cure is possible from each state, effectively removing individuals from the model system, but regression followed by progression is not possible. Depending on the setting, they find the mean duration of disease to vary from less than one year to up to three years. Emery *et al.* have followed up this study with an analysis on the likely infectiousness of subclinical disease as compared to clinical disease, by using the durations of disease reported by Ku *et al.*.[19,59] Due to variations in the durations of disease, similarly wide variations were found in the estimates of infectiousness (a 95% prediction interval of 0.62 - 6.18) and in reality, values much over 1 seem highly unlikely.[59] However, with the lower bound of the prediction interval being much greater than one, it is another piece of evidence towards the importance of considering the infectiousness of subclinical disease.[59–61]

Ryckman *et al.* have created a model that combines historical and contemporary data to understand the natural history of TB. The historical data used is that of mortality presented by Tiemersma *et al.*, and adjusted by Ragonnet *et al.*, and the contemporary data are from prevalence surveys where bacteriological data were split by smear status.[2,8,20] They created a two dimensional model structure with both transitions between smear-positive and smear-negative disease, as well as transitions between symptomatic and asymptomatic disease. By using the Tiemersma and Ragonnet data on mortality, the assumption that "open" TB is equivalent to smear-positive TB is propagated further.[2,8,20] As there are very limited data to inform transitions between these states, they chose to restrict the parameter space to make the

model identifiable by matching relative risks. This has created a parameterisation dependent on assumptions rather than data. This can be seen most obviously in the posterior distributions for symptom progression and regression and spontaneous resolution, where the posterior distributions are sharply cut off on the upper bound, due to restrictions from the prior. The limited data can also lead to circular arguments within the fitting process; to ensure that the difference in mortality between smear-negative and smear-positive disease is observed, smear status must be persistent, and so they report in their findings that smear status is persistent. Both this and the prior assumptions drive the fit that ultimately says that recovery from smear-negative, subclinical disease is very fast whereas regression from smear-positive disease is much slower. Ultimately, although they used a complex fitting mechanism and combined data from historical and contemporary sources, the analysis is flawed due to an over-reliance on unjustifiable assumptions.

The final implementation of the spectrum of disease is in an analysis of the impact of vaccine strategies for vaccines currently in development.[62] This implementation of the spectrum is much more simplistic but is an example of the potential in new model structures. Clark *et al.* have maintained the widely used structure of fast and slow latent states, but then split disease into subclinical and clinical. Ultimately, the model structure is very similar to the full structure I presented in Chapter 6. The biggest difference is the lack of regression within disease states. However, their definition of recovery is different to the one I have used, as they allow for reactivation of disease from a recovered state. The prior used for progression from subclinical to clinical disease was a uniform distribution between 0 and 1 per year, which encompasses my final estimate for that progression.

7.4 Open gaps and future work

By design, the systematic review (Chapter 3), and the subsequent modelling (Chapter 4 and Chapter 6) focussed on pulmonary TB in adults. This leaves gaps in understanding the natural history of pulmonary TB in children and extra-pulmonary TB in all ages. There is literature on the development of TB in children as seen from the eight studies excluded at full text Chapter 3 as well as the previous review on intra-thoracic TB in children.[1] However as pulmonary TB is the most commonly observed form in adults, the limited data found in the review suggests that the likelihood of finding similar amounts of high quality data on each different presentation of TB is slim.[63] Whilst treatment of adults with extra-pulmonary TB and children with all forms of TB is important, these forms of disease are not believed to be key drivers of transmission, and so understanding the natural history for that purpose is less important.[4]

As discussed previously, drug-resistance and HIV present a further layer of complexity that was not possible to incorporate within my analyses. As neither were recognised before treatment was available, a similar review to find data on them would likely come up with no data on untreated progression. For HIV, it is plausible that there could be studies following progression from infection to disease and minimal to subclinical or clinical, where treatment is given once that progression has been made, (like some of the studies in the systematic review), however, searching for these was outside the scope of my review.[35,64] From the difference between my baseline parameterisation, and the suggested parameterisation from those data, estimates on the change in progression could be made. However, this is a complicated analysis, as HIV status has been split into up to 11 levels in mathematical models.[41] This would not help with understanding regression of disease without treatment. For drug-resistance, there may be studies where there are no further treatments to try, and so observational studies of progress after treatment could be used, but it is unlikely that treatment will have stopped, and more likely that an individualised approach to treating the TB would have been started, making each person's trajectory through disease incomparable.[65]

Whilst all the gaps that this work did not fill are starting points for future work, the gaps that this work did fill provide pathways to further work, either as an extension of work presented here or as new work that can utilise the estimates and results for new purposes.

Following the initial successful trials of TB chemotherapy, starting in the 1940s, came the concept of treating people before they became ill from TB.[37,66] Viewing this choice from the previous dichotomised model of *M.tb* infection followed by TB disease would suggest that these preventative therapy trials were providing treatment to prevent progression from infection or even initial infection. However, with the model of TB I have proposed, the potential treatment situation could be very different. In one of the studies, based in remote villages in Alaska, investigators expected that a year of isoniazid treatment would just delay progression to, rather than prevent, active disease.[67] However, in the following year there were fewer cases than expected, and the difference between treatment and control arms were large enough, that a few years later, everyone within the villages, regardless of original randomisation, was offered another one-year course of isoniazid.[68] These villages had a large number of people with suspected tuberculosis based on x-rays, although potentially deemed not worthy of treatment due to perceived inactivity, but with no local medical service, it was likely that outside of mass screening opportunities, only those with symptoms were offered full treatment for disease.[31,32] Therefore, it seems likely that people with minimal or even subclinical disease were offered prophylaxis. It is possible that this then treated these people, reduced the prevalent disease and transmission, and thus the unexpected effect was from a combination of these and the

prevention of progression from infection and prevention of infection. With my new model structure combined with the understanding on relative transmission by Emery *et al.*, it should be possible to simulate this trial with a model to start understanding the true effect of the prophylaxis trial.[59]

The screening model, presented in Chapter 5, can be turned into an open-source tool for estimating the impact of screening tools in new locations. The current assumptions and set up allow the model to produce estimates quickly, which makes it ideal for use as a simple tool to establish the likely impact. A further extension that can allow for a baseline of changing prevalence would be useful as most locations where community level interventions would be considered are not in a steady state. This would likely increase computation time, so either a new background code set-up would be required, or the use as a quick tool would be lost. Although the changing baseline prevalence may not be required for a comparison of likely impacts, understanding how those impacts change when the baseline is also changing would be a useful analysis to then pair up with the quicker version of the tool.

Without any adaption of the tool, it can alternatively be used to find the optimal threshold for screening tests. Given the ROC curve of diagnostic performance, different thresholds can be selected, and given a choice of which metrics to optimise, the best threshold can be selected. This can be seen in Chapter 5 where I compare two thresholds of CORTIS.[26] While RISK11 at the 60% threshold had a lower effect on prevalence and transmission, there were fewer than half as many people incorrectly treated as there were with RISK11 at the 26% threshold.

7.5 Implications for research and TB policies

The focus of TB policy has already shifted from smear status to bacteriological and symptom status, however, there are results from my work that could be incorporated to better inform policies.

7.5.1 Implications for modelling

As discussed in Chapter 2, many of the sources of data to parameterise TB models link back to one of four reviews.[7,67,69–71] Each of these reviews have overlapping studies included, that have selected from the literature, without a systematic search.[67,69–71] They also define disease poorly, and report infection from the point of testing rather than the point of conversion.[67,69–71] My systematic review in Chapter 3 has searched the literature for high quality data on progression through disease, and I have used these data in Chapter 4 and Chapter 6 to estimate progression through disease and from infection. Differentiating between subclinical and clinical disease is import when considering screening, acknowledging the rate of progression from minimal to

subclinical disease is important to accurately describe the incidence, new and returning, into infectious disease. For scenarios where it is important to consider all three, my model, and the data-based parameters, can form a framework on which to build a model. For models that need more simplification, the data collected in Chapter 3 can provide a source to differently parameterise, sourced systematically rather than the selection of studies reported in the four reviews.[67,69–71]

7.5.2 TB Burden Estimation

Estimates of incidence in 29 high burden countries are based on prevalence surveys and estimates of the duration of disease.[72] In 2019, these 29 countries were estimated to account for 66% of all global incident cases.[72] There are two approaches used to estimate the incident from prevalence surveys.

The first approximates incidence as prevalence divided by the mean duration of disease.[72] The mean duration of untreated disease used in these estimates is based on the two studies found in the review by Tiemersma *et al.* which estimated the mean duration of disease at 3 years. My work estimated a mean duration of closer to 1.5 years which would, at face value, double the estimated incidence of untreated disease.

The second method uses a simplified compartmental model of Susceptible, Untreated, and Treated.[72] This method finds durations closer to that estimated by both my work and that by Ku *et al.*.[19,72] Although the model structure does not differentiate between disease states, and so there is no need for progression and regression between disease states, it does include regression from disease.[72] The combined estimate of the proportion who recover from disease or who die before treatment is given a uniform distribution between 0 and 0.05. It is not clear whether this recovery is complete recovery beyond re-progression, or regression of symptoms or of bacteriology. Without treatment my model estimates about 5% of people recover from subclinical disease within the first year, with another 30% who recover to minimal disease. This proportion will reduce with treatment, but the people who regress the quickest are those who are least likely to receive treatment. Therefore, particularly if recovery from disease is to minimal disease, an estimate that up to 5% die or recover seems low. If it is true recovery, it seems a more reasonable assumption.

Although a strong active case finding programme may be able to find a high proportion of people with clinical disease and even subclinical disease, most places do not have active, symptom-agnostic screening processes, nor the resource capacity to set one up. I have estimated that 50% of people with subclinical disease may never progress to clinical disease,

and without a strong active case-finding programme, these cases are likely missed, and so the proportion who recover without treatment may well be higher.

Updating both of these methods of burden estimation to reflect the emerging understandings of disease progression across the spectrum may lead to better informed estimates of incidence. There is also potential to incorporate the spectrum of disease and estimates of incidence into each disease state. This leads to questions in how regression and re-progression are counted for incidence calculations.

7.5.3 Disease detection

As highlighted throughout, and in particular in Chapter 5 and Chapter 6, the sensitivities of screening tests are defined against sputum culture positive clinical disease, albeit in different populations. It is unlikely that a screening tool will perform as well throughout the entire spectrum of disease (e.g. culture based tests only detect bacteriologically positive disease, blood biomarkers are significantly better at detecting clinical disease than subclinical disease) but there are not good data on how most tests perform at different points across the spectrum.[26,73] However, there is still no confirmatory test that detects either minimal disease or M.tb infection. Without this gold standard test to compare screening tests against, the performance of screening tests for different TB disease phenotypes cannot accurately be determined. Therefore, before trying to determine how screening tests perform, particularly in minimal disease, more research is required to understand the state, its implications for long lasting health, and what could serve as a gold standard for detection.

There is a balance to be found between the personal effect on the individuals of recovering without treatment compared to undergoing the long treatment regimens with the attached stigma along with the population level effects on transmission. If there is extensive lung damage from minimal and subclinical disease that could be reversed by earlier treatment, then it is important to treat people to reduce morbidity for those individuals. Similarly, if a large proportion of transmission comes from people with long term subclinical disease that never progresses to clinical disease before resolving, then at a population level, treatment is important. At a population level, I have estimated that treating subclinical disease has a large impact on transmission, when comparing the reductions in transmission from symptom screening and radiographic screening in Chapter 5. However, the same analysis found a much smaller difference in transmission when also treating minimal disease (comparing the two TPPs). Therefore, the balance is likely for treatment for subclinical disease but depends on the burden of treatment and stigma compared to the potential long term lung damage for minimal disease. Until the mid-1960s, many countries were using mass radiography in what may well

have been an effort to detect as many people with both clinical and subclinical disease as possible, and a return to this may be a solution to reducing prevalence.

7.6 Conclusions

Subclinical TB disease makes up a large proportion of all prevalent infectious TB disease.[6,35] It is not a given that subclinical disease progresses to clinical disease. Therefore, since subclinical disease likely does contribute to a large proportion of transmission, it is important to develop and deploy case-finding strategies that can detect disease without waiting for the development of symptoms.[59–61] Most novel screening tools (blood-biomarkers and CRP) available at the moment do not meet the WHO TPP sensitivity and specificity targets and so the development of a cheap, rapid screening test could help inform which people should be referred for further testing and treatment. However, pending the development of such a tool, a return to repeated mass x-ray screening or similar in high-risk populations, with treatment available to all, may help reduce the prevalence of disease significantly without requiring population level Xpert testing. Developing a test for viable *M.tb* infection and increasing understanding of minimal disease could help target preventative therapies to those who will progress, or have progressed beyond infection.

7.7 References

1. Marais, B. J. *et al.* The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era. 11 (2004).

2. Tiemersma, E. W., van der Werf, M. J., Borgdorff, M. W., Williams, B. G. & Nagelkerke, N. J. D. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PloS One* **6**, e17601 (2011).

3. World Health Organization. *Global tuberculosis control: WHO report 2010*. https://apps.who.int/iris/handle/10665/44425 (2010).

4. World Health Organization. Global tuberculosis report 2021. (World Health Organization, 2021).

5. Onozaki, I. *et al.* National tuberculosis prevalence surveys in asia, 1990-2012: An overview of results and lessons learned. *Trop Med Int Health* **20**, 1128–1145 (2015).

6. Frascella, B. *et al.* Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. *Clinical Infectious Diseases* **73**, e830–e841 (2021).

7. 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* **18**, e228–e238 (2018).

8. Ragonnet, R. *et al.* Revisiting the Natural History of Pulmonary Tuberculosis: A Bayesian Estimation of Natural Recovery and Mortality Rates. *Clinical Infectious Diseases* **73**, e88–e96 (2021).

9. Szucs, E. Sputum Examination in So-called Closed Tuberculosis. *JAMA: The Journal of the American Medical Association* **86**, 946 (1926).

10. Blahd, M., Leslie, E. I. & Rosenthal, S. R. Infectiousness of the "closed case" in tuberculosis. *American Journal of Public Health and the Nations Health* **36**, 723–726 (1946).

11. Tattersall, W. H. The Survival of Sputum-Positive Consumptives: A study of 1,192 cases in a county borough between 1914 and 1940. *Tubercle* **28**, 85–96 (1947).

12. Krebs, W. Die fälle von lungentuberkulose in der aargauischen heilstätte barmelweid aus den jahren 1912-1927. *Beitr Klin Tuberk* **74**, 345–379 (1930).

13. Rathbun, W. L. The classification of pulmonary tuberculosis. *American Review of Tuberculosis* **1**, 1–15 (1917).

14. Pamra, S. P., Goyal, S. S. & Mathur, G. P. Changes in prevalence and incidence of pulmonary tuberculosis in Delhi in recent years. *Indian Journal of Tuberculosis* 57–64 (1973).

15. National Tuberculosis Institute, B. Tuberculosis in a rural population of South India: a five-year epidemiological study. *Bulletin of the World Health Organization* **51**, 473–88 (1974).

16. Armstrong, D. B. Four years of the framingham demonstration. *American Review of Tuberculosis* **4**, 908–919 (1921).

17. Mungai, B. N. *et al.* 'If not TB, what could it be?' Chest X-ray findings from the 2016 Kenya Tuberculosis Prevalence Survey. *Thorax* **76**, 607–614 (2021).

18. Murray, L. M. Bayesian state-space modelling on high-performance hardware using LibBi. *arXiv:1306.3277 [stat]* (2013).

19. Ku, C.-C. *et al.* Durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease with a Bayesian analysis of prevalence survey and notification data. *BMC Medicine* **19**, 298 (2021).

20. Ryckman, T. S., Dowdy, D. W. & Kendall, E. A. Infectious and Clinical Tuberculosis Trajectories: Bayesian modeling with case finding implications. doi:10.1101/2022.06.27.22276965.

21. World Health Organization. *WHO expert committee on tuberculosis - eighth report.* https://apps.who.int/iris/bitstream/handle/10665/40606/WHO_TRS_290.pdf?sequence=1&isAllowed=y (1964).

22. World Health Organization. *WHO expert committee on tuberculosis - ninth report.* https://apps.who.int/iris/bitstream/handle/10665/41095/WHO_TRS_552_eng.pdf?sequence=1&isAllowed=y (1974).

23. Marks, G. B. *et al.* Community-wide Screening for Tuberculosis in a High-Prevalence Setting. *The New England Journal of Medicine* **381**, 1347–1357 (2019).

24. Linh, N. N., Miller, C. & Marks, G. B. Potential benefits of active case finding to reduce the burden of TB. *The International Journal of Tuberculosis and Lung Disease* **25**, 93–94 (2021).

25. Burke, R. M. *et al.* Community-based active case-finding interventions for tuberculosis: A systematic review. *The Lancet Public Health* **6**, e283–e299 (2021).

26. Scriba, T. J. *et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): A randomised controlled trial. *The Lancet Infectious Diseases* **0**, (2021).

27. Yoon, C., Davis, J. L. & Cattamanchi, A. C-reactive protein and tuberculosis screening: a new trick for an old dog? *Int J Tuberc Lung Dis* **17**, 1656 (2013).

28. Mungai, B. *et al. Accuracy of computer-aided chest X-ray screening in the Kenya National Tuberculosis Prevalence Survey*. http://medrxiv.org/lookup/doi/10.1101/2021.10.21.21265321 (2021) doi:10.1101/2021.10.21.21265321.

29. World Health Organization. WHO consolidated guidelines on tuberculosis Module 2: Screening Systematic screening for tuberculosis disease. (World Health Organization, 2021).

30. Pai, M. *et al.* Tuberculosis. *Nat Rev Dis Primers* **2**, 16076 (2016).

31. Comstock, G. W. Isoniazid prophylaxis in an undeveloped area. *The American Review of Respiratory Disease* **86**, 810–822 (1962).

32. Comstock, G. W., Ferebee, S. H. & Hammes, L. M. A controlled trial of community-wide isoniazid prophylaxis in Alaska. *The American Review of Respiratory Disease* **95**, 935–943 (1967).

33. WHO. *Systematic screening for active tuberculosis: Principles and recommendations*. (2013).

34. Esmail, H., Dodd, P. J. & Houben, R. M. G. J. Tuberculosis transmission during the subclinical period: Could unrelated cough play a part? *The Lancet Respiratory Medicine* **6**, 244–246 (2018).

35. Okada, K. *et al.* Epidemiological impact of mass tuberculosis screening: A 2 year follow-up after a national prevalence survey. **NA**, (2012).

36. Monto, A. S. & Sullivan, K. M. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiology and Infection* **110**, 145–160 (1993).

37. Fox, W., Ellard, G. A. & Mitchison, D. A. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease* **3**, S231–279 (1999).

38. Sharp, P. M. & Hahn, B. H. The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 2487–2494 (2010).

39. Ricks, S., Denkinger, C. M., Schumacher, S. G., Hallett, T. B. & Arinaminpathy, N. The potential impact of urine-LAM diagnostics on tuberculosis incidence and mortality: A modelling analysis. *PLOS Medicine* **17**, e1003466 (2020).

40. Blower, S. M., Small, P. M. & Hopewell, P. C. Control Strategies for Tuberculosis Epidemics: New Models for Old Problems. *Science* **273**, 497–500 (1996).

41. Houben, R. M. G. J. *et al.* TIME impact a new user-friendly tuberculosis (TB) model to inform TB policy decisions. *BMC Medicine* **14**, (2016).

42. Yoon, C. *et al.* Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *The Lancet Infectious Diseases* **17**, 1285–1292 (2017).

43. Yoon, C. *et al.* Yield and efficiency of novel intensified tuberculosis case-finding algorithms for people living with HIV. *American Journal of Respiratory and Critical Care Medicine* **199**, 643–650 (2018).

44. Kendall, E. A. *et al.* The spectrum of tuberculosis disease in an urban ugandan community and its health facilities. *Clinical Infectious Diseases* **72**, e1035–e1043 (2021).

45. Meyer, A. J. *et al.* C-reactive protein testing for active tuberculosis among inpatients without HIV in uganda: A diagnostic accuracy study. *Journal of Clinical Microbiology* **59**, e02162–20 (2020).

46. Gothi, G. D. Natural History of Tuberculosis. *Indian Journal of Tuberculosis* **25**, (1977).

47. Barry, C. E. *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Reviews. Microbiology* **7**, 845–855 (2009).

48. Lin, P. L. & Flynn, J. L. The End of the Binary Era: Revisiting the Spectrum of Tuberculosis. *The Journal of Immunology* **201**, 2541–2548 (2018).

49. Houben, R. M. G. J. *et al.* Spotting the old foe — revisiting the case definition for TB. *The Lancet Respiratory Medicine* **7**, 199–201 (2019).

50. Sossen, B. *et al.* The Natural History of Untreated Pulmonary Tuberculosis in Adults: A Systematic Review and Meta-Analysis. doi:10.1101/2022.08.30.22279374.

51. Waaler, H. T. A dynamic model for the epidemiology of tuberculosis. *American Review of Respiratory Disease* **98**, 591–600 (1968).

52. Houben, R. M. G. J. & Dodd, P. J. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS medicine* **13**, e1002152 (2016).

53. Behr, M. A., Edelstein, P. H. & Ramakrishnan, L. Revisiting the timetable of tuberculosis. *The BMJ* **362**, (2018).

54. Behr, M. A., Edelstein, P. H. & Ramakrishnan, L. Is Mycobacterium tuberculosis infection life long? *BMJ* **367**, l5770 (2019).

55. Emery, J. C. *et al.* Self-clearance of mycobacterium tuberculosis infection: Implications for lifetime risk and population at-risk of tuberculosis disease. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20201635 (2021).

56. Lillebaek, T. *et al.* Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection. *The Journal of Infectious Diseases* **185**, 401–404 (2002).

57. Byng-Maddick, R. & Noursadeghi, M. Does tuberculosis threaten our ageing populations? *BMC Infectious Diseases* **16**, 119 (2016).

58. Ankrah, A. O. *et al.* PET/CT imaging of mycobacterium tuberculosis infection. *Clinical and Translational Imaging* **4**, 131–144 (2016).

59. Emery, J. C. *et al.* Estimating the contribution of subclinical tuberculosis disease to transmission an individual patient data analysis from prevalence surveys. doi:10.1101/2022.06.09.22276188.

60. Williams, C. M. *et al.* Exhaled mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: Prospective observational studies. *The Lancet Infectious Diseases* **20**, 607–617 (2020).

61. Dinkele, R. *et al.* Aerosolization of mycobacterium tuberculosis by tidal breathing. *American Journal of Respiratory and Critical Care Medicine* **206**, 206–216 (2022).

62. Clark, R. A. *et al.* The impact of alternative delivery strategies for novel tuberculosis vaccines in low- and middle-income countries: a modelling study. doi:10.1101/2022.04.16.22273762.

63. World Health Organization. *Global tuberculosis report 2020*. https://www.who.int/publications-detail-redirect/9789240013131 (2020).

64. Norregaard, J. Abacillary pulmonary tuberculosis. *Tubercle* **71**, 35–38 (1990).

65. World Health Organization. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. (World Health Organization, 2014).

66. Ferebee, S. H., Mount, F. W. & Palmer, C. E. Tuberculosis Prophylaxis Trials in Preview. *Public Health Reports* **72**, 703–704 (1957).

67. Ferebee, S. H. Controlled chemoprophylaxis trials in tuberculosis. A general review. *Bibl Tuberc* **26**, 28–106 (1970).

68. Hanson, M. L., Comstock, G. W. & Haley, C. E. Community isoniazid prophylaxis program in an underdeveloped area of alaska. *Public Health Reports* **82**, 1045–1056 (1967).

69. Comstock, G. W. Epidemiology of tuberculosis. *The American Review of Respiratory Disease* **125**, 8–15 (1982).

70. Styblo, K. in *Advances in Respiratory Medicine* (eds. Flenley, D. C. & Petty, T. L.) 77–108 (Churchill Livingstone, Edinburgh, 1986).

71. Styblo, K. Selected papers (vol 24). *KNCV Tuberculosis Foundation - Selected papers* (1991).

72. Glaziou, P., Dodd, P. J., Dean, A. & Floyd, K. Methods used by WHO to estimate the global burden of TB disease. (2020).

73. Zifodya, J. S. *et al.* Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. *The Cochrane Database of Systematic Reviews* **2**, CD009593 (2021).

Appendix A Official write up of Chapter 3

A1 Cover Sheet

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

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

SECTION A – Student Details

Student ID Number	1700322	**Title**	Miss
First Name(s)	Alexandra S		
Surname/Family Name	Richards		
Thesis Title	Rediscovering the natural history of tuberculosis using modelling to combine historical and contemporary data		
Primary Supervisor	Rein Houben		

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

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

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

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	The Lancet Respiratory Medicine
Please list the paper's authors in the intended authorship order:	Bianca Sossen, Alexandra Richards, Torben Heinsohn, Beatrice Frascella, Federica Balzarini, Aurea Oradini-Alacreu, Anna Odone, Ewelina Rogozińska, Brit Häcker, Frank Cobelens, Katharina Kranzer, Rein MGJ Houben, Hanif Esmail

Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	This work was originally conceived by Hanif Esmail and Rein Houben. I helped develop the aims and the protocol with Bianca Sossen, Hanif Esmail, and Rein Houben. I have done the largest share of the work at all stages of the review, screening the titles, finding and reviewing the full texts, assessing bias and extracting data. This work has been shared with the original group and through a team of eight other people who I guided and managed; three helped source the full texts and search the index medicus, five assisted with the full text screening, bias screening, and data extraction. Every study that has had data extracted, I either extracted myself or approved the extraction by someone else. I prepared the figures for the meta-analyses.

SECTION E

Student Signature	
Date	19-09-2022

Supervisor Signature	Rein Houben 
Date	19-Sep-22

Authors

Bianca Sossen (MBChB)[1,2,#], Alexandra Richards (MMathPhys)[3,4,#], Torben Heinsohn (BMBCh)[2,5], Beatrice Frascella (MD)[6], Federica Balzarini (MD)[6], Aurea Oradini-Alacreu (MD)[6], Prof. Anna Odone (PhD)[7], Ewelina Rogozinska (PhD)[8], Brit Häcker (Dr)[9], Prof. Frank Cobelens (PhD)[10,11], Prof. Katharina Kranzer (PhD)[12-14], Prof. Rein MGJ Houben (PhD)[3,4;&], Hanif Esmail (PhD)[2,8,15;&,¥]

[1] Department of Medicine, Faculty of Health Sciences, University of Cape Town, South Africa

[2] Institute for Global Health, University College London, London, UK

[3] TB Modelling Group, TB Centre, London School of Hygiene and Tropical Medicine, London, UK

[4] Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, London, UK

[5] Helmholtz Centre for Infection Research, Department of Epidemiology

[6] School of Public Health, Vita-Salute San Raffaele University, Milan, Italy

[7] Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia, Italy

[8] MRC Clinical Trials Unit at University College London, UK

[9] German Central Committee against Tuberculosis (DZK), Berlin, German

[10] Amsterdam University Medical Centers location University of Amsterdam, Department of Global Health

[11] Amsterdam Institute for Global Health and Development, Amsterdam, The Netherlands

[12] Clinical Research Department, Faculty of Infectious and Tropical Disease, London School of Hygiene and

Tropical Medicine, London, UK

[13] Biomedical Research and Training Institute, Harare, Zimbabwe

[14] Division of Infectious Diseases and Tropical Medicine, University Hospital, LMU Munich, Munich,

Germany

[15] Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Diseases and

Molecular Medicine, University of Cape Town, South Africa

# Contributed Equally

& Contributed Equally

¥ Corresponding Author: Dr Hanif Esmail at MRC Clinical Trials Unit, 90 High Holborn, London, WC1V 6LJ or h.esmail@ucl.ac.uk

SUMMARY

BACKGROUND

Key stages in TB disease can be delineated by radiology, microbiology and symptoms, but transition between relevant stages remains unclear. We sought to quantify progression and regression across the spectrum of TB disease by systematically reviewing studies of individuals with untreated TB undergoing follow up.

METHODS

We searched PubMED, EMBASE and Web of Science until December 31st 1960, the Index Medicus between 1895 and 1945, and extensive investigator collections without date restriction - in English and German. Eligible studies were observational cohorts and clinical trials, presenting adults/adolescents with TB or recent TB exposure, undergoing follow-up for at least 12 months without therapeutic intervention. Two authors independently reviewed titles/abstracts and full texts for inclusion. Quality was assessed with a modified Newcastle-Ottawa Score, excluding highly biased studies. Summary estimates were extracted to align with TB disease transitions in a conceptual model, and we used meta-analysis of proportions with random-effects to synthesise the extracted data. This study is registered with PROSPERO (CRD42019152585).

FINDINGS

10477 titles were screened and 1648 full texts reviewed. 223 met inclusion criteria. 109 were excluded for high risk of bias and 90 did not have extractable data. 24 studies (34 cohorts) were included. Progression from microbiologically negative to positive disease in those with radiographic TB evidence occurred at an annualized rate of 9.71% (95% CI:6.17-13.34) with "active" TB imaging, and 1.06% (95% CI:0.31-1.82) with "inactive" TB imaging. Reversion from microbiologically-positive to -undetectable in prospective cohorts occurred at an annualized rate of 12.40% (95% CI: 6.81-17.99). Studies reported symptoms poorly not allowing for direct estimation of transitions for subclinical (asymptomatic, culture positive) disease.

INTERPRETATION

We present the risk of progression in those with radiographic evidence of disease and the rate of self-cure for microbiologically positive disease to inform global disease burden estimates, clinical guidelines and policy decisions.

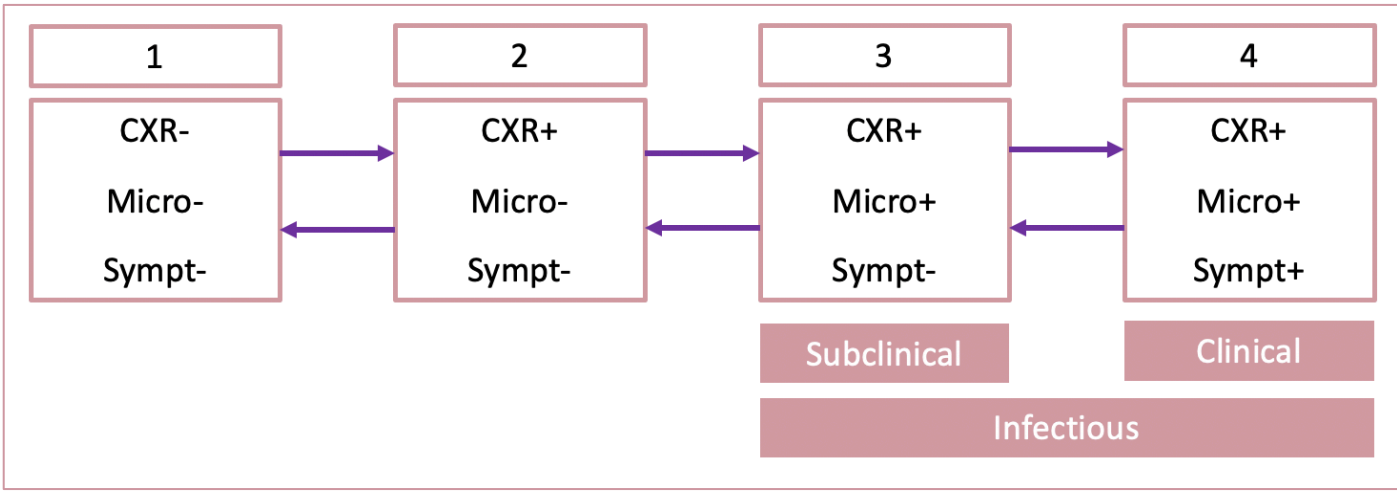
INTRODUCTION

Despite a clinical awareness of tuberculosis (TB) for centuries, its natural history is incompletely understood. We have oscillated between characterizing TB with binary states of latent infection and active disease, to a condition existing on a dynamic continuum(1–4). In the early 20th century, TB control relied on early identification of those with evidence of disease, particularly through chest X-ray (CXR) screening. Researchers were able to highlight the heterogeneity and dynamics of disease evolution between individuals, through longitudinal assessment(5–8). With the discovery of effective treatment in the mid-20th century and driven by the need for scalable, programmatic treatment algorithms, a binary description of disease states reflecting two extremes ('latent infection' and 'active disease') became established(9). Although this provided a useful paradigm, the more nuanced understanding of disease natural history was arguably forgotten.

An accurate understanding of the kinetics of TB natural history is now increasingly critical at both population and individual level, with implications for disease management, population-level prevention and control, and disease burden estimations. Treatment of patients that fall between active and latent TB - for instance having abnormalities suggestive of active disease on X-ray but microbiologically negative - is not adequately covered by management algorithms, but progress could be driven by adequate understanding of the risk of disease progression. A better understanding of this natural history is also a key priority for vaccine development(10). In addition, estimates of TB burden currently rely strongly on assumptions around the progression, regression and mortality from untreated TB, of which only mortality estimates are informed by systematic review of available literature(11–13). Furthermore, estimation methods do not cater to different stages of TB which are detected in disease prevalence surveys, including individuals who have culture positive disease but a negative symptom screen (referred to as subclinical), or those with TB suggestive X-rays(14). Given the implications for health care seeking and potential

for interrupting or preventing transmission, a better understanding of this natural history is key to inform TB burden estimation and policies for care and prevention.

Within the disease continuum, key stages in the evolution of pulmonary TB can be marked by diagnostic tests that have been available for over a century, to allow for categorization within a widely accepted conceptual framework (Figure 1)(1,2). The emergence of disease pathology is first visible by typical radiographic features. Microbiological detection in sputum signals presence of bacilli (and potential infectiousness), and the reporting of symptoms marks the development of active, clinical disease. Transitions across all of these stages can only be fully studied in the absence of treatment and hence can no longer be ethically investigated. We conducted a systematic review focusing on articles from the pre-chemotherapy era to determine which of the transitions could be adequately described by existing literature, with the aim of providing parameters for the rate of progression and regression of disease across the spectrum.



Panel: Research in context

Evidence before this study

Certain aspects of the natural history of TB in the absence of treatment have been previously estimated through systematic review of the literature. Several authors have assessed the progression of latent infection (defined by TST or Interferon-gamma releasing assay (IGRA)) to active disease (usually defined by bacteriological positivity). However, these studies took a binary approach to categorizing TB, not recognizing the spectrum of disease states. Tiemersma *et al* took a systematic approach to determine the mortality of untreated TB disease distinguishing the outcome of 'open' and 'closed' disease, interpreted as smear-positive and smear-negative disease respectively. However, they did not focus on the dynamics of progression and regression of early stages of disease. Marais *et al* have investigated the natural history of TB in childhood, focusing on age specific rates and nature of disease progression following primary infection in children. Key gaps in our understanding of disease natural history therefore remain to be systematically assessed. Most notably, in individuals with changes consistent with early TB on CXR (as found on national prevalence surveys), the risk of progression to bacteriologically positive disease has not been sufficiently summarized, nor do we know the progression and regression from bacteriologically positive, symptom screen negative disease, or adequately understand regression (or 'self-cure) from classic 'active disease', i.e. bacteriologically positive, symptom screen positive.

Added value of this study

This systematic review has used the enormous body of historical literature to better capture progression and regression across the spectrum of TB disease stages: defined

by chest imaging, sputum microbiology and symptom status. We identified 34 cohorts with a combined sample of 139,212 participants that contributed to our analysis. We show progression to bacteriologically positive disease of 10%/year from individuals with CXRs suggestive of active TB, and quantify reversion (self-cure) from microbiologically-positive disease, but could not directly quantify progression into and regression from subclinical disease

Implications of all the available evidence

With the currently available and newly generated evidence, we have a better understanding of the kinetics of TB natural history. High progression from CXR positive disease calls for a reconsideration of treatment guidelines. In addition, the comprehensive data collected could enable modelling to inform kinetics around subclinical disease as well. Finally, this improved understanding will allow for efforts to be refined when attempting to study and improve diagnostic and prognostic biomarkers, and likely allow for greater benefit to be derived from targeted TB prevention and treatment efforts.

METHODS

Search strategy and selection criteria

This systematic review and meta-analysis was conducted following a protocol registered at PROSPERO (CRD42019152585). The study is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRIMSA) guidelines(17). We searched for articles from the pre-chemotherapy era combining electronic and manual searches. Electronic searches were conducted in Medline (via PubMED), EMBASE and Web of Science from the start of the database (1946, 1947, and 1900 respectively) until 31st December 1960, in two languages with high yield for study designs of interest in this period: English and German. Additionally, we manually searched titles from Index Medicus between 1903 and 1945; volumes from 1895-1902 were not available. The systematic search was restricted to manuscripts published prior to 1960 to include cohorts observed from the pre-chemotherapy era while allowing for a publication delay of earlier cohorts. Furthermore, supplementary searches were conducted in extensive author collections. Further references were snowballed from those articles that met the criteria for data extraction and from key review articles. Personal libraries and snowballed references were searched without date restriction.

Electronic search terms used both modern and historical terminology in English and German (full search strategies in supplementary appendix). All titles were imported into Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia). After de-duplication, titles and abstracts were screened for relevance by two independent reviewers, with a third reviewer resolving conflicts. Full text articles were sought online, within the library stores at the Wellcome and British libraries (English articles) and the library of the German Central Committee against Tuberculosis (DZK) and the German Tuberculosis Archive (DTA) (German articles), and on online archive websites (e.g HathiTrust.org and archive.org). If manuscripts could not be found through any of these sources, they were not included. At full-text stage, two independent reviewers

reviewed eligibility. Articles were included if they presented a longitudinal cohort of at least 25 adolescents (≥10 years) and/or adults followed up (radiologically, microbiologically and clinically) for at least 12 months from the point of either (1) positive TST following recent TB exposure, (2) radiographic changes suggestive of TB or (3) positive microbiology for TB (smear microscopy and/or mycobacterial culture). A minimum of 12 months was selected in order to ensure an adequate number of events. Articles were excluded if they made no attempt at microbiological confirmation of disease, presented no new data (i.e. review article), all participants received a therapeutic (medical or surgical) intervention or those who did not receive a therapeutic intervention could not have data extracted separately, or where ≥5% of the cohort were paediatric (<10 years) and these children could not be separated from the adolescent/adult data.

Eligible articles were assessed for risk of bias with an adapted Newcastle-Ottawa Scale (NOS) to a maximum of seven stars (NOS - General Quality Assessment) by two reviewers per language (supplementary table 1) with conflicts resolved by consensus. To pass the quality assessment, studies could only lose two stars in the "Study Selection" and "Outcome" domains of the NOS. The "comparability" domain was not assessed as this systematic review did not use control groups. An additional quality assessment tool was designed to assess the quality of specific diagnostic compartments in study cohorts i.e. radiological, microbiological and symptoms (supplementary table 1). While this Specific Quality Assessment was captured to get a sense of quality of the study designs, it did not inform study eligibility. Those that passed the NOS were extracted in a standardized electronic tool by one reviewer and then datapoints confirmed by a second reviewer with conflicts were resolved by consensus, involving input from additional reviewers if needed.

Data extraction and analysis

We extracted data corresponding to the proportion of individuals in the cohort transitioning between diagnostic states (figure 1) over a specified period of time. Recognizing that description of symptom status in particular may not always be explicit by current standards this could be recorded as unknown as long as microbiological status was clear. Where authors differentiated abnormal chest imaging that was suggestive of TB versus not suggestive then we only extracted the TB-suggestive group as abnormal. In addition, where authors provided a subgroup of abnormal chest x-rays that were limited to only calcified nodules then we did not deem these to be an abnormal x-ray for the purpose of this review. The clinical classification method of the National Tuberculosis Association Diagnostic Standards and Classification of Tuberculosis; facilitated extraction of the data(18).

Certain studies presented the proportion of individuals who progressed within a window of time rather than a specific time point; in these cases, we have presented datapoints as at the midpoint of the time window provided. All summary estimates are presented with 95% confidence intervals, calculated from the point data provided. To allow for exploration of the data and any heterogeneity, we attempted to collect data on variables of interest, namely: age distribution, sex, frequency of follow up visits, microbiological test used (i.e. culture versus smear), CXR characteristics described by the historical study's authors, TST data, local disease burden as per today's WHO classification(19), features of the study design (i.e. passive versus active versus mixed case finding and whether the data was generated from two cross-sectional assessments of participants ("single follow-up") or through a cumulative count of events over time ("cumulative count")), the enrollment setting, and symptom status.

To allow comparison of the varying follow-up times, the last data point of each study was annualised and the expected number transitioning in the first year calculated. The variance of the annualised rate was then calculated using the escalc function from the metafor package(20), specifying the raw proportion measure. Meta-analysis was then conducted using the rma function

with the study outcome and variance as inputs. By default each study was weighted proportional to the inverse of the variance calculated in the previous step. The forest plots were created using the forest function from the meta package. Confidence interval proportions were limited to between 0 and 1 by the observation limit argument within the forest function. Sub analyses were also conducted using the rma function and added to the forest plot using the addpoly function from metafor. Heterogeneity was assessed with the I^2 and tau^2 statistics.

Role of the funding source

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

RESULTS

After de-duplication a total of 10477 titles and abstracts were screened of which 8829 were deemed not relevant (figure 2). 145/1648 (8.8%) full texts could not be sourced. A further 1280 studies were deemed to meet exclusion criteria, leaving 223 for bias assessment. A high risk of bias was found in 109 studies and an additional 90 could not reliably have data extracted and therefore did not contribute to our results. In total, 22 English and two German articles, with a combined sample of 139,212 participants contributed 34 cohorts for analysis. Eight of the 24 studies scored maximal scores on the General Quality Assessment. The quality of data on symptom status was generally poor, with 10 studies scoring zero stars in the Specific Quality Assessment (table 1).

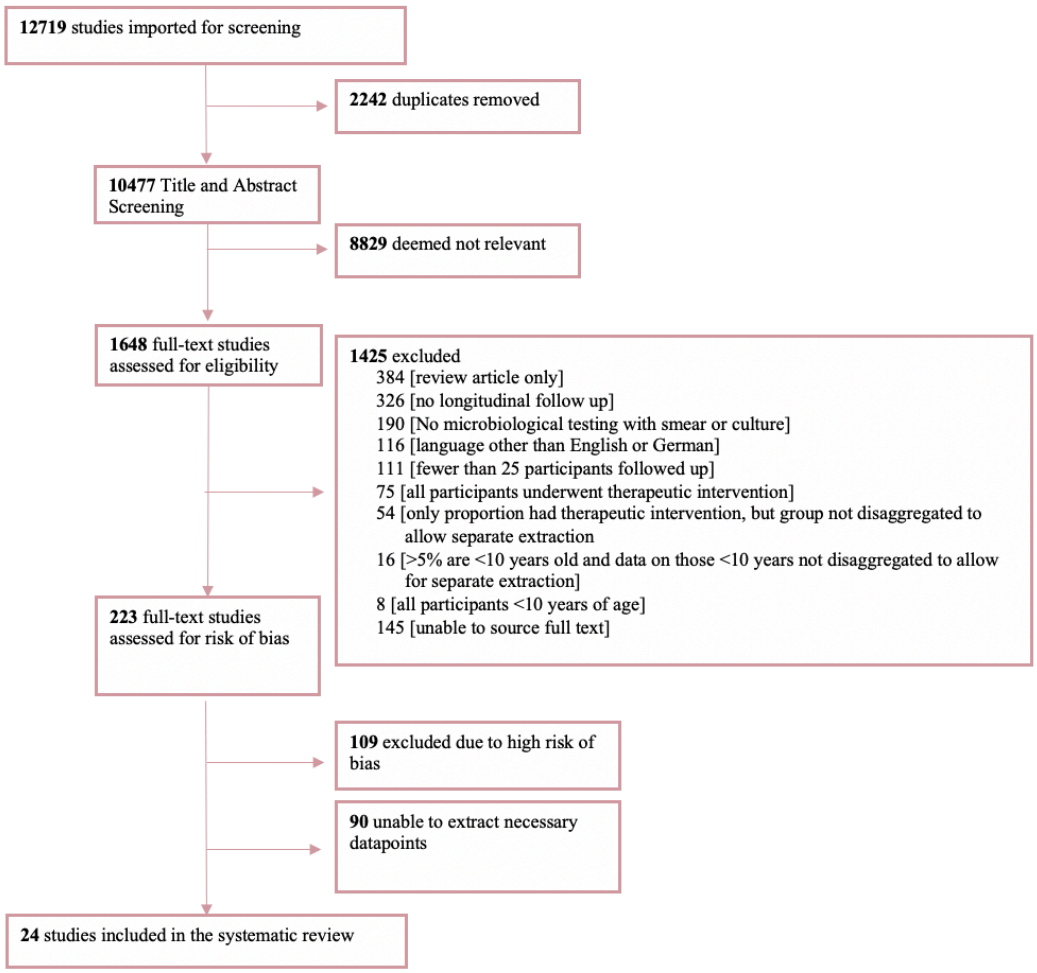


Figure 2: Study Selection

Table 1: Bias Assessments*

	General Quality Assessment†		Specific Quality Assessment		
	Study Selection‡	Outcome Assessment	Radiography	Microbiology	Symptoms
1. Alling(21)	0	0 0 0 X	0 0 0 X	0 0 X	0 X
2. Anastasatu(22)	X	0 0 0 0	0 0 X X	0 0 0	X X
3. Aneja(23)	0	0 0 0 0	0 0 0 0	0 0 0	0 0
4. Beeuwkes(24)	0	0 0 0 X	0 0 X X	0 0 X	0 X
5. Bobrowitz(25,26)	X	0 0 0 0	0 0 0 0	0 0 0	0 0
6. Borgen(27,28)	0 0 0	0 0 0 0	0 0 0 X	0 0 X	0 0
7. Breu(29)	0	0 0 0 0	0 0 0 X	0 0 0	X X
8. Cowie(30)	0	0 0 0 0	0 0 0 X	0 0 0	X X
9. Downes(31)	X	0 0 0 X	0 0 0 X	X X X	0 0
10. Frimodt-Moller(32)	0	0 0 0 0	0 0 0 0	0 0 0	X X
11. Hong Kong Chest Service(33–36)	X	0 0 0 0	0 0 0 X	0 0 0	0 0
12. IUAT Committee on Prophylaxis(37)	X	0 0 0 0	0 0 0 X	0 0 0	X X
13. Lincoln(21,38)	0	0 0 0 X	0 0 0 X	0 0 X	0 X
14. Manser(39)	X	0 0 0 X	0 0 0 X	0 0 0	0 0
15. Marshall(40)	0	0 0 0 0	0 0 0 0	0 0 0	0 0
16. National Tuberculosis Institute(41–49)	0 0 0	0 0 0 0	0 0 0 0	0 0 0	X X
17. Norregaard(50)	X	0 0 0 0	0 0 0 X	0 0 0	0 X
18. Okada(51)	0 0 X	0 0 0 0	0 0 0 0	0 0 0	0 0
19. Orrego Puelma(52)	X	0 0 0 0	0 0 0 X	0 0 0	0 X

20. Pamra(53)	0	0 0 0 0	0 0 0 X	0 0 0	X X
21. Puffer(54)	0	0 0 0 X	0 0 0 X	X X X	X X
22. Sikand(55)	0 0 0	0 0 0 X	0 0 0 X	0 0 0	X X
23. Styblo(56)	0 0 0	0 0 X X	0 0 0 X	0 0 X	0 X
24. Tuberculosis Society of Scotland(57,58)	X	0 0 0 0	0 0 0 X	0 0 0	X X

* This table only includes the quality assessments for studies included in the meta-analysis, with a 0 representing a positive score and X representing a negative score.

† Studies could only lose two stars in the General Quality Assessment to proceed to data extraction. Further details of the quality assessment tools are available in supplementary material.

‡ Studies could score a maximum of 3 stars if participants entered the cohort with no evidence of disease (chest x-ray negative, microbiologically negative and asymptomatic) or a maximum of 1 star if participants already had some evidence of disease at entry.

The setting for the 34 longitudinal cohorts were as follows: workplace or university screening (n=5), general community screening (n=7), from household contact studies (n=4), clinical cohorts at clinics or sanatoria (n=9) and control arms of therapeutic interventions (n=9) (table 2 and supplementary table 4). Cohorts were conducted in Europe (n=10), Asia (n=11), North America (n=11), Africa (n=1) and South America (n=1). Eleven of the 34 cohorts provided an estimate of the local burden of TB disease in the study setting and related time period. The majority (n=9/11) of these settings would be classified as endemic or high burden TB settings, and the remainder (n=2/11) as medium burden, based on today's WHO classification(19). Cohorts were conducted between 1923 and 2004 with 20/34 (58.8%) prior to 1960.

We did not identify any cohorts, meeting our inclusion and quality criteria, closely following up confirmed recent TST converters where transition from normal chest X-ray (CXR) to CXR suggestive of TB was reported. We identified four cohorts following up participants with normal radiography, negative microbiological testing where the timepoint of initial infection was unclear,

with either no evidence of symptoms (n=3 (75.0%)) or unrecorded symptom status (n=1 (25.0%)) (table 2). We identified 24 cohorts following-up participants with evidence of radiographic abnormalities and negative microbiology but with either no symptoms (n=8 (33.3%)), symptoms (n=3 (12.5%)) or mixed/unknown symptoms (n=13 (54.2%)) initially. Of these 24 cohorts, the radiographic abnormalities were specified by the original authors as either active (n=9 (37.5%)) or inactive/fibrotic (n=7 (33.3%)), with the remaining being mixed or not specified (n=8 (29.1%)). We identified six cohorts following participants with microbiologically detectable tuberculosis either initially with symptoms (n=4 (66.7%)) or those with an unknown symptom status (n=2 (33.3%)), however there were no cohorts found in which patients were documented to be asymptomatic.

Table 2: Extracted studies

	Cohort Years	Age distribution	Percentage Female	Local Burden of Disease; Setting	Study Type	Data Collection	Cohort Size (n)	Micro. quality (design*)	X-ray Description†	Follow up‡	Starting point§	Endpoint§
1. Alling(21)	1938 - 1948	Average: 52 years	Unknown	Not specified; USA	Clinic/Hospital / Sanatorium cohort	Retrospective	58	Unclear (passive)	Inactive	Cumulative	cxr.pos | micro.neg | sympt.unk [arrested¶]	cxr.pos | micro.pos | sympt.unk
2. Anastasatu(22)	Pub‖: 1985	Unknown	Unknown	Not specified; Romania	Control/Placebo arm	Prospective	143	>1 culture (active)	Inactive	Cumulative	cxr.pos | micro.neg | sympt.neg	cxr.pos | micro.pos | sympt.unk
3. Aneja(23)	1975 - 1977	Minimum: 12 years	41.8	Not specified; India	Control/Placebo arm	Prospective	110	>1 culture (active)	Not specified	Single	cxr.pos | micro.neg | sympt.pos	cxr.pos | micro.pos | sympt.unk
4. Beeuwkes(24)	1933 - 1938	Unknown	Unknown	Not specified; USA	Household Contact Study	Prospective	784	>1 culture (not specified)	Neg	Single	cxr.neg | micro.neg | sympt.neg	cxr.unk | micro.pos | sympt.pos
							79		Inactive	Single	cxr.pos | micro.neg | sympt.neg	cxr.unk | micro.pos | sympt.pos
							43		Active	Single	cxr.pos | micro.neg | sympt.pos	cxr.unk | micro.pos | sympt.pos
							28		Active	Single	cxr.pos | micro.pos | sympt.pos	cxr.unk | micro.neg | sympt.unk
5. Bobrowitz(25,26)	1938 - 1945	Unknown	58	Not specified; USA	Clinic/Hospital / Sanatorium cohort	Prospective	191	≥1 culture (passive)	Mixed	Single	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
6. Borgen(27,28)	1947 - 1949	Minimum: 15 years	20.5	High; Norway	Occupational /Student Screening	Prospective	24	Unclear (active)	Active	Single	cxr.pos | micro.neg | sympt.pos	cxr.pos | micro.pos | sympt.pos
							120		Active	Single	cxr.pos | micro.neg | sympt.neg	cxr.pos | micro.pos | sympt.pos

7. Breu(29)	1949 - 1952	Minimum: 15 years	Unknown	Medium; Germany	General Community Survey	Prospective	904	>1 culture (active)	Mixed	Single	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
8. Cowie(30)	1979-1984	Unknown	0	Not specified; South Africa	Occupational /Student Screening	Prospective	152	>1 culture (passive)	Active	Cumulative	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
9. Downes(31)	1923 - 1935	Range: 15-69 years	50	Not specified; USA	Clinic/Hospital / Sanatorium cohort	Retrospective	342	Microscopy Only (passive)	Active	Cumulative	cxr.pos | micro.pos | sympt.pos	cxr.pos | micro.neg | sympt.neg
10. Frimodt-Moller(32)	1960 - 1961	Minimum: 15 years	29	High; India	Control/Placebo arm	Prospective	86	>1 culture (active)	Active	Cumulative	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
11. Hong Kong Chest Service(33–36)	Pub‖: 1979-1981	Range: 15-75	30.68	Not Specified; Hong Kong	Control/Placebo arm	Prospective	176	>1 culture (active)	Active	Cumulative	cxr.pos | micro.neg | sympt.mix	cxr.pos | micro.pos | sympt.unk
12. IUAT Committee on Prophylaxis(37)	Pub‖: 1982	Average: 50 years	47	Not specified; Europe	Control/Placebo arm	Prospective	6990	>1 culture (active)	Inactive	Cumulative	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
13. Lincoln(38)	1937 - 1947	Minimum: 14 years Average: 24 years	49.8	Not specified; USA	Clinic/Hospital / Sanatorium cohort	Retrospective	314	Unclear (passive)	Inactive	Cumulative	cxr.pos | micro.neg | sympt.unk [arrested¶]	cxr.pos | micro.pos | sympt.unk
14. Manser(39)	1941 - 1951	Range: 60-83 years	Unknown	Not specified; Switzerland	Clinic/Hospital / Sanatorium cohort	Retrospective	40	Unclear (active)	Active	Single	cxr.pos | micro.pos | sympt.unk	cxr.pos | micro.neg | sympt.unk
15. Marshall(40)	1947-1948	Range: 15-30 years	59.6	Not specified; United Kingdom	Control/Placebo arm	Prospective	52	>1 culture (active)	Active	Cumulative	cxr.pos | micro.pos | sympt.pos	cxr.pos | micro.neg | sympt.unk
16. National Tuberculosis	1961 - 1968	Minimum: 5 years	49	High; India	General Community Survey	Prospective	31490	>1 culture (passive)	Neg	Single	cxr.neg | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk

Institute(42–49)							329	>1 culture (active)	Active	Single	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
							269	>1 culture (active)	Active	Single	cxr.pos | micro.pos | sympt.unk	cxr.pos | micro.neg | sympt.unk
17. Norregaard(50)	1978 - 1985	Minimum: 20 years	39	Medium; Denmark	Control/Placebo arm	Prospective	28	>1 culture (active)	Active	Cumulative	cxr.pos | micro.neg | sympt.mix	cxr.pos | micro.pos | sympt.neg cxr.pos | micro.pos | sympt.pos
18. Okada(51)	2002 - 2004	Minimum: 10 years Median: 30.6 years	54	High; Cambodia	General Community Survey	Prospective	309	>1 culture (active)	Active	Single	cxr.pos | micro.neg | sympt.neg	cxr.pos | micro.pos | sympt.unk
								>1 culture (passive)				cxr.pos | micro.pos | sympt.pos
								>1 culture (active)				cxr.neg | micro.neg | sympt.unk
							21580	>1 culture (passive)	Neg	Single	cxr.neg | micro.neg | sympt.neg	cxr.pos | micro.pos | sympt.unk
19. Orrego Puelma(52)	Pub‖: 1945	Minimum: 15	42	Not specified; Chile	Clinic/Hospital / Sanatorium cohort	Retrospective	67	>1 culture (passive)	Mixed	Single	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
20. Pamra(53)	1958 - 1968	Range: 15-45 years	11	Not specified; India	Control/Placebo arm	Prospective	178	>1 culture (active)	Inactive	Cumulative	cxr.pos | micro.neg | sympt.neg	cxr.pos | micro.pos | sympt.pos cxr.pos | micro.pos | sympt.neg

21. Puffer(54)	1931 - 1943	Unknown	60.6	Not specified; USA	Clinic/Hospital /Sanatorium cohort	Retrospective	261	>1 culture (passive)	Mixed	Single	cxr.pos | micro.neg | sympt.neg	cxr.pos | micro.pos | sympt.pos
							267		Inactive	Single	cxr.pos | micro.neg | sympt.neg [arrested¶]	cxr.pos | micro.pos | sympt.pos
							384		Active	Single	cxr.pos | micro.pos | sympt.pos	cxr.pos | micro.neg | sympt.neg
22. Sikand(55)	1952 - 1958	Minimum: 15 years	Unknown	Not specified; India	Occupational/ Student Screening	Prospective	167	Unclear (not specified)	Inactive	Cumulative	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
							152		Mixed	Cumulative	cxr.pos | micro.neg | sympt.unk	cxr.pos | micro.pos | sympt.unk
23. Styblo(56)	1961 - 1965	Minimum: 15 years	53	High; Czechoslovakia	General Community Survey	Prospective	73000	>1 culture (passive)	Neg	Cumulative	cxr.neg | micro.neg | sympt.neg	cxr.unk | micro.pos | sympt.pos
												cxr.unk | micro.pos | sympt.unk
24. Tuberculosis Society of Scotland(57,58)	1954 - 1959	Minimum: 15	38.95	Not specified; Scotland	Control/Placebo arm	Prospective	95	>1 culture (active)	Inactive	Single	cxr.pos | micro.neg | sympt.neg	cxr.pos | micro.pos | sympt.unk

* "Active" microbiological follow up refers to those that systematically tested at specified intervals (or attempted testing in those able to produce sputum). "Passive" microbiological follow up refers to those where results either relied only on screening of TB notification or clinical records, or if they only assessed microbiologically when clinical or radiological deterioration was noted

† Based on historical authors' description/classification. Further descriptions of how the authors made these assessments are available in the supplementary material.

‡ "Single" follow up refers to those with two cross-sectional assessments of the group of participants, whereas "cumulative" refers to those studies which cumulatively captured events over time

§ Start and end points have three characteristics or states including radiological (i.e. cxr.neg or cxr.pos or cxr.unk), microbiological (i.e. micro.neg, micro.pos, micro.unk, or micro.mixed), and symptom status (i.e. sympt.neg, sympt.pos, sympt.unk or sympt.mixed). Within these, “neg” denotes negative, “pos” denotes positive, “unk” denotes unknown and “mix” denotes mixed

¶ This group were known to have been microbiologically positive on a prior occasion, then documented arrested disease, followed by a relapse

|| “Pub” denotes those studies where the years in which studies were conducted are not known and therefore publication year/s are listed instead

ATT=Antituberculosis Therapy; IUAT=International Union Against Tuberculosis; Micro.=Microbiology; USA=United States of America

Progression to microbiologically positive disease in those with abnormal chest x-ray at baseline

From the 24 cohorts with abnormal chest radiography but no evidence of *M. tb* on respiratory sampling at baseline representing 11,185 participants, development of microbiologically-detectable disease occurred in between 1.1 – 57.9% of individuals with half the studies reporting a follow-up of up to three years (range 12-156 months) (figure 3). Considerable statistical heterogeneity was seen across cohorts (I^2 = 97.3%, tau^2=0.001, p<0.01). We considered that the radiographic abnormalities categorized as active versus inactive TB (as specified by the original authors; supplementary table 3) could represent distinct pathological states contributing to clinical variability of studies. Therefore we did not pool these studies in meta-analysis, but rather conducted stratified meta-analysis to describe the progression of these two states separately. The annualized rate of transition from microbiologically negative to positive was 9.7% (95% CI: 6.2-13.3) for those in the nine cohorts described to have active changes on radiography compared to 1.1% (95% CI: 0.3-1.8) for those in the seven cohorts with inactive changes (figure 3). Over a three-year period, this would equate to 26% (95%CI: 17-35) of those with active TB changes vs 3% (95%CI: 1-5) with inactive TB changes progressing from microbiologically positive to negative disease. Statistical heterogeneity in the active and inactive TB subgroups was lower than in all cohorts taken together, I^2 = 77.4% and I^2 = 53.2% respectively. The annual progression in cohorts with "mixed" radiographic changes was 6.3% (95% CI: 1.5-11.1) - in between the values for inactive and active strata.

Out of 24 cohorts that contributed patients to this group, 18 (75%) used culture as part of microbiological work-up and the remainder (n=6/25) did not specify the microbiological tests undertaken. Restricting this analysis to the 18/24 cohorts explicitly using culture had little impact on these results (supplementary figure 1). Only 11 cohorts provided data on symptom status. Of the 9 cohorts described to have active TB changes on radiography, three were in symptomatic individuals. Progression in this subgroup was at an annualized rate of 11.74% (95% CI: 2.73-20.75) (supplementary figure 2). There was only one cohort describing active TB changes on radiography in an asymptomatic group with the remainder unknown.

In the four cohorts following up those with no radiographic changes suggestive of any TB (table 2), transition to microbiologically positive occurred at an annualized rate of 0.14% (95% CI: 0.11-0.17) (supplementary figure

3). In “single follow up” and “cumulative count” studies, those with active TB changes showed similar annual progression (supplementary figure 4).

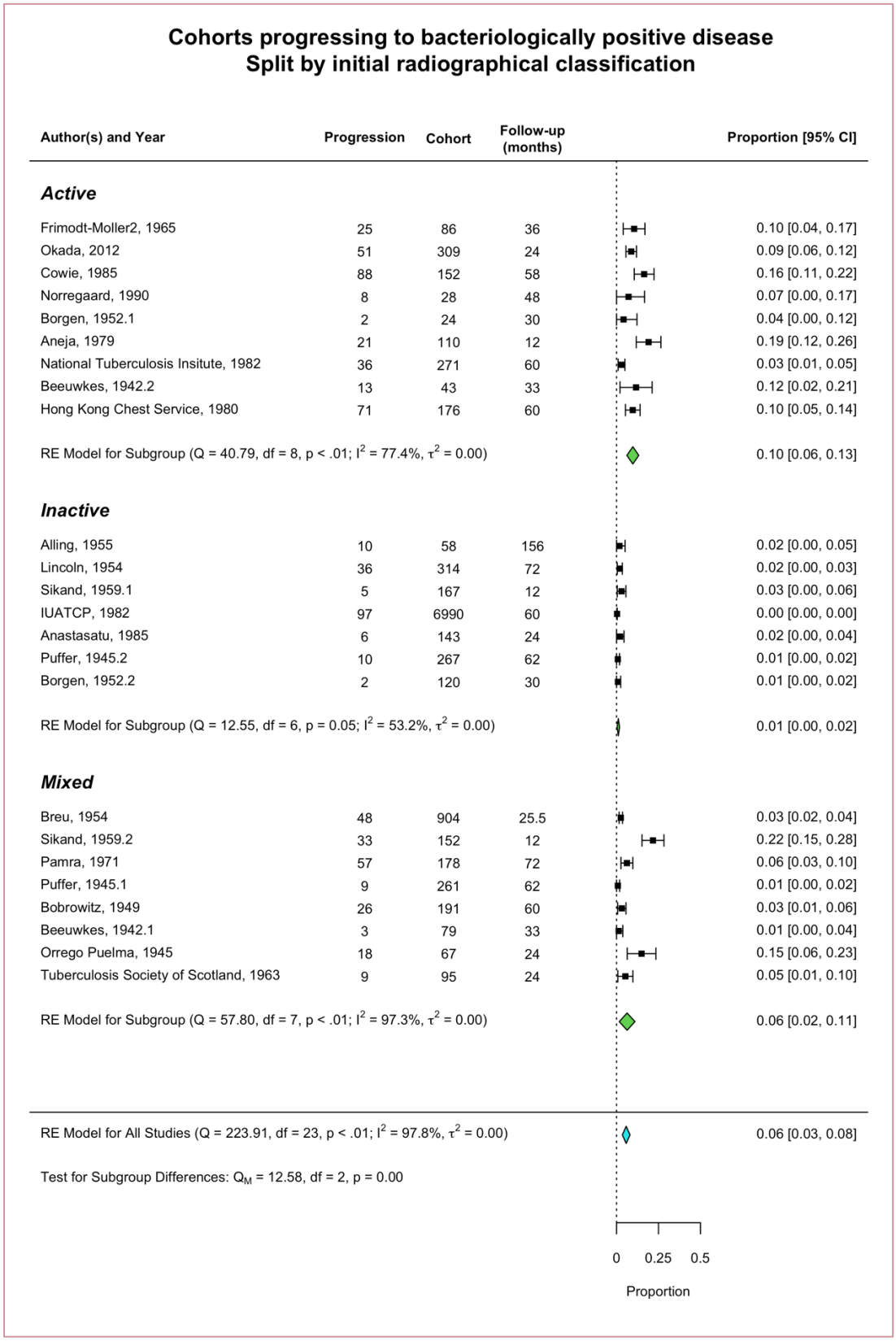


Figure 3: participants entering cohorts with abnormal chest X-rays and negative microbiology, transitioning to positive microbiology: forest plot of the random effects meta-analysis of annualized rates (as described fully in methods section) with proportion and 95% confidence intervals for subgroups. Subgroups are as per the historical authors' provided data on radiographic classification being either "active", "inactive" or where the group was "mixed".

Regression to negative microbiology in those with positive microbiology at baseline

Six cohorts followed a total of 1115 participants with evidence of *M. tb* in respiratory samples at baseline and assessed the proportion transitioning to a microbiologically undetectable state without treatment or intervention, with two thirds of studies reporting up to three years of follow up (range 6-26 months). The majority of these cohorts included participants with limited or minimal disease on CXR - either due to this being entry criteria into the original study or due to the eligibility criteria of this systematic review. No studies were able to adequately describe symptom status of the participants. Three out of six were retrospective cohorts from TB hospitals or sanatoria and three were prospective cohorts from general community/household surveys or a placebo arm of a trial. In four of the six cohorts, culture was used to assess microbiological status of participants while in two cohorts, both retrospective, either microscopy was used or nature of microbiological investigations was not specified. With meta-analysis, this transition occurred at an annualized rate of 18.3% (95% CI: 3.0-33.7) (figure 4), but there was considerable heterogeneity across these studies (I^2 = 98.1%, tau^2=0.03, p<0.01). We then restricted the meta-analysis to prospective studies, hence removing the three retrospective hospital/sanitoria cohorts, where culture had also not be used in two instances, and showed an annualized rate of 12.4% (95% CI: 6.8-18.0) with reduced statistical heterogeneity I^2 = 35.1%. Over three years this would equate to 33% (95% CI: 19-45) of those initially with culture positive TB becoming culture negative.

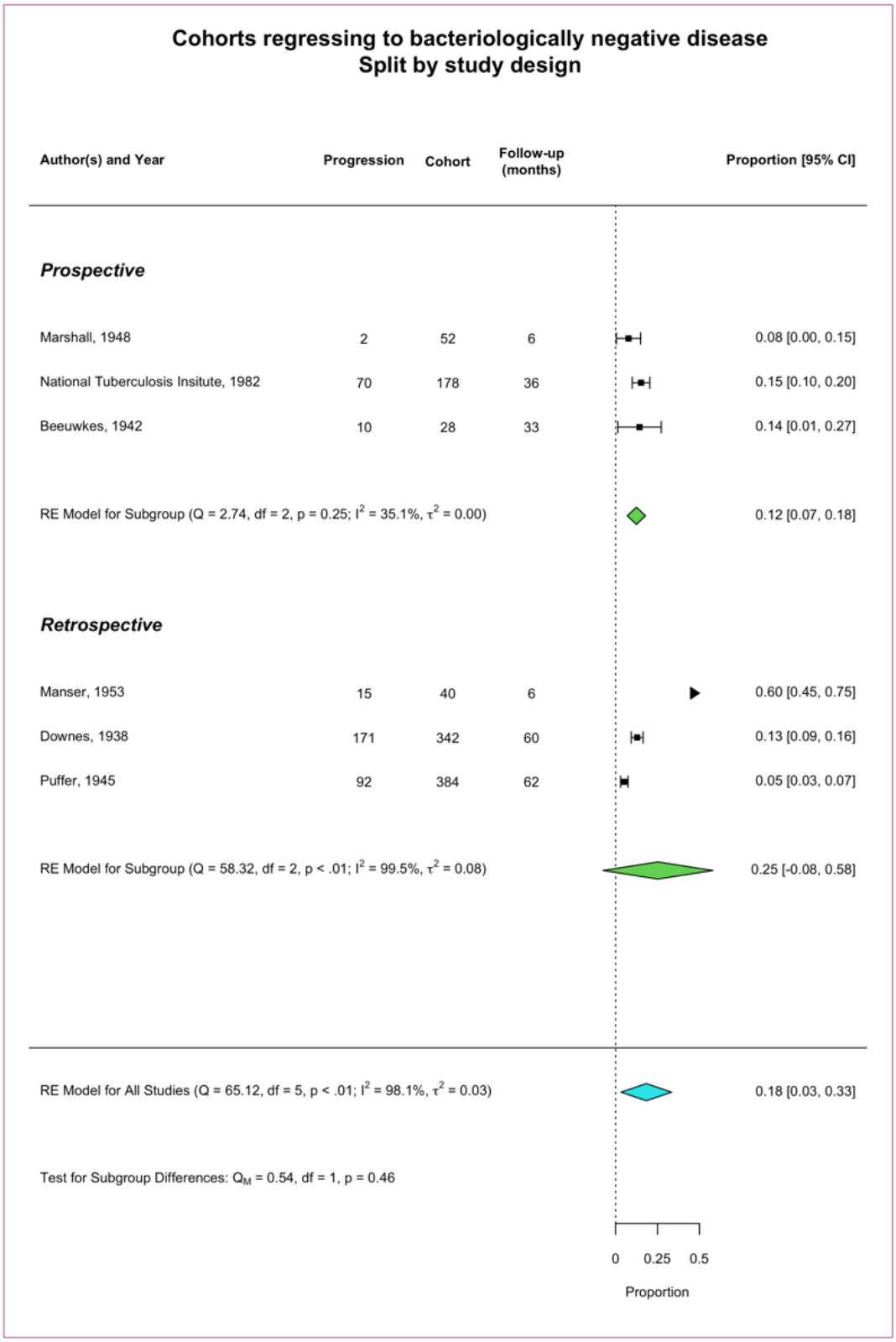


Figure 4: participants entering cohorts with positive microbiology, transitioning to negative microbiology: forest plots of the random effects meta-analysis of annualized rates (as described fully in methods section) with proportion and 95% confidence intervals for subgroups, according to study design (figure 4A) and when limiting to those studies that documented using culture (figure 4B).

DISCUSSION

This review is the first to systematically summarize key aspects of the kinetics of the natural history of untreated tuberculosis in adults, outside of the rate of mortality, making full use of historical literature in English and German. Through meta-analysis we provide estimates of the risk of progression to microbiologically positive disease in those with initially negative microbiology at an annualized rate of approximately 10% in those with "active" radiographic TB changes and 1% in those with "inactive" or fibrotic changes. For comparison, progression was approximately 0.1% for those with normal CXRs, while recognizing that this rate would be affected factors such as local burden of disease. In addition to this we provided an estimate for the reversion from culture-positive disease to culture negative without treatment (also referred to as 'self-cure) as 12.4% per year.

These results highlight that individuals with CXR changes suggestive of active TB but who are found to be initially microbiologically negative are at considerable risk of disease progression. Guidance on how to manage this group is often unsatisfactory with a tension between providing empirical treatment, or monitoring for bacteriological progression which can be resource intensive or unfeasible in resource constrained settings. Better understanding of the risk of disease progression provided by this study could assist decision making and inform the planning of necessary clinical trials. In addition, our study is the first to determine an estimate for this transition which will be of use to modellers wanting to understand the implications on intervening in this population. We also have shown that approximately a third of those with culture positive disease could revert to culture negative without treatment over a 3-year period. While this may not inform clinical management, our results may refine parameters in models used to estimate disease incidence from prevalence survey data where the probability of so-called "self-cure" needs to be factored in. Our annual rate of approximately 12% provides empirical foundation to the slightly higher rates of 15% and 20% used by Dye to parameterize "self-cure" – which was informed by a review of literature although not systematic(59,60).

We used a widely accepted conceptual framework to guide our data collection which required determination of the microbiological, radiological and symptom status of participants over follow-up. We found that no single

study systematically recorded these three features over the entire course of disease from exposure to final outcome. In addition we found that the recording of symptoms in these studies was not explicit, particularly during follow up - meaning there was insufficient empirical data to directly determine the trajectory around subclinical (asymptomatic, microbiologically positive) TB. Subclinical TB is a commonly identified state through CXR-based active case finding but conducting contemporary natural history studies to determine the rates of progression and regression would present ethical challenges with the availability of treatment. However, the substantial additional data uncovered in this review should allow inference of the kinetics around subclinical TB, which Richards *et al* have explored using a Bayesian framework to utilize the information from all available data simultaneously(61).

There are several key limitations to consider when interpreting the findings of this systematic review. HIV is a significant role-player in the epidemiology of TB in certain settings today and 22 of 24 of our studies were set prior to the discovery of the virus. It is likely that people living with HIV progress along the disease spectrum with different kinetics, also influenced by immune status (62–64). Secondly the nature of this research question and the historical focus resulted in studies being included from a period spanning almost 80 years; over this time period, microbiological and radiographic methods evolved. However, from a microbiological perspective included studies predominately used culture and where they did not, we conducted sensitivity analyses. For radiology, even where studies used mass miniature radiography or fluoroscopy for screening, findings were typically confirmed with conventional chest radiography which informed data extraction. The majority of studies were conducted over fifty years ago, when socioeconomic, health access, comorbidity distribution and prevalence of TB were likely very different to what they are today. However, these study environments may to a certain extent remain representative of many contemporary settings with a high TB burden.

There are also considerable methodological challenges in conducting a systematic review involving historical research. It is notable that 1503/1648 (91.2%) of studies were retrieved for full text review, however for 95 studies that met eligibility and bias criteria, manuscript style did not allow for data extraction and authors could not be contacted for assistance. Although our work focused on the period 1895-1960, through extensive investigator collections and snowballing of references we are confident we were able to identify key literature post-1960 as evidenced by nearly half of our final 24 studies being after this date.

Through our extensive review, we find that the natural history of TB is a dynamic, heterogenous process which is not adequately represented by a single 'active disease' state, and quantified three key transitions. Importantly, this review provides a much-needed foundation of empirical data for our ongoing re-discovery of the complexity of TB natural history, enabling a grounding for new preconceptions or dogmas, and a drive toward new clinical guidelines and policies for those suffering from TB.

CONTRIBUTORS

HE, RH, BS, AR, FC, and KK conceptualised the study protocol. BS, AR, TH, BF, FB, AO, and BH carried out the literature search and data collection. AR and BS carried out the statistical analysis and verified the final data with input and oversight from HE, RH and ER. BS wrote the first draft of the manuscript with input from AR, RH and HE. All authors subsequently reviewed and edited the manuscript. All authors had full access to the study data and had final responsibility for the decision to submit for publication.

DECLARATION OF INTERESTS

We declare no competing interests.

ACKNOWLEDGEMENTS

This work was funded by TB Modelling and Analysis Consortium (MAC), Medical Research Council (Grant ref: MR/V00476X/1) and the European Research Council (Starting Grant Action Number 757699). We would like to acknowledge Adrienne Burrough, Bridget Chivers and Rohisha Luchin for their assistance in performing the hand searches of Index Medicus and in collecting literature from various English libraries, and would like to thank the team of the German Tuberculosis Archives in Heidelberg for assisting with collection of German literature.

REFERENCES

1. Barry CE, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol. 2009 Dec;7(12):845–55.

2. Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, et al. Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. Clin Microbiol Rev. 2018 Jul 18;31(4):e00021-18.

3. Gothi DGD. NATURAL HISTORY OF TUBERCULOSIS. Indian J Tuberc Suppl. 1978;25(2):13.

4. Behr MA, Kaufmann E, Duffin J, Edelstein PH, Ramakrishnan L. Latent Tuberculosis: Two Centuries of Confusion. Am J Respir Crit Care Med. 2021 Mar 24;rccm.202011-4239PP.

5. Squire JE. Prognosis in Pulmonary Tuberculosis. Hospital (Rio J). 1909;4.

6. Gillespie JR. Prognosis in pulmonary tuberculosis. Br Med J. 1928 Jun;1928(1):436–7.

7. Amberson JB, Jones JM. Prognosis and treatment in minimal pulmonary tuberculosis. XIIth Conf Int Union Tuberc. 1952;Report and co-reports on the second subject:1–19.

8. Dijkstra C. Results of a frequent and longlasting sputum examination of cultures, performed on 445 patients suffering from tertiary pulmonary tuberculosis. Acta Tuberc Scand. 1952;26:301–14.

9. World Health Organization. What is DOTS? A guide to understanding the WHO-recommended TB Control Strategy [Internet]. 1999. Available from: https://apps.who.int/iris/bitstream/handle/10665/65979/WHO_CDS_CPC_TB_99.270.pdf?sequence=1&isAllowed=y

10. Cobelens F, Suri RK, Helinski M, Makanga M, Weinberg AL, Schaffmeister B, et al. Accelerating research and development of new vaccines against tuberculosis: a global roadmap. Lancet Infect Dis. 2022 Apr;22(4):e108–20.

11. World Health Organisation. Global tuberculosis report 2019. 2019.

12. Glaziou P, Dodd P, Dean A, Floyd K. Methods used by WHO to estimate the global burden of TB disease. 2020 Oct 14;

13. Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJ. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS ONE. 2011;6(4):e17601.

14. Frascella B, Richards A, Sossen B, Emery J, Odone A, Law I, et al. Subclinical tuberculosis disease - a review and analysis of prevalence surveys to inform definitions, burden, associations and screening methodology. Clin Infect Dis. 2020;43.

15. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Primer. 2016 Dec;2(1):16076.

16. Esmail H, Barry CE, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Philos Trans R Soc B Biol Sci. 2014 Jun 19;369(1645):20130437.

17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;n71.

18. Long ER, Hopkins FD. History of Diagnostic Standards and Classification of Tuberculosis of the National Tuberculosis Association. Am Rev Tuberc. 1952 Apr;65(4):494–504.

19. World Health Organization. WHO global lists of high burden countries for tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB), 2021–2025: background document [Internet]. Geneva: World Health Organization; 2021 [cited 2021 Nov 22]. Available from: https://apps.who.int/iris/handle/10665/341980

20. Naike Wang. How to Conduct a Meta-Analysis of Proportions in R: A Comprehensive Tutorial. 2018 [cited 2021 Oct 18]; Available from: http://rgdoi.net/10.13140/RG.2.2.27199.00161

21. Alling D, Bosworth E, Lincoln N. The after-history of pulmonary tuberculosis: V. Moderately Advanced Tuberculosis. Am Rev Tuberc. 1955;71(4):519–28.

22. Anastasatu C, Berceea O, Corlan E. Controlled clinical trial on smear negative, x-ray positive new cases, with the view to establishing if and how to treat them. Bull Int Union Tub. 1985;60:108–9.

23. Aneja KS, Gothi GD, Samuel GER. Controlled study of the effect of spcific treatment on bacteriological status of "suspect cases." Indian J Tuberc. 1979;26:50–7.

24. Beeuwkes H, Hahn RG, Putnam P. A survey of persons exposed to tuberculosis in the household. Am Rev Tuberc. 1942;45(2):165–93.

25. Bobrowitz ID, Hurst A, Martin M. Minimal Tuberculosis; the prognosis and clinical significance of a sanatorium treated group. Am Rev Tuberc. 1947;56(2):16.

26. Bobrowitz ID, Hurst A. Minimal Tuberculosis: Problems in Roentgenologic Interpretation. Radiology. 1949 Apr;52(4):519–32.

27. Borgen L., Meyer SN., Refsum E. A preliminary report of tuberculin tests and mass chest surveys carried out in Aker. Acta Tuberc Scand Suppl. 1950;26:159–68.

28. Borgen L., Meyer SN., Refsum E. Mass photofluorography, tuberculin testing, and BCG vaccination in the district of Aker (Norway) 1947-49. Acta Tuberc Scand. 1951;25(4):327–55.

29. Breu K. [Public health x-ray diagnosis of closed pulmonary tuberculosis later proved contagious]. Beitrage Zur Klin Tuberk Spezifischen Tuberk-Forsch. 1954;111(5):437–44.

30. Cowie L, Langton ME. Diagnosis of sputum smear- and sputum culture-negative pulmonary tuberculosis. SAMJ. 1985;68:1.

31. Downes J. A Study of Mortality among Individuals with Active Pulmonary Tuberculosis. Milbank Mem Fund Q. 1938 Jul;16(3):304.

32. Frimodt-Moller J, Parthasarathy R, Thomas J. Results of treatment of non-bacillary tuberculosis in a domiciliary treatment prrogramme,-a preliminary report. 1960;

33. Hong Kong Chest Service. SPUTUM-SMEAR-NEGATIVE PULMONARY TUBERCULOSIS CONTROLLED TRIAL OF 3-MONTH AND 2-MONTH REGIMENS OF CHEMOTHERAPY. The Lancet. 1979 Jun;313(8131):1361–3.

34. Hong Kong Chest Service. A controlled trial of 2-month, 3-month, and 12-month regimens of chemotherapy for sputum smear-negative pulmonary tuberculosis: the results at 30 months. 1981;

35. Hong Kong Chest Service, Centre TR. A Controlled Trial of 2-Month, 3-Month, and 12-Month Regimens of Chemotherapy for Sputum-Smear-Negative Pulmonary Tuberculosis: Results at 60 months. Am Rev Respir Dis. 1984;6.

36. Hong Kong Chest Service, Tuberculosis Research Centre, Mad, British Medical Research Council. A study of the characteristics and course of sputum smear-negative pulmonary tuberculosis. Tubercle. 1981 Sep;62(3):155–67.

37. International Union Against Tuberculosis Committee on Prophylaxis. Efficacy of various duration of isoniazid preventive therapy for tuberculosis: five years of follow-up in the IUAT trial. Bull World Health Org. 1982;60(4):555–64.

38. Lincoln NS, Bosworth EB, Alling DW. The After-History of Pulmonary Tuberculosis - III: Minimal tuberculosis. Am Rev Tuberc. 1954;70(1):17.

39. Manser H. [Tuberculosis in aged and its course during sanatorium treatment]. Schweiz Z Tuberk Rev Suisse Tuberc Riv Svizzera Della Tubercolosi. 1953;10(2):65–82.

40. Medical Research Council. Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. BMJ. 1948 Oct 30;2(4582):769–82.

41. National Tuberculosis Institute B. Tuberculosis in a rural population of South India: a five-year epidemiological study. Bull World Health Organ. 1974;51(5):473–88.

42. Chakaraborty, A K, Singh, H, Srikantan, K, Rangaswamy, K R, Krishnamurthy, MS, Stephen, J A. Tuberculosis in a rural population of South India: Report on five surveys. Indian J Tuberc. 1982;29(3):153–67.

43. CHAKRABORTY AK, GOTHI GD, DWARKANATH S, SINGH H. Tuberculosis mortality rate in a South Indian rural population. Indian J Tuberc. 1978;25(3):181–6.

44. Krishnamurthy V, Nair S, Gothi G. A comparison of new cases (incidence cases) who had come from different epidemiological groups in the population. Indian J Tuberc. 1978;25(3).

45. Gothi GD, Chakaraborty AK, Jayalakshmi MJ. INCIDENCE OF SPUTUM POSITIVE TUBERCULOSIS IN DIFFERENT EPIDEMIOLOGICAL GROUPS DURING FIVE YEAR FOLLOW UP OF A RURAL POPULATION IN SOUTH INDIA. Indian J Tuberc. 1978;25(2):9.

46. Krishnamurthy VV, Nair SS, Gothi GD, Chakraborty AK. Incidence of Tuberculosis among newly infected population and in relation to the duration of infected status. Indian J Tuberc. 1976;23(1).

47. Gothi GD, Nair SS, Chakraborty AK, Ganapathy KT. Five year incidence of tuberculosis and crude mortality in relation to non-specific tuberculin senssitivity. Indian J Tuberc. 1976;23(2).

48. Chakraborty AK, Gothi GD. RELAPSES AMONG NATURALLY CURED CASES OF PULMONARY TUBERCULOSIS. Indian J Tuberc. 1976;23(1):6.

49. Chakraborty AK. INTERPRETATION OF PHOTOFLUOROGRAMS OF ACTIVE PULMONARY TUBERCULOSIS PATIENTS FOUND IN EPIDEMIOLOGICAL SURVEY AND THEIR FIVE YEAR FATE. Indian J Tuberc. 1974;21:8.

50. Norregaard J. Abacillary pulmonary tuberculosis. Tubercle. 1990;71:35–8.

51. Okada K, Onozaki I, Yamada N, Yoshiyama T, Miura T, Saint S, et al. Epidemiological impact of mass tuberculosis screening: a 2-year follow-up after a national prevalence survey. :7.

52. Orrego Puelma H, Grebe G. Analysis of One Hundred Cases of Minimal Pulmonary Tuberculosis. Dis Chest. 1945 Sep;11(5):375–9.

53. Pamra SP, Mathur GP. Effects of Chemoprophylaxis on Minimal Pulmonary Tuberculosis Lesions of Doubtful Activity. Bull World Health Org. 1971;45(5):10.

54. Puffer RR, Stewart HC, Gass RS. Tuberculosis according to age, sex, family history, and contact. Am Rev Tuberc. 1945;51(4):295–311.

55. Sikand BK, Narain R, Mathur GP. Incidence of TB as Judged by Re-surveys. A study of Delhi Police. Indian J Tuberc. 1959;6(3):73–83.

56. Styblo K, Dajkova D, Kubik A, Langerova M, Radkovsky J. Epidemiological and Clinical Study of Tuberculosis in the District of Kolin, Czechoslovakia. Bull World Health Org. 1967;37:56.

57. Tuberculosis Society of Scotland. A controlled trial of chemotherapy in pulmonary tuberculosis of doubtful activity. Tubercle. 1958 Jun;39(3):129–37.

58. Scottish Thoracic Society. A controlled trial of chemotherapy in pulmonary tuberculosis of doubtful activity: Five year follow-up. Tubercle. 1963 Mar;44(1):39–46.

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

60. Dye C. The Population Biology of Tuberculosis. United Kingdom: Princeton Univerity Press; 2015.

61. Richards A, Sossen B, Emery J, Horton K, Heinsohn T, Frascella B, et al. The natural history of TB diseease - a synthesis of data to quantify progression and regression across the spectrum. medRxiv. 2021;

62. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Coussens AK, et al. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18F]fluoro-D-glucose positron emission and computed tomography. Nat Med. 2016 Oct;22(10):1090–3.

63. Lawn S, Kerkhoff A, Wood R. Progression of subclinical culture-positive tuberculosis to symptomatic disease in HIV-infected individuals. AIDS. 25(17):2190–1.

64. Oni T, Burke R, Tsekela R, Bangani N, Seldon R, Gideon HP, et al. High prevalence of subclinical tuberculosis in HIV-1-infected persons without advanced immunodeficiency: implications for TB screening. Thorax. 2011 Aug 1;66(8):669–73.

65. Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. The Lancet. 2011 Apr;377(9776):1495–505.

Supplementary Appendix for:

Natural History of Untreated Pulmonary Tuberculosis in Adults: A Systematic Review and Meta-Analysis

Bianca Sossen (MBChB)[1,2,#], Alexandra Richards (MMathPhys)[3,4,#], Torben Heinsohn (BMBCh)[2,5], Beatrice Frascella (MD)[6], Federica Balzarini (MD)[6], Aurea Oradini-Alacreu (MD)[6], Prof. Anna Odone (PhD)[7], Ewelina Rogozinska (PhD)[8], Brit Häcker (Dr)[9], Prof. Frank Cobelens (PhD)[10,11], Prof. Katharina Kranzer (PhD)[12-14], Prof. Rein MGJ Houben (PhD)[3,4;&], Hanif Esmail (PhD)[2,8,15;&,¥]

[1] Department of Medicine, Faculty of Health Sciences, University of Cape Town, South Africa

[2] Institute for Global Health, University College London, London, UK
[3] TB Modelling Group, TB Centre, London School of Hygiene and Tropical Medicine, London, UK
[4] Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, London, UK
[5] Helmholtz Centre for Infection Research, Department of Epidemiology
[6] School of Public Health, Vita-Salute San Raffaele University, Milan, Italy
[7] Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia, Italy
[8] MRC Clinical Trials Unit at University College London, UK
[9] German Central Committee against Tuberculosis (DZK), Berlin, German
[10] Amsterdam University Medical Centers location University of Amsterdam, Department of Global Health
[11] Amsterdam Institute for Global Health and Development, Amsterdam, The Netherlands
[12] Clinical Research Department, Faculty of Infectious and Tropical Disease, London School of Hygiene and
Tropical Medicine, London, UK
[13] Biomedical Research and Training Institute, Harare, Zimbabwe
[14] Division of Infectious Diseases and Tropical Medicine, University Hospital, LMU Munich, Munich,
Germany
[15] Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Diseases and
Molecular Medicine, University of Cape Town, South Africa

# Contributed Equally
& Contributed Equally
¥ Corresponding Author: Dr Hanif Esmail at MRC Clinical Trials Unit, 90 High Holborn, London, WC1V 6LJ or h.esmail@ucl.ac.uk

Supplementary Table 1: Adapted Newcastle Ottawa Quality Assessment Tool

GENERAL QUALITY ASSESSMENT	MAXIMUM STARS
STUDY SELECTION	
1. Representativeness of 'exposed' cohort	
Truly representative of general population e.g. identified through prospective recruitment following mass screening or household screening	1 star
Somewhat representative of general population e.g. identified through occupational screening	
Selected group e.g. sanatorium or hospital patients	0 stars
No description of the derivation of the cohort	
2. Selection of non-exposed cohort	
NOT APPLICABLE	
3. Ascertainment of exposure*	
Documented TST conversion	1 star
Exposure appropriately inferred from significant exposure to active TB e.g. in household or occupational setting	
No clear significant exposure to active TB	0 stars
No description	
4. Demonstration that outcome of interest not present at start of study*	
Yes - evidence of TB excluded by CXR at beginning of follow-up, or documented as not being present on CXR within previous 12 months	1 star
Not stated	0 stars
COMPARABILITY	
NOT APPLICABLE	
OUTCOME	
5. Assessment of outcome	
Disease defined on the basis of microbiological demonstration of organism	1 star
Disease defined without microbiological evidence	0 stars
No description	
6. Could intervention prevent development of outcome	
No medical or surgical treatment provided	1 star
Medical or surgical treatment provided to a subgroup of clearly identified individuals analysed separately and no selection bias	
Medical or surgical treatment provided to a subgroup (<10%) and that treatment deemed unlikely to influence outcome	

Medical or surgical treatment provided to a large number of individuals OR unable to be analysed separately OR selection bias	0 stars
7. Was total follow-up long enough for outcomes to occur	
Yes - follow-up equal to or longer than 12 months	1 star
No	0 stars
8. Adequacy of follow-up of cohorts	
Complete follow-up	1 star
Participants lost to follow up unlikely to introduce bias: <20% of participants lost to follow-up per annum, or description provided for those lost, making introduction of bias unlikely	
Lost to follow-up rate > 20% per annum and no description of those lost	0 stars
No statement	

* These sections were only applicable to those cohorts of patients that started with CXR and microbiology negative for TB

ADDITIONAL SPECIFIC QUALITY ASSESSMENT *- If study meets general quality criteria above, used to determine if microbiological, radiographic and symptoms are recorded to sufficient quality*	MAXIMUM STARS
RADIOGRAPHY	
1. Technology	
Radiography (X-Ray, Roentgenography)	1 star
Mass Miniature radiography (Photofluorography, Aburography)	
Other (please specify)	0 stars
2. Diagnostic criteria used	
Clear use of recognised diagnostic criteria (e.g. ATS)	1 star
Clear description of study specific methodology (not recognised diagnostic criteria)	
No or limited description of criteria or methodology	0 stars
No imaging	
3. Methodology	
Double or triple independent read	1 star
Single reader	0 stars
Not specified	
4. Follow-up	
Complete follow-up	1 star
<20% of participants per annum in follow-up did not receive imaging OR description for those not receiving imaging, making introduction of bias unlikely	
<80% followed up received imaging and no description of those lost	0 stars

No statement	
MICROBIOLOGY	
1. Technology	
Culture - solid	1 star
Culture - liquid	
Smear - with concentration	
Smear - without concentration	
Guinea pig inoculation	
Other (please specify)	
Not specified	0 stars
2. Sampling method	
24-hour sputum collection	1 star
48-hour sputum collection	
72-hour sputum collection	
Induced sputum (e.g. physiotherapy or saline)	
Spot sputum collection (including <24 hour collection	
Gastric aspirate	
Nasal/Laryngeal/Tracheal swab	
Other (please specify)	
Not specified	0 stars
3. Follow-up	
Complete follow-up of those able to produce sputum	1 star
<20% of per annum of participants had sputum collected (of those able to produce) for microbiological investigation OR description provided for that lost, making introduction of bias unlikely	
Follow-up rate <80% and no description of those lost	0 stars
No statement	
SYMPTOMS	
1. Methodology	
Standardised symptom screen described in methodology	1 star
Clear statement of symptomatic vs asymptomatic	
Standardised definitions from published criteria used with symptoms included in definition	
No clear statement	0 stars
2. Follow-up	
Complete follow-up	1 star

<20% of participants per annum in follow-up did not get symptom assessment OR description provided for those lost, making introduction of bias unlikely	
Follow-up rate <80% and no description of those lost	0 stars
No statement	

Supplementary Table 2: Extracted studies

	Population	Study Type	Dates of study	Imaging Assessment	Microbiological Assessment	Symptom Assessment	Follow up methods	Transition*	Proportion making transition (time, months)
1. Alling(1) (1955) USA	• Diagnosis of moderate tuberculosis • Followed up within the study hospital • Sourced from referral from private physician or mass occupational radiographic surveys	Clinic/ Hospital/ Sanatorium Cohort	1938 - 1948	Chest X-ray (CXR) reviewed according to NTA† criteria with double, independent reading. All "moderate" tuberculosis at baseline	Sputum sent for unknown method. All deemed "arrested" in this data row.	Delineation of symptom status is not well defined but all individuals in this data row were deemed "arrested"	Retrospective collection of routinely collected clinical data from their hospital. Follow up was therefore not systematically done. Definitions of disease groups required repeat CXR and sputum testing over periods of time.	cxr.pos | micro.neg | sympt.unk [arrested‡] to cxr.pos | micro.pos | sympt.unk	• 8/58 (60) • 10/58 (60-156)
2. Anastasatu(2) (1985) Romania	Smear negative with minimal or moderately advanced pulmonary lesions	Control/ Placebo arm	Unknown	CXR	Sputum smear and culture	Not specified but described to be "without radio-clinical symptoms of activity"	Follow up details not clear but authors describe follow up to 24 months, with sputum testing being performed	cxr.pos | micro.neg | sympt.neg to cxr.pos | micro.pos | sympt.unk	6/143 (24)
3. Aneja(3) (1979)	• Symptomatic for TB and abnormal photofluorogram	Control/ Placebo arm	1968 - 1972	CXR that was double-read ± a third umpire	Sputum for microscopy and culture	Review of clinical symptoms e.g. cough	Repeat CXR and spot sputum for microscopy and	cxr.pos | micro.neg | sympt.pos	21/110 (12)

India	• Negative sputum microscopy • No TB treatment, or less than 2 weeks of treatment prior to entry into trial • Well enough for ambulant care			in case of discrepancies			culture were done at the end of the 2nd, 4th, 6th, 9th and 12th months of follow up	to cxr.pos | micro.pos | sympt.unk	
4. Beeuwkes(4) (1942) USA	Household contacts with at least 6 months of follow up data	Household Contact Study	1933-1938	CXR	72 hour concentrated sputum sample for microscopy. Some also had culture and/or inoculation.	Based on medical history and examination to have been "asymptomatic"	Re-examined 6-12 monthly plus additional unscheduled visits if developing symptoms. Reviews included medical history and examination, chest x-ray and sputum sampling.	cxr.neg | micro.neg |symp.neg to cxr.unk | micro.pos | sympt.pos	1/784 (6-60)
								cxr.neg | micro.neg | symp.neg to cxr.unk | micro.neg | sympt.pos	3/784 (6-60)
								cxr.pos | micro.neg | sympt.neg to cxr.unk | micro.pos | sympt.pos	3/79 (6-60)

								cxr.pos | micro.neg | sympt.neg to cxr.unk | micro.neg | sympt.pos	5/79 (6-60)
								cxr.pos |. micro.neg | sympt.neg to cxr.unk | micro.pos | sympt.pos	13/43 (6-60)
								cxr.pos | micro.pos | sympt.pos to cxr.unk | micro.neg | sympt.unk	10/28 (6-60)
5. Bobrowitz(5,6) (1947; 1949) USA	• Patients at study site with minimal disease on Chest X-ray • At least 6 months of follow up data	Clinic/ Hospital/ Sanatorium Cohort	1938 - 1945	CXR	Sputum sample or gastric concentrates or cultures (authors noted that gastric cultures became more frequent from 1939 onwards)	Symptoms and clinical review	Radiographic, mycobacteriological and clinical review during Sanatorium stay but intervals not clearly defined	cxr.pos | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	26/191 (60)

6. B**orgen(7,8)** (1951;1952) Norway	• Age: >15 years • Factory workers who participated in a repeat survey	Occupational/ Student Screening	1947 - 1949	Photofluorography: no lesion versus suggestive tuberculous lesion	It is unclear whether systematic sputum testing was done at baseline, irrespective of photofluorography results	Clinical review including symptoms such as loss of weight and signs such as pyrexia	Factory workers had repeat survey at 2 years follow up point. Included repeat photofluorography. Also included sputum testing but unclear as to whether this was performed in all or only those with radiographic abnormalities.	cxr.neg | micro.unk | sympt.unk to cxr.pos | micro.pos | sympt.pos	4/6884 (30)
								cxr.pos |. micro.neg | sympt.pos to cxr.pos | micro.pos | sympt.pos	2/24 (30)
								cxr.pos | micro.neg | sympt.neg to cxr.pos | micro.pos | sympt.pos	2/120 (30)
7. Breu(9) (1954) Germany	• Patients that were assessed as part of mass screening in Ludwigsburg, Germany • Demonstrated abnormal chest X-ray at baseline but micro negative	General Community Survey	1949 - 1952	Photofluorography, CXR, tomography	Sputum for culture if either symptomatic or CXR abnormality suggestive of TB (but not all patients were able to produce sputum)	Delineation of symptom status is not well defined. All patients had an ESR performed at baseline.	Followed up "regularly" but interval not specified. Follow up included CXR, CT, microbiology (for sputum, laryngeal swab, gastric lavage)	cxr.pos | micro.pos | sympt.unk to cxr.pos | micro.pos | sympt.unk	48/904 (3-48)

8. Cowie(10) (1985) South Africa	• New or enlarging apical lesions on 6-monthly Chest X-ray occupational screening • Apical lesion still present on repeat Chest X-ray 2 months later • Concentrated smear X3 (baseline) and culture X2 (2-months later) of sputum was negative for MTB • Employee at the Gold Mine	Occupational/ Student Screening	1979 - 1984	CXR with new or enlarging apical lesions	Early morning sputa for concentrated microscopy and culture	The authors do not comment on the clinical status/symptoms of the cohort	• 3-monthly CXRs for 3 years followed by 6-monthly until 5 year endpoint in the study • Sputum for smear or culture if there was any progression radiographically or if the original lesion was thought to suggest bacteriologically positive disease • Histology from bronchoscopy was also performed in some	cxr.pos | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	88/152 (3-58)
9. Downes(11) (1938) USA	• Resident of Cattaraugus County • Diagnosed with active TB at some stage during the cohort years	Clinic/ Hospital/ Sanatorium Cohort	1923 - 1935	CXR	Microscopy on sputum (to confirm certain definitions, sputum was also concentrated)	Met the criteria for 'active' disease which was based on symptoms and examination	Retrospective data collection of clinical records, sanatoria outcomes and self-reporting. Data collection included that of mycobacteriology, clinical status and imaging.	cxr.pos | micro.pos | sympt.pos to cxr.pos | micro.neg | sympt.neg	• 27/342 (12) • 104/342 (24) • 140/342 (36) • 158/342 (48)

									• 171/342 (60)
10. Frimodt-Moller(12) (1965) India	• Abnormalities on Chest X-ray at baseline (and deemed probably tuberculous aetiologygy by two reviewers) • Negative smear and culture from sputum and larryngeal swab	Control/ Placebo arm	1960-1961	CXR deemed probably tuberculous by two readers	Microscopy and culture of sputum and laryngeal swab	Delineation of symptom status is not well-defined	3-monthly follow-ups occurred over a 3 year period including CXR and microbiological assessment	cxr.pos | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	• 11/86 (12) • 18/86 (24) • 25/86 (36)
11. Hong Kong Chest Service(13–16) (1979, 1981, 1981, 1984) HONG KONG	• Age: 15-75 years • Radiographic evidence of 'active pulmonary tuberculosis' by a Hong Kong Chest Physician (not those considered fibrotic/inactive) • No previous ATT • At least 5 sputum smears negative over approximately one week	Control/ Placebo arm	Not specified	CXR	Microscopy and culture of sputum X 5	Cough, sputum production, haemoptysis	• 1-2 sputum specimens for microscopy and culture at monthly reviews in the first year and then quarterly reviews up to 30 months • 20 CXRs done between baseline and month 60 • Clinical reports made at each sampling/imaging visit • Unscheduled additional reviews	cxr.pos | micro.neg | sympt.mixed to cxr.pos | micro.pos | sympt.unk	• 40/176 (3) • 49/176 (6) • 61/176 (12) • 67/176 (18) • 69/176 (24) • 70/176 (30) • 71/176 (36) • 71/176 (60)

							with imaging & sputum sampling occurred if deterioration suspected		
12. IUAT Committee on Prophylaxis (17) (1982) Europe	• Fibrotic lesions on Chest X-ray that had been stable during the year prior to study entry • Positive Mantoux • No previous ATT • Not previously mycobacteriologically positive and not culture positive at entry to the trial	Control/ Placebo arm	Unknown	CXR showing fibrotic lesions	2 Sputum cultures at entry	Not well-defined	• Annual reviews were done for 5 years including sputum culture, CXR and/or vital status was recorded • Records were also reviewed annually to see whether participants had been started on TB treatment	cxr.pos | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	97/6990 (12-60)
13. Lincoln(18) (1954) USA	• No history of prior diagnosis of TB • Now Chest X-ray suggestive of minimal TB • Followed up within the study hospital for at least one year • Sourced from referral from	Clinic/ Hospital/ Sanatorium Cohort	1937 - 1947	CXR reviewed according to NTA criteria with double, independent reading. All "minimal" tuberculosis at baseline in this data row.	Sputum sent for unknown method. 128/134 had sputum examination for tubercle bacilli. 112/128 had at least one "positive" sputum by some method.	The authors' definition of 'active' disease makes use of the NTA definitions which include symptom activity within their definitions. The majority of those with 'active' disease were	Retrospective collection of routinely collected clinical data from their hospital. Follow up was therefore not systematically done. Definitions of disease groups required repeat CXR	cxr.pos | micro.mixed | sympt.pos to cxr.pos | micro.neg | sympt.unk	• 45/134 (24) • 71/134 (36) • 80/134 (48) • 83/134 (60) • 86/134 (72) • 87/134 (84)

	private physician (due to symptoms or a known TB contact) or from mass occupational radiographic surveys					also noted to have self-presented for care rather than having being picked up during general community screening.	and sputum testing over periods of time.	cxr.pos | micro.neg | sympt.unk [arrested‡] to cxr.pos | micro.pos | sympt.unk	• 15/314 (24) • 25/314 (36) • 32/314 (48) • 35/314 (60) • 36/314 (72)
14. Manser(19) (1953) Switzerland	• Patients from the study sanatorium's records • Over 60 years of age	Clinic/ Hospital/ Sanatorium Cohort	1941 - 1951	CXR	Sputum or gastric lavage for unspecified method	Symptom questioning, fever, temperature, weight; but the authors do not present breakdown of data	Frequency of follow up was not specified, but included physical exam, weight, ESR, CXR/CT, and sputum or gastric lavage	cxr.pos | micro.pos | sympt.unk to cxr.pos | micro.neg | sympt.unk	15/40 (6)
15. Marshall(20) (1948) UK	• Acute bilateral pulmonary tuberculosis of presumably recent origin • Old-standing disease and disease with thick-walled cavities excluded • Bacteriologically proven • Unsuitable for	Control/ Placebo arm	1947 - 1948	CXR	Sputum (direct smear and culture) ± laryngeal swab or gastric lavage if unable to produce (for culture)	General condition, temperature, weight	• All remained admitted to the sanatorium for 6 months and outcomes were assessed at this point • Follow up included monthly clinical reviews • CXR (by blinded radiologist) and	cxr.pos | micro.pos | sympt.pos to cxr.pos | micro.neg | sympt.unk	1/52 (6)

	collapse therapy • 15 to 30 years of age						mycobacterial investigations done by clinical team were recorded		
16. National Tuberculosis Institute(21–28) (1974; 1976; 1976; 1976; 1978; 1978; 1978; 1978; 1982) India	• Randomly selected households in Bangalore district entered into a mass survey • Excluded those with BCG scars	General Community Survey	1961 - 1968	CXR (this data row: normal or "considerred to be of nontuberculous etiology")	Sputum (spot & overnight) for microscopy and culture in those with abnormal CXR or CXR not interpretable	The authors do not comment on the clinical status/symptoms of the cohort	Surveyed every 18-24 months including: • Photofluorography • Two sputum specimens if any history of abnormal CXR or CXR not interpretable	cxr.neg | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	• 44/31490 (18) • 99/17936 (60)
				CXR (this data row: "suggestive of tuberculous etiology")				cxr.pos | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	• 23/329(18) • 36/271 (60)
				CXR				cxr.pos | micro.pos | sympt.unk to cxr.pos | micro.neg | sympt.unk	• 86/269 (18) • 70/178 (36)
17. Norregaard (29) (1990)	• Abnormal Chest X-ray • Sputum smear-negative	Control/ Placebo arm	1978 - 1985	CXR	Sputum or gastric lavage for microscopy and culture X 6	Medical history: e.g. cough, haemoptysis, fever	CXR and culture of sputum or gastric lavage monthly for the first three	cxr.pos | micro.neg | sympt.mixed to	6/28 (48)

Denmark							months, followed by every second month for six months and annually thereafter for at least three years. In this data row: "slight or no symptoms" were deemed asymptomatic.	cxr.pos | micro.pos | sympt.neg	
								cxr.pos | micro.neg | sympt.mixed to cxr.pos | micro.pos | sympt.pos	2/28 (48)
18. Okada(30) (2012) Cambodia	• >10 years of age • Living in survey area • Two culture-negative specimens	General Community Survey	2002 - 2004	CXR that was deemed "TB suggestive"	Sputum for smear and solid culture X 2	Medical history and examination	Reviewed TB registers for the development of incident TB cases	cxr.pos | micro.neg | sympt.neg to cxr.pos | micro.pos | sympt.pos	5/309 (1-24)
							Conducted a repeat review at 2 years of all those with abnormal CXR at baseline, including medical history, CXR and two sputum specimens for smear and culture	cxr.pos | micro.neg | sympt.neg to cxr.pos | micro.pos | sympt.unk	46/309 (24)
				CXR that was deemed to be normal (n=20407) or to have an			Reviewed TB registers for the development of incident TB cases	cxr.neg | micro.neg | sympt.neg to	32/21580 (1-24)

				abnormality that was not suggestive of TB (n=1173)				cxr.pos | micro.pos | sympt.unk	
				CXR that was deemed "TB suggestive"			Conducted a repeat review at 2 years of all those with abnormal CXR at baseline, including medical history, CXR and two sputum specimens for smear and culture	cxr.pos | micro.neg | sympt.neg to cxr.neg | micro.neg | sympt.unk	26/309 (24)
19. Orrego Puelma(31) (1945) CHILE	• Abnormal CXR of minimal extent • Had at least 2 years of follow up data available for retrospective review	Clinic/ Hospital/ Sanatorium cohort	Unknown	CXR	Sputum and/or gastric washings for microscopy, culture and/or inoculation	Cough, loss of weight, haemoptysis, back pain	Repeat CXR and sputum testing were done at unspecified intervals over the two years	cxr.pos | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	18/67 (24)
20. Pamra(32) (1971) India	• no history of previous treatment for Tb • No manifest tuberculous lesion in any organ other than the lungs • No evidence of diabetes or any other non-tuberculous	Control/ Placebo arm	1958 - 1968	CXR	At least 2 sputum and laryngeal swab specimens smear and culture negative	Not well-described	• Laryngeal swab culture and CXR were done 3-monthly during year 1 • Laryngeal swab culture and CXR were done 6-monthly during year 2-5	cxr.pos | micro.neg | sympt.neg to cxr.pos | micro.pos | sympt.pos	2/178 (72)
								cxr.pos | micro.neg | sympt.neg to	55/178 (72)

	disease • Apparent radiological stability of the lesions during the period of observation before inclusion in the study, this period being not less than 3 months or more than 3 years • Constantly negative results on direct smear and culture examination of at least 2 sputum and laryngeal swab specimens during the period of observation							cxr.pos | micro.pos | sympt.neg	
21. Puffer(33) (1945) USA	• Following up at study clinic at the time of study • Deemed to have "reinfection type" Tuberculosis	Clinic/ Hospital/ Sanatorium cohort	1931 - 1943	CXR (in this data row, all deemed to have apical disease)	As per Opie criteria: "Arrested" or "Apparently arrested" is diagnosed when all symptoms have disappeared for	Review of "physical signs and/or symptoms"	Outcome for all study participants was assessed at 1 Jan 1943 irrespective of when they started in the study. It is not clear if follow up review was systematic or	cxr.pos | micro.neg | sympt.neg to cxr.pos | micro.pos | sympt.pos	9/261 (62)
				CXR				cxr.pos | micro.neg | sympt.neg [arrested‡]	10/267 (62)

					a period of at least 6 months and the physical signs are those of a healed lesion, the disease is regarded as arrested or, when some uncertainty still exists, the designation "apparently arrested" has been regarded as preferable. When the patient first comes under observation of the clinic, arrested lesions are not infrequently recognisable only by radiological examination, but are separable from "latent lesions" because they are known		only based on routinely collected clinical data. In order to meet study definitions, follow up would have had to meet NTA criteria.	to cxr.pos | micro.pos | sympt.pos	
				CXR				cxr.pos | micro.pos | sympt.pos to cxr.pos | micro.neg | sympt.neg	92/384 (62)

					to have passed through a period of manifest disease with the usual symptoms of pulmonary tuberculosis. When, however, a tuberculous pulmonary lesion, recognised by roentgenographic examination, has been accompanied by no symptoms of the disease discoverable by inquiry, but is associated with physical signs of a healed lesion, such as impaired resonance or diminished breath sound, it should be classified as arrested tuberculosis"				

22. Sikand(34) (1959) India	Employed as police officer in study setting	Occupational/ Student Screening	1952 - 1958	CXR deemed 'inactive' tuberculosis or tuberculosis of 'doubtful activity'	Sputum or laryngeal swab for unknown method	Delineation of symptom status is not well defined	• CXR repeated at two timepoints: o 17 months and o 5-6 years • Indications and regularity of mycobacteriological testing not clearly described	cxr.pos | micro.neg | sympt.unk to cxr.pos | micro.pos | sympt.unk	38/319 (12)
				CXR deemed 'Negative' for TB				cxr.neg | micro.unk | sympt.unk to cxr.pos | micro.pos | sympt.unk	89/11268 (17-69)
								cxr.neg | micro.unk | sympt.unk to cxr.pos | micro.neg | sympt.unk	251/11268 (17-69)
23. Styblo(35) (1967) Czechoslovakia	• 15 years and older • Resident of study district	General Community Survey	1961 - 1965	CXR with two blinded, independent reviewers	Microscopy of sputum and culture of sputum/laryngeal swabs on 1-3 occasions in all with symptoms or abnormal CXR and in a random	Clinical History	• Reviewed monthly if hospitalised • If ambulatory, CXR annually ± bacteriology if symptomatic or CXR changes	cxr.neg | micro.neg | sympt.neg to cxr.unk | micro.pos | sympt.pos	• 17/73000 (1-12) • 33/73000 (1-24) • 49/73000 (1-36)

					selection not meeting these criteria				• 66/73 000 (1-48)
								cxr.neg | micro.neg | sympt.neg to cxr.unk | micro.pos | sympt.unk	• 102/73000 (1-12) • 123/73000 (1-24) • 168/73000 (1-36) • 175/73000 (1-48)
24. Tuberculosis Society of Scotland(36,37) (1958, 1963) SCOTLAND	• Age: >15 years • European • No bacilli on sputum smear or culture • Abnormal CXR but of "doubtful activity" with no cavitation • No known active extrapulmonary disease • Not pregnant/ recently post-	Control/ Placebo arm	1954 - 1959	CXR	Sputum, gastric lavage or laryngeal swab for microscopy and/or culture	Authors noted that "chest radiograph was the only important sign of disease"	• 3-monthly reviews, including CXR, weight checks and ESR testing • Follow up sputum testing was done in all those able to produce a sputum sample; laryngeal swabs or gastric lavage was only done if	cxr.pos | micro.neg | sympt.neg to cxr.pos | micro.pos | sympt.unk	9/95 (24)

	partum • Not diabetic						deterioration suspected		

* Start and end points are labeled with three characteristics or states including a Chest Xray status (i.e. cxr.neg or cxr.pos or cxr.unk), a microbiological status (i.e. micro.neg, micro.pos, micro.unk, or micro.mixed), and a symptom status (i.e. sympt.neg, sympt.pos, sympt.unk or sympt.mixed). In these states, "neg" denotes negative, "pos" denotes positive, "unk" denotes unknown and "mixed" denotes a mixed group.

† The National Tuberculosis Association (NTA) developed criteria for assigning a clinical status to individuals with tuberculosis(38). These criteria and nomenclature changed over time but generally included groups such as "apparently cured", "arrested", "apparently arrested", "quiescent", and "active". The definition of these groups included radiographic, microbiological and clinical criteria over defined periods of time.

‡ This group were known to have been microbiologically positive on a prior occasion, then documented arrested disease, followed by a relapse

ATT=Antituberculosis Therapy; IUAT=International Union Against Tuberculosis; NTA=National Tuberculosis Association; UK=United Kingdom; USA=United States of America

Supplementary Table 3: Descriptors on imaging assessments

Author (Publication Year) Country	Imaging Assessment	Descriptor/quoted text from study
1. Alling (1955) USA	Inactive	"The clinical status of each person at the time of diagnosis was assessed on the basis of roentgenographic and bacteriologic data collected during the first six months following diagnosis…according to the Diagnostic Standards (1940 edition) published by the National Tuberculosis Association (NTA)." "Arrested: For six months or more prior to the anniversary, the patient's sputum has been free of acid-fast bacilli and his serial chest roentgenograms have been compatible with stable disease. When more than a year has elapsed between clinic examinations and the chest roentgenograms show a stable lesion, the patient is continued in the arrested category throughout the interim."
2. Anastasatu (1985) Romania	Inactive	"A group with an X-ray lesion extension smaller than 10cm^2 and with no epidemiological risk [and]… no "radiological activity"
3. Aneja (1979) India	Active	"They had an abnormal shadow on a chest photofluorogram, read as pulmonary tuberculosis (TBP) by the Medical Officer" "suspected to have active tuberculosis on X -ray evidence on single 70mm photofluorogram"
4. Beeuwkes (1942) USA	Mixed	"Latent (asymptomatic) lesions of first infection or childhood type were designated as calcified nodules, and latent infiltrations, childhood; latent lesions of reinfection or adult type, as latent apical tuberculosis." "Latent (asymptomatic) infiltrations of either the childhood or adult type could be active (progressive or retrogressive), but they were classified as asymptomatic until the patient had symptoms, or physical signs of disease were demonstrated. At that time the diagnosis was changed to manifest (clinical) disease."
	Active	Manifest (clinical) tuberculosis, that which gives rise to physical signs or symptoms or both, was designated as childhood (first infection type) or adult (reinfection type) depending on the usual anatomical and clinical characteristics of the lesion
5. Bobrowitz (1947; 1949) USA	Mixed	"Only patients with minimal tuberculosis according to the classification of the National Tuberculosis Association and with the characteristic roentgenological picture of pulmonary tuberculosis of the reinfection type were acceptable. These cases were further subdivided into four groups, according to the size of the infiltrations, In grouping the patients by type of lesion, we utilized the excellent definitions for the roentgenological character of the infiltration described by Reisner and Downes - Exudative, Productive and fibrotic, Exudative-productive, Fibro-calcific form"
6. B**orgen** (1951;1952) Norway	Active	"Cases submitted to control examination [were] grouped according to roentgenographic and clinical findings into 10 categories including: Active non-destructive tuberculous pulmonary lesions - This group includes cases with pulmonary infiltrates which, according to the roentgenographic findings (soft, indistinct

		parenchymal lesions) and clinical findings (elevated temperature, increased sedimentation rate, loss of weight, etc.) were considered active lesions but where direct smears or culture failed to reveal tubercle bacilli in the sputum and after an observation period of 1.5-3.5 years no other etiology could reasonably be assumed."
	Inactive	"Inactive pulmonary lesions - This group includes cases with clear-cut, dense, streak-formed and, sometimes calcified lesions. In these cases the process showed no signs of activity during the observation period"
7. Breu (1954) Germany	Mixed	"Bewusst beschränke ich mich bei der Aufzeigung der Entwicklungsreihen aus der geschlossenen in die bakteriologisch offene Tuberkulose, bzw. auf die angeführten röntgenologischen Erscheinungsformen (=Ia-b-Fälle der Fürsorgestatisik)...";"nur [...] als Ic/IIa angesprochenen Tuberkulosefälle [...], die sich später zu einer ansteckenden Lungentuberkulose entwickelten." *Intentionally, in presenting the progression from closed to bacteriological open tuberculosis, meaning the roentgenological appearances (Ia-b-cases of the register); only as Ic/IIa classified cases that later progressed to a contagious tuberculosis. [According to old German tuberculosis classification, where Ia/b open, active on xray and contagious with or without bacteriology; Ic open, active but not contagious, with fuzzy appearance on xray; IIa closed, inactive with clear demarcation of lesion on xray]*
8. Cowie (1985) South Africa	Active	"New or enlarging apical lung lesions detected by routine 6-monthly chest radiography"
9. Downes (1938) USA	Active (Also micro pos)	Retrospectively captured imaging findings, based on clinical records
10. Frimodt-Moller (1965) India	Active	Cases showing lung pathology on the survey films according to two independent readers were referred for bacteriological examination in the field with laryngeal swab cultures and sputum collection for microscopy and sputum culture as well as for having another 70 mm X-ray film taken. The new film was compared with the former and cases thought to be of a non-tuberculous aetiology, or had only calcified lesions, were excluded. Other cases were ranked according to extent of the parenchymal lesion The pretreatment films of the cases have for the purpose of this report been re-read and classified by the senior author with regard to extent of disease and cavitation according to the pattern of classification used by Fox and Sutherland (1956). The assessor did not know what treatment had been given or the fate of the individual cases studied.
11. Hong Kong Chest Service (1979(13), 1981(14), 1981(15), 1984(16)) HONG KONG	Active	"Patients admitted to the study…were diagnosed at the regular routine meetings of physicians of the Hong Kong Chest Service as having radiologically active pulmonary tuberculosis, not previously treated"
12. IUAT Committee on Prophylaxis (1982) Europe	Inactive	"For the purpose of entry to the trial, fibrotic lesions were defined as well-delineated radiographic lesions of probable tuberculous origin, usually in the upper half of the lung, which had been stable during the year prior to entry"

13. Lincoln (1954) USA	Inactive	"A review was made of all original diagnoses of minimal pulmonary tuberculosis made by the hospital staff during the years 1937 through 1947. In no case had a diagnosis of pulmonary tuberculosis been previously made" "Two physicians, independently of one another and without knowledge of the subsequent clinical course, classified the diagnostic chest roentgenogram of each case according to the presence of pulmonary tuberculosis and the stage of disease. The classification of the stage of disease was based on the Diagnostic Standards (1940 edition), published by the National Tuberculosis Association (NTA). In the few cases in which there was disagreement, consultation was held to effect a mutually acceptable decision." "Arrested: For six months or more prior to the anniversary, the patient's sputum has been free of acid-fast bacilli and his serial chest roentgenograms have been compatible with stable disease. When more than a year has elapsed between clinic examinations and the chest roentgenograms show a stable lesion, the patient is continued in the arrested category throughout the interim."
14. Manser (1953) Switzerland	Active (Also micro pos)	Retrospectively captured imaging findings, based on clinical records
15. Marshall (1948) UK	Active (Also micro pos)	Study included patients with acute bilateral pulmonary tuberculosis of presumably recent origin, and patients who were unsuitable for collapse therapy. Old-standing disease and disease with thick-walled cavities excluded. Xrays were reviewed by a blinded radiologist.
16. National Tuberculosis Institute (1974; 1976; 1976; 1976; 1978; 1978; 1978; 1978; 1982) India	Active	"Suggestive of tuberculous aetiology judged to be possibly, probably or definitely active tuberculous"
17. Norregaard (1990) Denmark	Active	"The admission criteria were usually small X-ray shadows compatible with active pulmonary tuberculosis in smear negative patients. This evaluation was made at conference by the consultants…. All patients had infiltrates in the upper lobe of one or both lungs"
18. Okada (2012) Cambodia	Active	"A panel of experts composed of two respiratory physicians and/or radiologists made a decision on the final radiological findings. Based on the Japanese CXR classification for TB" "A CXR suggestive of active TB was categorised as 'TB-suggestive CXR'"
19. Orrego Puelma (1945) CHILE	Mixed	"Minimum or superficial lesions without apparent cavity, limited to a small area in one or both lungs. The total involvement, ignoring its distribution, should not exceed the equivalent in volume of the lung tissue which lies above the second chondrosternal junction and the spine of the fourth or body of the fifth thoracic vertebra on one side." "Furthermore, we studied only those cases which had two years' evolution, at least."
20. Pamra (1971) India	Mixed	(a) No history of previous treatment for tuberculosis. (b) No manifest tuberculous lesion in any organ other than the lungs. (c) No evidence of diabetes or any other non-tuberculous disease.

		(d) Apparent radiological stability of the lesions during the period of observation before inclusion in the study, this period being not less than 3 months or more than 3 years.
21. Puffer (1945) USA	Mixed	"persons with lesions demonstrable by X-ray examination had no physical signs nor symptoms and gave no history of illness indicative of tuberculosis, they are classed as latent apical cases" "The classification of lesions as minimal, moderately advanced and far advanced is in accordance with that recommended by the National Tuberculosis Association."
	Inactive	"manifest cases were believed to be arrested at the time of first examination in this clinic." "The classification of lesions as minimal, moderately advanced and far advanced is in accordance with that recommended by the National Tuberculosis Association."
22. Sikand (1959) India	Inactive	Inactive Pulmonary Lesions (i) Sputum negative by all methods or L. S. negative by culture; (ii) Hardish shadows in the skiagrams; (iii) No clinical or laboratory evidence suggestive of activity.
	Mixed	Lesions of Doubtful Activity (i) Small soft infiltrations, stationary of regressive but not progressive; (ii) Primary type of tuberculosis; (iii) Known pulmonary cases under successful therapy.
23. Styblo (1967) Czechoslovakia	N/A	Xray negative group
24. Tuberculosis Society of Scotland (1958, 1963) SCOTLAND	Mixed	"Patients were accepted for the trial if the chest radiograph showed an abnormality considered to be tuberculous, this being the only important sign of disease" "certain conditions also excluded acceptance- ... cavitation demonstrable on an ordinary posteroanterior radiograph" "The clinicians were asked to classify their cases as 'acute' without evidence of fibrosis, or 'chronic', with fibrosis present."

Supplementary Table 4: Descriptors of cohorts included in the Systematic Review

	n (%) or median (IQR)
Cohort Types, n=34	
• Clinical Cohort	9 (24.5)
• Control arm of therapeutic intervention	9 (24.5)
• General Community Survey	7 (20.6)
• Household Contact Study	4 (11.8)
• Occupational or Student Screening	5 (14.7)
Study Location, n=34	
• Africa	1 (2.9)
• Asia	11 (32.4)
• Europe	10 (29.4)
• South America	1 (2.9)
• North America	11 (32.4)
Status at entry to cohorts, n=34	
• Normal Chest X-ray/imaging	4 (11.8)
• Abnormal Chest X-ray/imaging	30 (88.2)
○ Abnormal Chest X-ray/imaging, where the authors state that the extent of disease was active	15 (44.1)
○ Abnormal Chest X-ray/imaging, where the authors state that the extent of disease was inactive/fibrotic	9 (26.5)
○ Abnormal Chest X-ray/imaging, where the authors state that the extent of disease was "mixed" in this group or they did not describe/specify	6 (17.6)
• Microbiological status† negative	28 (82.4)
• Microbiological status† positive	6 (17.6)
• Microbiological status† mixed or unknown	0 (0.0)
• Symptomatic‡	7 (20.6)
• Asymptomatic‡	11 (32.4)

• Symptom status† mixed or unknown	16 (47.1)
Demographics, n=34	
• Sex distribution	
○ Equal distribution of males and females	5 (14.7)
○ More females	8 (23.5)
○ More males	11 (32.4)
○ Unknown sex distribution	10 (29.4)
• Age distribution	
○ Minimum	5 years
○ Maximum	83 years
○ No details on age distribution specified	10 (29.4)
Local Burden of TB Disease at time of cohort§, n=34	
• Low	0 (0.0)
• Mid	2 (5.9)
• High	9 (26.5)
• Not specified or unknown	23 (67.6)

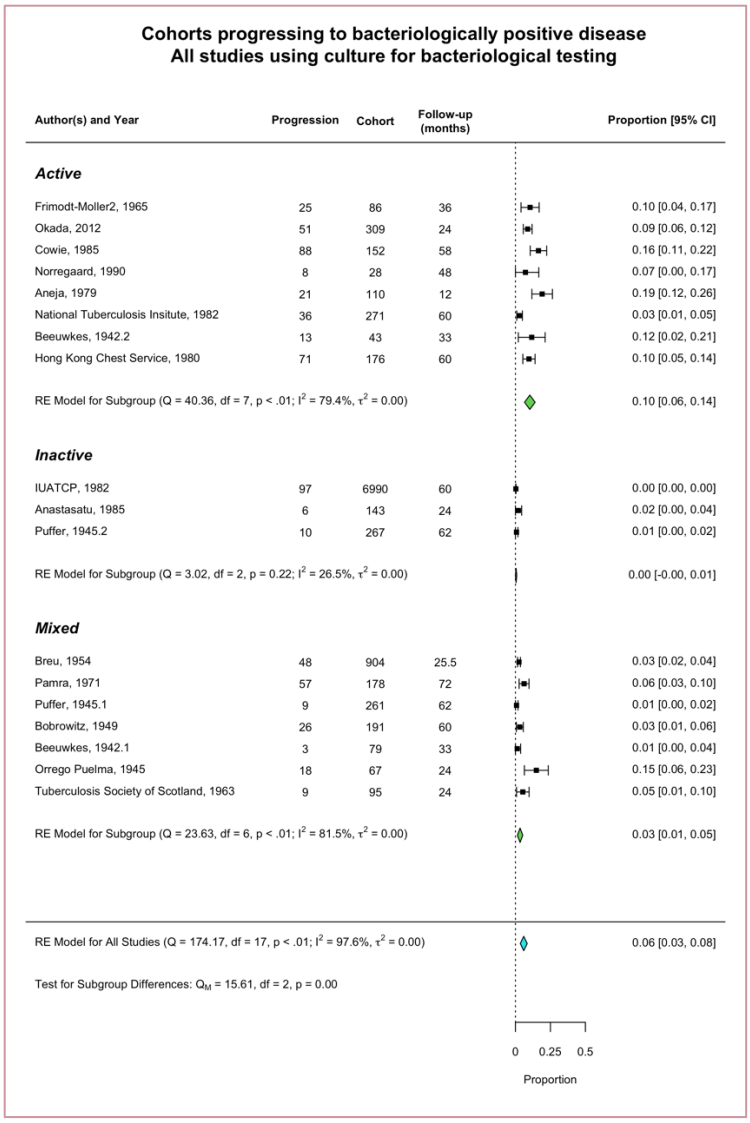
* Studies were able to contribute ≥1 cohort

† To establish microbiological status, authors used a variety of methods which included sputum (spot or extended collection periods e.g. overnight), laryngeal swabs, and gastric aspirates

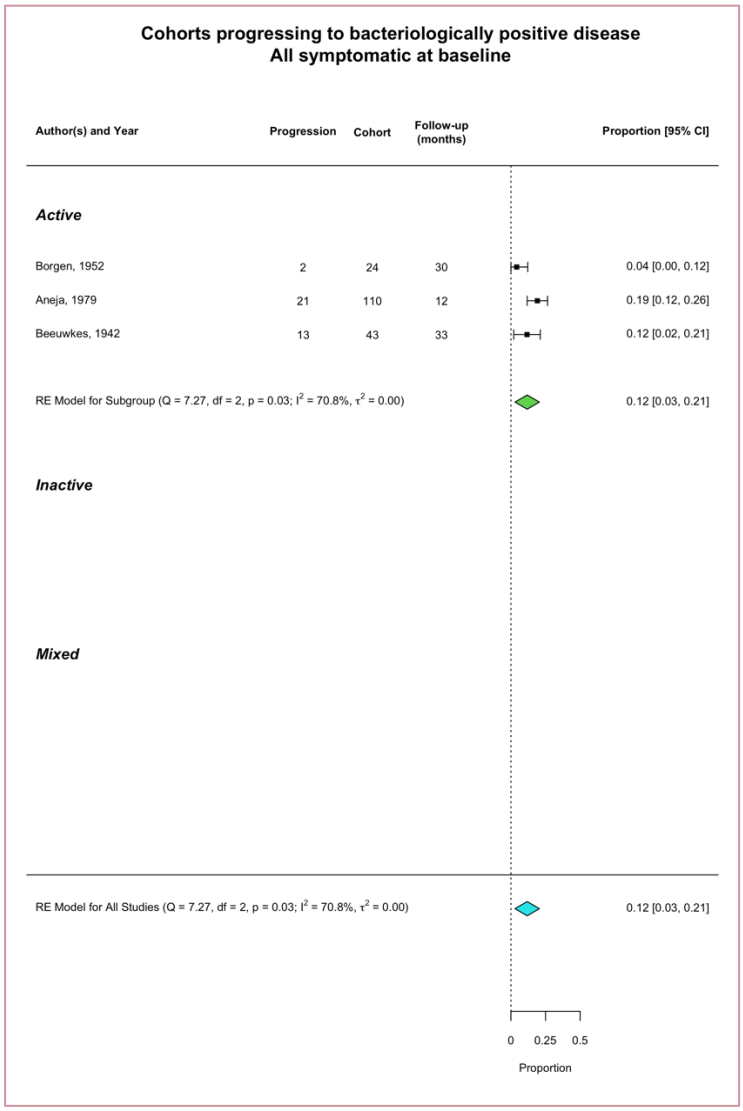
‡ Symptom status was as per author classification

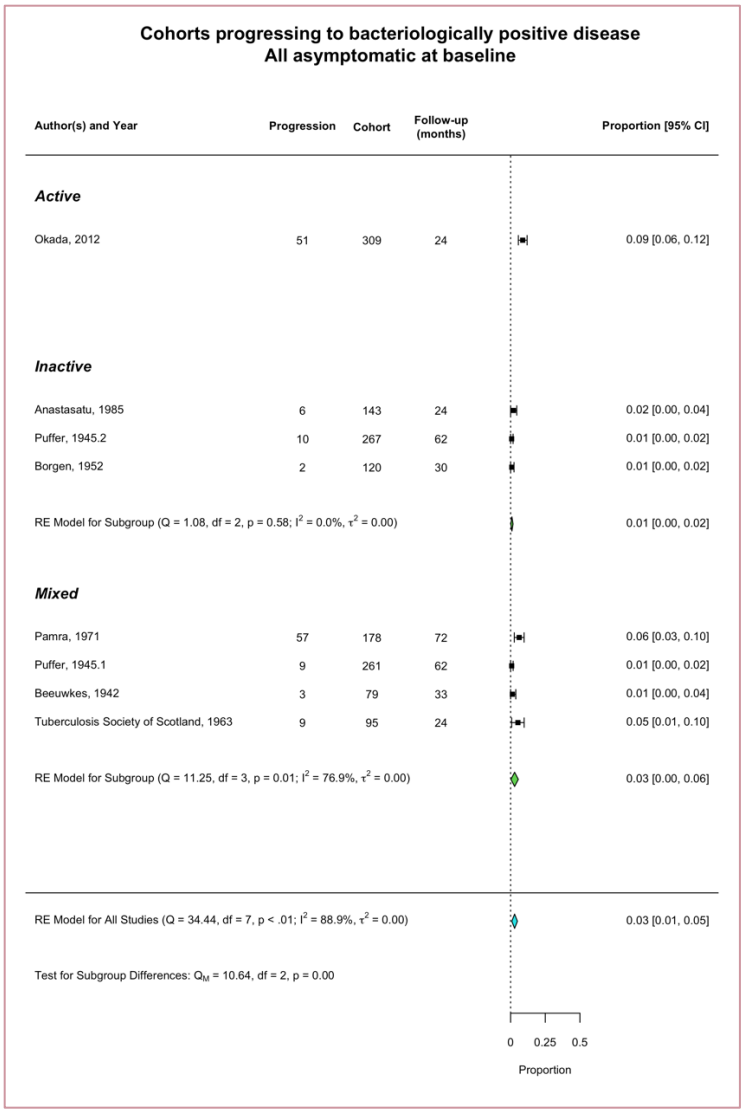
§ As per World Health Organisation stratification

IQR = Interquartile Range; TB = Tuberculosis

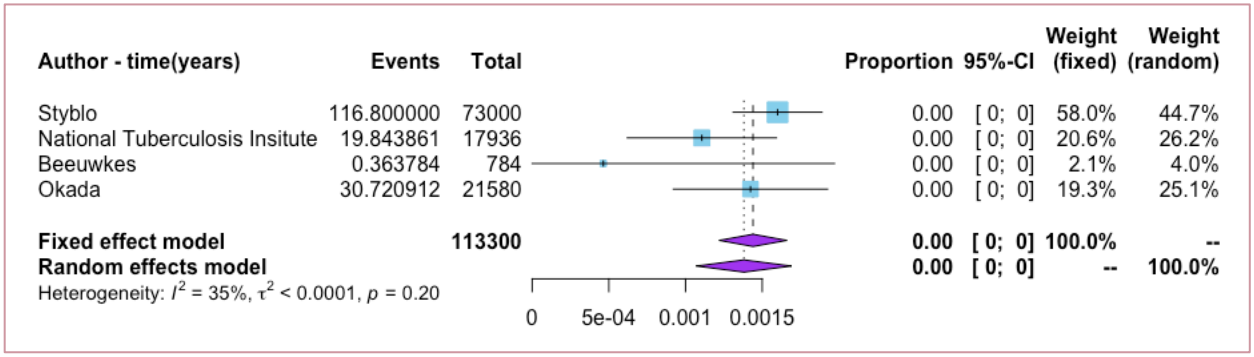


Supplementary Figure 1: Micro negative to positive, in those studies that used culture

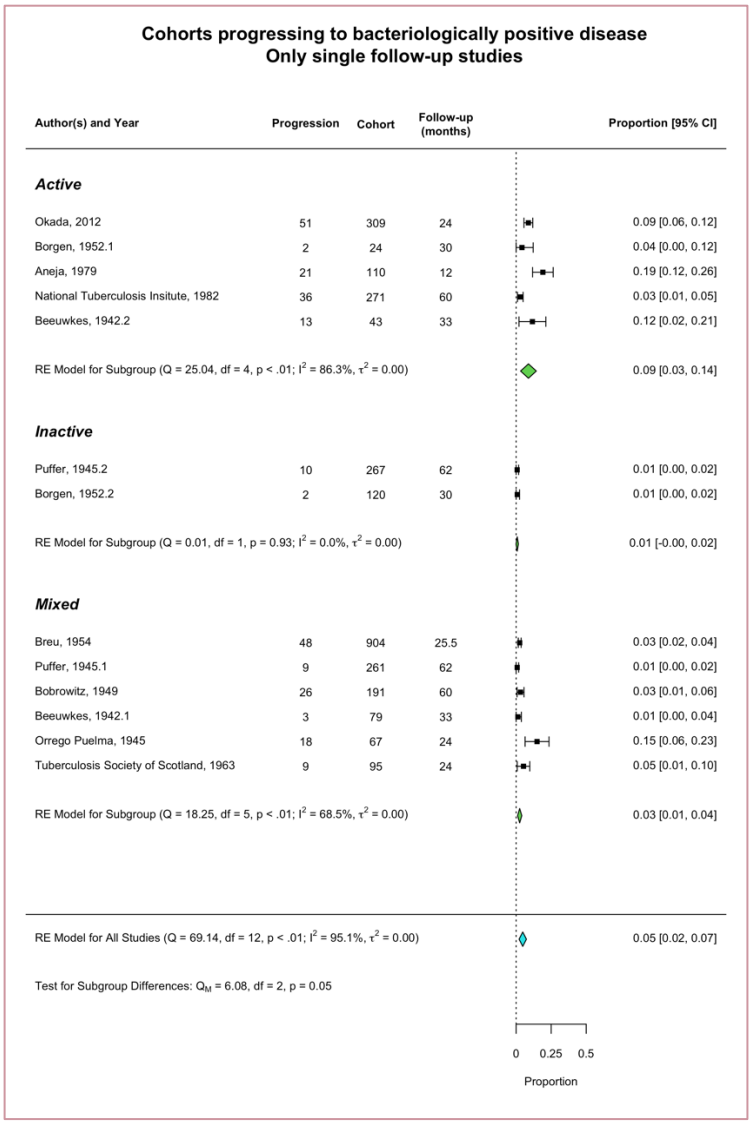

Cohorts progressing to bacteriologically positive disease
All symptomatic at baseline
Author(s) and Year
Progression
Cohort
Follow-up (months)
Proportion [95% CI]
Active
Borgen, 1952 2 24 30 0.04 [0.00, 0.12]
Aneja, 1979 21 110 12 0.19 [0.12, 0.26]
Beeuwkes, 1942 13 43 33 0.12 [0.02, 0.21]
RE Model for Subgroup (Q = 7.27, df = 2, p = 0.03; I² = 70.8%, τ² = 0.00) 0.12 [0.03, 0.21]
Inactive
Mixed
RE Model for All Studies (Q = 7.27, df = 2, p = 0.03; I² = 70.8%, τ² = 0.00) 0.12 [0.03, 0.21]
0
0.25
0.5
Proportion

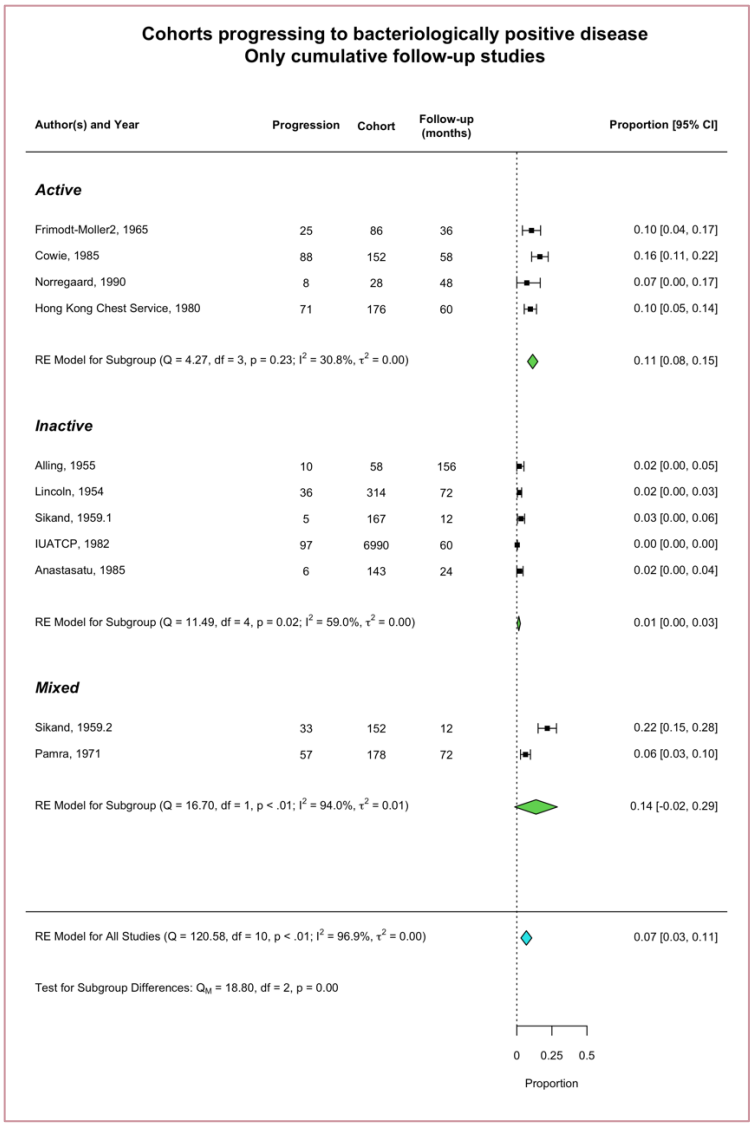


Supplementary Figure 2: Micro negative to positive, stratified by symptom and CXR status at baseline



Supplementary Figure 3: CXR negative and Micro negative, to Micro positive


Cohorts progressing to bacteriologically positive disease
Only single follow-up studies
Author(s) and Year
Progression
Cohort
Follow-up (months)
Proportion [95% CI]
Active
Okada, 2012 51 309 24 0.09 [0.06, 0.12]
Borgen, 1952.1 2 24 30 0.04 [0.00, 0.12]
Aneja, 1979 21 110 12 0.19 [0.12, 0.26]
National Tuberculosis Insitute, 1982 36 271 60 0.03 [0.01, 0.05]
Beeuwkes, 1942.2 13 43 33 0.12 [0.02, 0.21]
RE Model for Subgroup (Q = 25.04, df = 4, p < .01; I² = 86.3%, τ² = 0.00) 0.09 [0.03, 0.14]
Inactive
Puffer, 1945.2 10 267 62 0.01 [0.00, 0.02]
Borgen, 1952.2 2 120 30 0.01 [0.00, 0.02]
RE Model for Subgroup (Q = 0.01, df = 1, p = 0.93; I² = 0.0%, τ² = 0.00) 0.01 [-0.00, 0.02]
Mixed
Breu, 1954 48 904 25.5 0.03 [0.02, 0.04]
Puffer, 1945.1 9 261 62 0.01 [0.00, 0.02]
Bobrowitz, 1949 26 191 60 0.03 [0.01, 0.06]
Beeuwkes, 1942.1 3 79 33 0.01 [0.00, 0.04]
Orrego Puelma, 1945 18 67 24 0.15 [0.06, 0.23]
Tuberculosis Society of Scotland, 1963 9 95 24 0.05 [0.01, 0.10]
RE Model for Subgroup (Q = 18.25, df = 5, p < .01; I² = 68.5%, τ² = 0.00) 0.03 [0.01, 0.04]
RE Model for All Studies (Q = 69.14, df = 12, p < .01; I² = 95.1%, τ² = 0.00) 0.05 [0.02, 0.07]
Test for Subgroup Differences: Q_M = 6.08, df = 2, p = 0.05
0
0.25
0.5
Proportion



Supplementary Figure 4: Micro negative to positive, stratified by structure of follow up

Full Search Strategy: English

Filters to be applied within each:

- Languages: English, German & “Unknown language” (latter possible in Pubmed)
- Years: 1895 - 1960

a) Pubmed, Medline and Old Medline

	Category	Search terms
ENGLISH	**Population:**	Pulmonary tuberculosis OR “Tuberculosis, Pulmonary”[Mesh] OR Incipient tuberculosis OR Phthisis OR Minimal tuberculosis OR Moderate tuberculosis OR Advanced tuberculosis
	Intervention:	Follow up studies OR Course OR Biological Evolution OR After-history OR Supervision OR “Early Diagnosis”[Mesh] OR “Epidemiological Study”[Mesh] OR Epidemiologic study OR Prognosis OR Prospective OR Longitudinal OR Roentgenographic Survey OR Radiography OR Observation OR Progression OR Progress OR Diagnosis

(((((((("tuberculosis, pulmonary"[MeSH Terms] OR ("tuberculosis"[All Fields] AND "pulmonary"[All Fields]) OR "pulmonary tuberculosis"[All Fields] OR ("pulmonary"[All Fields] AND "tuberculosis"[All Fields])) OR "Tuberculosis, Pulmonary"[Mesh]) OR (Incipient[All Fields] AND ("tuberculosis"[MeSH Terms] OR "tuberculosis"[All Fields]))) OR ("tuberculosis, pulmonary"[MeSH Terms] OR ("tuberculosis"[All Fields] AND "pulmonary"[All Fields]) OR "pulmonary tuberculosis"[All Fields] OR "phthisis"[All Fields])) OR (Minimal[All Fields] AND ("tuberculosis"[MeSH Terms] OR "tuberculosis"[All Fields]))) OR (Moderate[All Fields] AND ("tuberculosis"[MeSH Terms] OR "tuberculosis"[All Fields]))) OR (Advanced[All Fields] AND ("tuberculosis"[MeSH Terms] OR "tuberculosis"[All Fields]))) AND ((((((((((((((("follow-up studies"[MeSH Terms] OR ("follow-up"[All Fields] AND "studies"[All Fields]) OR "follow-up studies"[All Fields] OR ("follow"[All Fields] AND "up"[All Fields] AND "studies"[All Fields]) OR "follow up studies"[All Fields]) OR Course[All Fields]) OR ("biological evolution"[MeSH Terms] OR ("biological"[All Fields] AND "evolution"[All Fields]) OR "biological evolution"[All Fields])) OR After-history[All Fields]) OR ("organization and administration"[MeSH Terms] OR ("organization"[All Fields] AND "administration"[All Fields]) OR "organization and administration"[All Fields] OR "supervision"[All Fields])) OR "Early Diagnosis"[Mesh]) OR ("epidemiologic studies"[MeSH Terms] OR ("epidemiologic"[All Fields] AND "studies"[All Fields]) OR "epidemiologic studies"[All Fields] OR ("epidemiologic"[All Fields] AND "study"[All Fields]) OR "epidemiologic study"[All Fields])) OR ("prognosis"[MeSH Terms] OR "prognosis"[All Fields])) OR ("longitudinal studies"[MeSH Terms] OR ("longitudinal"[All Fields] AND "studies"[All Fields]) OR "longitudinal studies"[All Fields] OR "prospective"[All Fields])) OR Longitudinal[All Fields]) OR (Roentgenographic[All Fields] AND ("surveys and questionnaires"[MeSH Terms] OR ("surveys"[All Fields] AND "questionnaires"[All Fields]) OR "surveys and questionnaires"[All Fields] OR

"survey"[All Fields])))) OR ("diagnostic imaging"[Subheading] OR ("diagnostic"[All Fields] AND "imaging"[All Fields]) OR "diagnostic imaging"[All Fields] OR "radiography"[All Fields] OR "radiography"[MeSH Terms])) OR ("observation"[MeSH Terms] OR "observation"[All Fields])) OR ("disease progression"[MeSH Terms] OR ("disease"[All Fields] AND "progression"[All Fields]) OR "disease progression"[All Fields] OR "progression"[All Fields])) OR Progress[All Fields]) OR ("diagnosis"[Subheading] OR "diagnosis"[All Fields] OR "diagnosis"[MeSH Terms])) AND (("1895/01/01"[PDAT] : "1960/12/31"[PDAT]) AND (Undetermined[lang] OR German[lang] OR English[lang]))

b) EMBASE and EMBASE CLASSIC

	Category	Search terms
ENGLISH	**Population:**	"Pulmonary tuberculosis" OR "Incipient tuberculosis" OR Phthisis OR Minimal tuberculosis OR Moderate tuberculosis OR Advanced tuberculosis
	Intervention:	"Follow-up stud*" OR "Follow up stud*" OR Course OR "Biological Evolution" OR "After-history" OR Supervision OR Epidemiological Stud* OR Epidemiology OR Prognosis OR Prospective OR Longitudinal OR "Roentgenographic survey" OR Radiography OR Observation* OR Progress* OR Diagnosis

c) Web of Science

	Category	Search terms
ENGLISH	**Population:**	"Pulmonary tuberculosis" OR "Incipient tuberculosis" OR Phthisis OR Minimal tuberculosis OR Moderate tuberculosis OR Advanced tuberculosis

	Intervention:	"Follow up stud*" OR "Follow-up stud*" OR "Course" OR "Biological Evolution" OR "After-history" OR Supervision OR "Epidemiological Stud*" OR Epidemiology OR Prognosis OR Prospective OR Longitudinal OR "Roentgenographic Survey" OR Radiography OR Observation* OR Progress* OR Diagnosis

Full Search Strategy: German

PUBMED

GERMAN	**Population:**	Pulmonale Tuberkulose OR Lungentuberkulose OR Lungen-Tuberkulose OR "Tuberkulose, pulmonal, pulmonale" [Mesh] OR anfangende Tuberkulose OR beginnende Tuberkulose OR Beginn der lungentuberkulose OR Fruehinfiltrat OR einsetzende Tuberkulose OR Schwindsucht OR Phthisische Entiwicklung* OR Initialherd* OR Tuberculosis inappercepta OR Spitzenschwielen OR Spitzenherd* OR begrenzte* OR Tuberkulinreihenpruefungen OR Lungenschwindsucht OR Entstenhung OR Entwicklung*
	Intervention:	Längsschnittstudie OR Longitudinalstudie OR Verlaufsstudie OR Verlauf OR Entwicklung OR Verlauf OR Nachverfolgung OR Beobachtung OR Frühdiagnose OR Früherkennung OR "epidemiologische Studie" [Mesh] OR Prognose OR prospektiv OR vorausschauend OR longitudinal OR längsschnitt OR Röntgen OR röntgenographisch OR radiographisch OR radiologisch OR Beobachtung OR Betrachtung OR Folge OR Entwicklung OR Progression OR Fortschreitung OR Verlauf OR Fortschritt OR Diagnose

EMBASE

GERMAN	**Population:**	"Pulmonale Tuberkulose" OR "Lungentuberkulose" OR "Lungen-Tuberkulose" OR "generalisierte Tuberkulose" OR "disseminierte Tuberkulose" OR "miliare Tuberkulose" OR "anfangende Tuberkulose"

		OR “beginnende Tuberkulose” OR “einsetzende Tuberkulose” OR Tuberk* OR Schwindsucht
	Intervention:	“Längsschnittstudie” OR “Längsschnitt-studie” OR “Longitudinalstudie” OR “Longitudinal-studie” OR “Verlaufsstudie” OR “Verlaufs-studie” OR Verlauf OR “Entwicklung” OR “Verlauf.” OR “Nachverfolgung” OR Beobachtung OR epidemiologische Studie OR Epidemiologie OR Prognose OR Prospektiv OR Vorausschauend OR longitudinal OR längsschnitt OR “Röntgen” OR “röntgenographische Studie” OR “röntgenographische Untersuchung” OR “radiologische Studie” OR “radiologische Untersuchung” OR Radiologie OR Röntgen OR Beobachtung* OR Fortschritt* OR Diagnose

Web of Science

GERMAN	**Population:**	“Pulmonale Tuberkulose” OR “Lungentuberkulose” OR “Lungen-Tuberkulose” OR “generalisierte Tuberkulose” OR “disseminierte Tuberkulose” OR “miliare Tuberkulose” OR “anfangende Tuberkulose” OR “beginnende Tuberkulose” OR “einsetzende Tuberkulose” OR Tuberk* OR Schwindsucht
	Intervention:	“Längsschnittstudie” OR “Längsschnitt- studie” OR “Longitudinalstudie” OR “Longitudinal-studie” OR “Verlaufsstudie” OR “Verlaufs-studie” OR “Verlauf” OR “Entwicklung” OR “Nachverfolgung” OR Beobachtung OR “epidemiologische Studie” OR Epidemiologie OR Prognose OR Prospektiv OR Vorausschauend OR longitudinal OR längsschnitt OR verlauf OR “Röntgen” OR “röntgenographische Studie” OR “röntgenographische Untersuchung” OR “radiologische Studie” OR “radiologische Untersuchung” OR Radiologie OR Röntgen OR Beobachtung* OR Fortschritt* OR Diagnose

PRISMA Checklist(39)

Section/topic	#	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.	3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-6
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6,9
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.	9
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.	9
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.	9
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	9 & Supplementary
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).	9-10
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.	10-11
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	11

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.	10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	10-11
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.	12

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	27
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	11
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	13 (figure 2)
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	16-18 & Supplementary
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	14
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 3,4 & Supplementary
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Figure 3,4 & Supplementary
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Supplementary figures
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	25

Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	26
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	25-27
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	12

References

1. Alling D, Bosworth E, Lincoln N. The after-history of pulmonary tuberculosis: V. Moderately Advanced Tuberculosis. Am Rev Tuberc. 1955;71(4):519–28.

2. Anastasatu C, Berceea O, Corlan E. Controlled clinical trial on smear negative, x-ray positive new cases, with the view to establishing if and how to treat them. Bull Int Union Tub. 1985;60:108–9.

3. Aneja KS, Gothi GD, Samuel GER. CONTROLLED STUDY OF THE EFFECT OF SPECIFIC TREATMENT ON BACTERIOLOGICAL STATUS OF 'SUSPECT CASES.' Indian J Tuberc. 1979;26:8.

4. Beeuwkes H, Hahn RG, Putnam P. A Survey of Persons Exposed to Tuberculosis in the Household1,. Am Rev Tuberc. 1942;29.

5. Bobrowitz ID, Hurst A, Martin M. Minimal Tuberculosis; the prognosis and clinical significance of a sanatorium treated group. Am Rev Tuberc. 1947;56(2):16.

6. Bobrowitz ID, Hurst A. Minimal Tuberculosis: Problems in Roentgenologic Interpretation. Radiology. 1949 Apr;52(4):519–32.

7. Borgen L, Meyer S, Refsum E. Mass photofluorography, tuberculin testing, and BCG vaccination in the district of Aker (Norway) 1947-49. Acta Tuberc Scand. 1951;25(4):327–55.

8. Borgen L., Meyer SN., Refsum E. A preliminary report of tuberculin tests and mass chest surveys carried out in Aker. Acta Tuberc Scand Suppl. 1950;26:159–68.

9. Breu K. [Public health x-ray diagnosis of closed pulmonary tuberculosis later proved contagious]. Beitrage Zur Klin Tuberk Spezifischen Tuberk-Forsch. 1954;111(5):437–44.

10. Cowie L, Langton ME. Diagnosis of sputum smear- and sputum culture-negative pulmonary tuberculosis. SAMJ. 1985;68:1.

11. Downes J. A Study of Mortality among Individuals with Active Pulmonary Tuberculosis. Milbank Mem Fund Q. 1938 Jul;16(3):304.

12. Frimodt-Moller J, Parthasarathy R, Thomas J. Results of treatment of non-bacillary tuberculosis in a domiciliary treatment prrogramme,-a preliminary report. 1960;

13. Hong Kong Chest Service. SPUTUM-SMEAR-NEGATIVE PULMONARY TUBERCULOSIS CONTROLLED TRIAL OF 3-MONTH AND 2-MONTH REGIMENS OF CHEMOTHERAPY. The Lancet. 1979 Jun;313(8131):1361–3.

14. Hong Kong Chest Service, Tuberculosis Research Centre, Mad, British Medical Research Council. A study of the characteristics and course of sputum smear-negative pulmonary tuberculosis. Tubercle. 1981 Sep;62(3):155–67.

15. Hong Kong Chest Service. A controlled trial of 2-month, 3-month, and 12-month regimens of chemotherapy for sputum smear-negative pulmonary tuberculosis: the results at 30 months. 1981;

16. Hong Kong Chest Service, Centre TR. A Controlled Trial of 2-Month, 3-Month, and 12-Month Regimens of Chemotherapy for Sputum-Smear-Negative Pulmonary Tuberculosis: Results at 60 months. Am Rev Respir Dis. 1984;6.

17. International Union Against Tuberculosis Committee on Prophylaxis. Efficacy of various duration of isoniazid preventive therapy for tuberculosis: five years of follow-up in the IUAT trial. Bull World Health Org. 1982;60(4):555–64.

18. Lincoln NS, Bosworth EB, Alling DW. The After-History of Pulmonary Tuberculosis - III: Minimal tuberculosis. Am Rev Tuberc. 1954;70(1):17.

19. Manser H. [Tuberculosis in aged and its course during sanatorium treatment]. Schweiz Z Tuberk Rev Suisse Tuberc Riv Svizzera Della Tubercolosi. 1953;10(2):65–82.

20. Medical Research Council. Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. BMJ. 1948 Oct 30;2(4582):769–82.

21. Chakaraborty, A K, Singh, H, Srikantan, K, Rangaswamy, K R, Krishnamurthy, MS, Stephen, J A. Tuberculosis in a rural population of South India: Report on five surveys. Indian J Tuberc. 1982;29(3):153–67.

22. Chakraborty, A K, Gothi, G D, Dwarkanath S, Singh H. Tubercculosis mortality rate in a South Indian rural population. Indian J Tuberc. 1978;25(3).

23. Krishnamurthy V, Nair S, Gothi G. A comparison of new cases (incidence cases) who had come from different epidemiological groups in the population. Indian J Tuberc. 1978;25(3).

24. Gothi GD, Chakaraborty AK, Jayalakshmi MJ. INCIDENCE OF SPUTUM POSITIVE TUBERCULOSIS IN DIFFERENT EPIDEMIOLOGICAL GROUPS DURING FIVE YEAR FOLLOW UP OF A RURAL POPULATION IN SOUTH INDIA. Indian J Tuberc. 1978;25(2):9.

25. Krishnamurthy VV, Nair SS, Gothi GD, Chakraborty AK. Incidence of Tuberculosis among newly infected population and in relation to the duration of infected status. Indian J Tuberc. 1976;23(1).

26. Gothi, G D, Chakraborty, A K, Ganapathy, K.T. Five year incidence of tuberculosis and crude mortality in relation to non-specific tuberculin sensitivity. Indian J Tuberc. 1976;23(2).

27. Chakraborty AK, Gothi GD. RELAPSES AMONG NATURALLY CURED CASES OF PULMONARY TUBERCULOSIS. Indian J Tuberc. 1976;23(1):6.

28. Chakraborty AK. INTERPRETATION OF PHOTOFLUOROGRAMS OF ACTIVE PULMONARY TUBERCULOSIS PATIENTS FOUND IN EPIDEMIOLOGICAL SURVEY AND THEIR FIVE YEAR FATE. Indian J Tuberc. 1974;21:8.

29. Norregaard J. Abacillary pulmonary tuberculosis. Tubercle. 1990;71:35–8.

30. Okada K, Onozaki I, Yamada N, Yoshiyama T, Miura T, Saint S, et al. Epidemiological impact of mass tuberculosis screening: a 2-year follow-up after a national prevalence survey. :7.

31. Orrego Puelma H, Grebe G. Analysis of One Hundred Cases of Minimal Pulmonary Tuberculosis. Dis Chest. 1945 Sep;11(5):375–9.

32. Pamra SP, Mathur GP. Effects of Chemoprophylaxis on Minimal Pulmonary Tuberculosis Lesions of Doubtful Activity. Bull World Health Org. 1971;45(5):10.

33. Puffer RR, Stewart HC, Gass RS. Tuberculosis according to age, sex, family history, and contact. Am Rev Tuberc. 1945;51(4):295–311.

34. Sikand B, Narrain R, Mathur G. Incidence of TB as Judged by Re-surveys. Indian J Tuberc. 1959;6:73–83.

35. Styblo K, Dajkova D, Kubik A, Langerova M, Radkovsky J. Epidemiological and Clinical Study of Tuberculosis in the District of Kolin, Czechoslovakia. Bull World Health Org. 1967;37:56.

36. Tuberculosis Society of Scotland. A controlled trial of chemotherapy in pulmonary tuberculosis of doubtful activity. Tubercle. 1958 Jun;39(3):129–37.

37. Scottish Thoracic Society. A controlled trial of chemotherapy in pulmonary tuberculosis of doubtful activity: Five year follow-up. Tubercle. 1963 Mar;44(1):39–46.

38. Long ER, Hopkins FD. History of Diagnostic Standards and Classification of Tuberculosis of the National Tuberculosis Association. Am Rev Tuberc. 1952 Apr;65(4):494–504.

39. Moher D, Liberati A, Tetzlaff J, Altman D, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Med. 6(7).

Appendix B Supplementary Materials for Chapter 4 "The natural history of TB disease-a synthesis of data to quantify progression and regression across the spectrum"

B1 Disease model structure

Our model structure focuses on the spectrum of disease, with infection and progression from infection already assumed to have happened. We split the progression into three states, as shown in figure B1; minimal, subclinical, and clinical disease.

Minimal disease is the earliest stage of disease from infection, being non-infectious, but with pathological changes to the lung visible on, originally, chest x-ray but also other forms of chest imaging such as computed tomography (CT). Along with being the first stage of disease after infection, minimal is the final stage before recovery, with regression back to minimal possible within the spectrum framework, and then natural recovery from disease possible.

In the forward progression, the stage after minimal is subclinical. This is an infectious disease state, but without sufficient symptoms to present for screening. In other words, this is an infectious but effectively asymptomatic state. Within the spectrum, there is progression to clinical disease (i.e. development of symptoms) and regression to minimal disease (i.e. becoming non-infectious) from subclinical disease.

Clinical disease, symptomatic and infectious, is the final disease state. This state can only be reached by passing through minimal and subclinical first. As with the other two states, there are two transition possibilities out of clinical disease, regression to subclinical disease (i.e. resolution of symptoms but remaining infectious), and death from TB.

In the model structure, there is possibility to both progress and regress, but in visualising the model, such as figure 1, arrows that point right indicate the disease progressing to a more severe state, and arrows that point left indicate the disease regressing to a less severe state.

Disease State	Minimal	Subclinical	Clinical
Clinical description	Radiological changes attributable to TB but sputum negative (non-infectious)	Sputum positive TB disease that does not pass a symptom screen (infectious)	Symptomatic and sputum positive TB diseaese (infectious)
Systematic review tests	Chest x-ray or fluoroscopy positive Smear or culture negative	Smear or culture positive Symptom negative	Smear or culture positive Symptom positive
Systematic review notation	cxr.pos micro.neg symp.pos/neg/unk	cxr.pos micro.pos symp.neg	cxr.pos micro.pos symp.pos

Figure B1: TB natural history model. States and transitions in red are considered in the model fitting, with dashed parameters holding a fixed value and solid transitions being estimated from the data. Each state is defined by a clinical description and the tests that would have been used in clinical settings at the time our data was recorded with the notation used for data collection in table S4 described in the final row. The black dashed lines out from minimal and clinical are recovery and TB mortality respectively.

B1.1 Alternative structures

We opted to parameterise a three state, linear model structure. There are data to support this choice, and as a sense of progression of disease, it matches with previous conceptualisations.[1–5] However, the data to support this structure could also inform other, more complex, model structures. For example, minimal disease could be split into two or more different states based on symptoms or perceived activity of lesions seen on radiography. However more disease states and transitions dilute the available data which, despite the systematic review that underpins it, is already limited. It is always possible to make a model structure more complex but here we looked for the most useful balance between simplicity, representing the data available, and providing answers to questions being asked. We feel the three state, linear model structure used here provides this balance.

B2 Open TB vs Clinical TB

In a systematic review, Tiemersma et al describe how study results were interpreted as having people with either smear-negative or smear-positive disease:[6]

> in studies where patients were described as having ‘‘open’’ tuberculosis or ‘‘bacillary tuberculosis’’ before 1930 (when culture became available) we assumed that these patients were smear-positive.

This definition is maintained in the work of Ragonnet et al.[7]

As we were not attempting to stratify bacteriological status any further by smear status, we instead interpreted bacillary tuberculosis as simply bacteriologically positive. The definition of ''open'' tuberculosis was less clear cut, with different studies choosing different explanations.

Szucs, in 1926, defines open and closed tuberculosis based solely on the symptom status of the individual:[8]

> Among others, we classify tuberculous patients as open and closed cases, based on the presence or absence of expectoration.

However, they point out that absence of symptoms does not indicate a negative bacteriological status, in essence diagnosing their patients with subclinical disease using more modern terminology.

> We succeeded in demonstrating the presence of tubercle bacilli in the spray of two of our recently admitted patients in the absence of any expectoration.

Another study in 1947 provides a different definition for open and closed. Tattersall compares his work with that of Lindhardt in Denmark, stating that:[9,10]

> These results accord closely with Lindhardt's finding in Denmark during the same period, which enhances the value of comparison of the present series of sputum +ve cases with the results of the Danish survivals of open cases.

This suggests that rather than open being equivalent to clinical and closed being equivalent to subclinical, open is actually all bacteriologically positive disease (i.e. subclinical and clinical).

Finally, Blahd, in 1946 states that Illinois state defines that all cases that were bacteriologically positive must be defined as open TB, using the same definition as Tattersall.[11] However he also considers the possibility of infection from closed TB.

> All cases in which a positive sputum has been shown must by Illinois law be considered open for a period of at least three months and thereafter until three successive specimens of sputum, collected at intervals of one week, contain no tubercle bacilli… Although a so-called "closed" case is not as grave a source of infection as a frankly open one, it is more insidious.

Our definition of clinical aligns closely with the National Tuberculosis Association's definition of "active" TB, as described in their diagnostic standards from 1940.[12] By this point there was less focus on open and closed definitions and more consideration of symptoms and bacteriology together:

> Symptoms unchanged, worse or less severe, but not completely abated. Lesions not completely healed or progressive according X-ray examination. Sputum almost always contains tubercle bacilli.

Overall, the definitions of "open" tuberculosis and the final definition of "active" tuberculosis do not differ significantly, and again are not dissimilar to our definition of clinical disease. Therefore, where Ragonnet et al have provided a TB mortality rate for smear-positive TB, we are instead using this number as the TB mortality rate from clinical TB disease.[7]

B3 Symptomatic Minimal

Minimal disease was classified as when an individual had radiological changes attributable to TB but negative bacteriology, regardless of symptoms. Although progressing from a potentially symptomatic bacteriologically negative state to an asymptomatic bacteriologically positive seems unlikely, a number of sources suggested that there was no need to consider an alternative progression for symptomatic minimal.

Firstly, there is insufficient data to show an obvious split in progression between symptomatic and asymptomatic bacteriologically negative individuals (see figure B2), the diagnostic standards from the 1940s placed significantly less weight on symptoms if they were not accompanied by a positive sputum, and the prevalence survey in Cambodia in 2002 found that symptoms in culture negative individuals were not associated with future bacteriological positivity.[12,13] We also know that TB symptoms are highly non-specific, and so there is no guarantee that symptoms occurring whilst an individual has minimal disease are actually caused by TB and not by something else.

Therefore, we have considered all bacteriologically negative individuals to be minimal, regardless of symptom status.

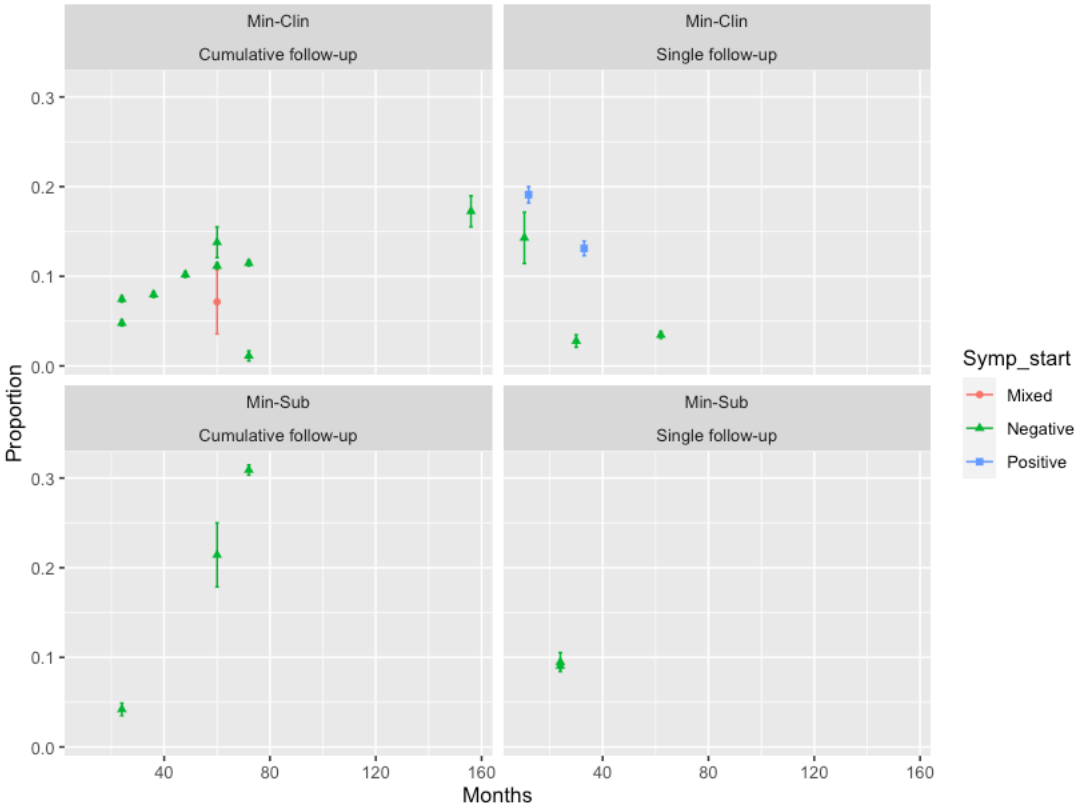


Figure B2: The different progression rates for minimal disease, based on symptoms, transition, and record type

B4 Data inclusion and exclusion

B4.1 Data types

As explained in the main text, there were two study types included. In table B1 we report the data types for each line of data, in column "Follow-up method". There were 38 data points reported as cumulative follow-up, and 16 reported as single follow-up.

B4.2 Timings

For single follow-up data, if an average follow-up time was given, that is the time used. Otherwise, if a minimum and maximum follow-up time was given, the times have been summed and halved to give the follow-up time used. For cumulative follow-up data, the maximum follow-up time given was used. The times in table B1, column "Months of follow-up" reflect these choices.

B4.3 Included data

All data that was included in the final data fit is in table B1.

These cohorts were followed for intervals between 1923 and 2004, with studies conducted in North America (6), Europe (7), Asia (7), and one each from South America and Africa. In total there were 5 data points from minimal to subclinical, 14 data points from minimal to clinical, 15 data points from clinical to minimal, 18 data points from minimal to infectious, and 2 data points from infectious to minimal.

Table B1: A table on all the studies included, the model transitions they parameterise, and, where applicable, notes on why a certain decision has been taken. The number of repeats column reports the number of data points that a study provides within a single transition. This is then used to calculate the effective number of people transitioned and the effective cohort size (divides the true numbers which are then rounded to the nearest integer) to appropriately weight each data point so that each study is considered as one.

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Downes[14]	North America	1935	cxr pos micro pos sympt pos	cxr pos micro neg sympt neg	27	342	12.0	cumulative	Clin-Min	5	5	68	
Downes[14]	North America	1935	cxr pos micro pos	cxr pos micro neg	104	342	24.0	cumulative	Clin-Min	5	21	68	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			sympt pos	sympt neg									
Downes[14]	North America	1935	cxr pos micro pos sympt pos	cxr pos micro neg sympt neg	140	342	36.0	cumulative	Clin-Min	5	28	68	
Downes[14]	North America	1935	cxr pos micro pos sympt pos	cxr pos micro neg sympt neg	158	342	48.0	cumulative	Clin-Min	5	32	68	
Downes[14]	North America	1935	cxr pos micro pos sympt pos	cxr pos micro neg sympt neg	171	342	60.0	cumulative	Clin-Min	5	34	68	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Beeuwkes[15]	North America	1938	cxr pos micro neg sympt neg	cxr unk micro pos sympt pos	16	122	33.0	single	Min-Clin	1	16	122	merged 2 groups exhibiting same start and end
Beeuwkes[15]	North America	1938	cxr pos micro pos sympt pos	cxr unk micro neg sympt unk	10	28	33.0	single	Clin-Min	1	10	28	
Puffer[16]	North America	1943	cxr pos micro neg sympt neg	cxr pos micro pos sympt pos	19	528	62.0	single	Min-Clin	1	19	528	merged 2 groups exhibiting same start and end
Puffer[16]	North America	1943	cxr pos	cxr pos	92	384	62.0	single	Clin-Min	1	92	384	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			micro pos sympt pos	micro neg sympt neg									
Puelma[17]	South America	1945	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	18	67	24.0	single	Min-Inf	1	18	67	
Bobrowitz[18,19]	North America	1945	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	26	191	60.0	single	Min-Inf	1	26	191	
Bosworth[20,21]	North America	1947	cxr pos micro mix	cxr pos micro neg	45	134	24.0	cumulative	Clin-Min	6	8	22	majority tested were micro pos and all

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			sympt unk	sympt unk									reported as active baseed on NTA definitions
Bosworth[20,21]	North America	1947	cxr pos micro mix sympt unk	cxr pos micro neg sympt unk	71	134	36.0	cumulative	Clin-Min	6	12	22	majority tested were micro pos and all reported as active baseed on NTA definitions
Bosworth[20,21]	North America	1947	cxr pos micro mix sympt unk	cxr pos micro neg sympt unk	80	134	48.0	cumulative	Clin-Min	6	13	22	majority tested were micro pos and all reported as active baseed on

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
													NTA definitions
Bosworth[20,21]	North America	1947	cxr pos micro mix sympt unk	cxr pos micro neg sympt unk	83	134	60.0	cumulative	Clin-Min	6	14	22	majority tested were micro pos and all reported as active baseed on NTA definitions
Bosworth[20,21]	North America	1947	cxr pos micro mix sympt unk	cxr pos micro neg sympt unk	86	134	72.0	cumulative	Clin-Min	6	14	22	majority tested were micro pos and all reported as active baseed on NTA definitions

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Bosworth[20,21]	North America	1947	cxr pos micro mix sympt unk	cxr pos micro neg sympt unk	87	134	84.0	cumulative	Clin-Min	6	15	22	majority tested were micro pos and all reported as active baseed on NTA definitions
Bosworth[20,21]	North America	1947	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	15	314	24.0	cumulative	Min-Clin	5	3	63	majority tested were micro pos and all reported as active baseed on NTA definitions
Bosworth[20,21]	North America	1947	cxr pos	cxr pos	25	314	36.0	cumulative	Min-Clin	5	5	63	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			micro neg sympt unk	micro pos sympt unk									
Bosworth[20,21]	North America	1947	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	32	314	48.0	cumulative	Min-Clin	5	6	63	
Bosworth[20,21]	North America	1947	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	35	314	60.0	cumulative	Min-Clin	5	7	63	
Bosworth[20,21]	North America	1947	cxr pos micro neg	cxr pos micro pos	36	314	72.0	cumulative	Min-Clin	5	7	63	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			sympt unk	sympt unk									
Bosworth[20,22]	North America	1948	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	8	58	60.0	cumulative	Min-Clin	2	4	29	
Bosworth[20,22]	North America	1948	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	10	58	156.0	cumulative	Min-Clin	2	5	29	
Marshall[23]	Europe	1948	cxr pos micro pos sympt pos	cxr pos micro neg sympt unk	2	52	6.0	cumulative	Clin-Min	1	2	52	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Borgen[24,25]	Europe	1949	cxr pos micro neg sympt pos	cxr pos micro pos sympt pos	4	144	30.0	single	Min-Clin	1	4	144	merged 2 groups exhibiting same start and end
Manser[26]	Europe	1951	cxr pos micro pos sympt unk	cxr pos micro neg sympt unk	15	40	6.0	single	Clin-Min	1	15	40	
Breu[27]	Europe	1952	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	48	904	25.5	single	Min-Inf	1	48	904	
Sikand[28]	Asia	1958	cxr pos	cxr pos	38	319	12.0	cumulative	Min-Inf	1	38	319	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			micro neg sympt unk	micro pos sympt unk									
Tuberculosis Society of Scotland[29,30]	Europe	1959	cxr pos micro neg sympt neg	cxr pos micro pos sympt unk	9	95	24.0	single	Min-Sub	1	9	95	assume lack of sysmptom persists
Frimodt-Moller[31]	Asia	1961	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	11	86	12.0	cumulative	Min-Inf	3	4	29	
Frimodt-Moller[31]	Asia	1961	cxr pos micro neg	cxr pos micro pos	18	86	24.0	cumulative	Min-Inf	3	6	29	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			sympt unk	sympt unk									
Frimodt-Moller[31]	Asia	1961	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	25	86	36.0	cumulative	Min-Inf	3	8	29	
Pamra[32]	Asia	1968	cxr pos micro neg sympt neg	cxr pos micro pos sympt pos	2	178	72.0	cumulative	Min-Clin	1	2	178	
Pamra[32]	Asia	1968	cxr pos micro neg sympt neg	cxr pos micro pos sympt neg	55	178	72.0	cumulative	Min-Sub	1	55	178	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
National Tuberculosis Institute[33–41]	Asia	1968	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	23	329	18.0	single	Min-Inf	2	12	165	
National Tuberculosis Institute[33–41]	Asia	1968	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	36	271	60.0	single	Min-Inf	2	18	136	
National Tuberculosis Institute[33–41]	Asia	1968	cxr pos micro pos sympt unk	cxr pos micro neg sympt unk	86	269	18.0	single	Inf-Min	2	43	135	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
National Tuberculosis Institute[33–41]	Asia	1968	cxr pos micro pos sympt unk	cxr pos micro neg sympt unk	70	178	36.0	single	Inf-Min	2	35	89	
Aneja[42]	Asia	1977	cxr pos micro neg sympt pos	cxr pos micro pos sympt unk	21	110	12.0	single	Min-Clin	1	21	110	assume symptoms persist
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	40	176	3.0	cumulative	Min-Inf	8	5	22	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	49	176	6.0	cumulative	Min-Inf	8	6	22	
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	61	176	12.0	cumulative	Min-Inf	8	8	22	
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	67	176	18.0	cumulative	Min-Inf	8	8	22	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	69	176	24.0	cumulative	Min-Inf	8	9	22	
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	70	176	30.0	cumulative	Min-Inf	8	9	22	
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	71	176	36.0	cumulative	Min-Inf	8	9	22	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Hong Kong Chest Service[43–46]	Asia	1981	cxr pos micro neg sympt mixed	cxr pos micro pos sympt unk	71	176	60.0	cumulative	Min-Inf	8	9	22	
Cowie[47]	Africa	1984	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	88	152	58.0	cumulative	Min-Inf	1	88	152	
Norregaard[48]	Europe	1985	cxr pos micro neg sympt mixed	cxr pos micro pos sympt neg	6	28	60.0	cumulative	Min-Sub	1	6	28	
Norregaard[48]	Europe	1985	cxr pos	cxr pos	2	28	60.0	cumulative	Min-Clin	1	2	28	

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
			micro neg sympt mixed	micro pos sympt pos									
Anastasatu[49]	Europe	1985	cxr pos micro neg sympt neg	cxr pos micro pos sympt unk	6	143	24.0	cumulative	Min-Sub	1	6	143	assume lack of sysmptom persists
Okada[13]	Asia	2004	cxr pos micro neg sympt neg	cxr pos micro pos sympt unk	28	309	24.0	single	Min-Sub	1	28	309	split group by symptoms dependent on proportion found in prevalence survey

First Author	Study Continent	Year	Start states	End states	Number transitioned	Cohort size	Months of follow-up	Follow-up method	Model transition	Number of repeats	Effective number transitioned	Effective cohort size	Notes
Okada[13]	Asia	2004	cxr pos micro neg sympt neg	cxr pos micro pos sympt pos	23	309	24.0	cumulative	Min-Clin	1	23	309	added in 39% (18) of those picked up at 2 year follow-up using known subclinical proportion at prevalence surveys

B4.4 Excluded data

As can be seen in table B2, some of the data that was originally extracted for the wider review was not eligible for this work. In total 6 cohorts were excluded, from from 10 different studies. Most (5) rows of data that were excluded, had an initial state with x-ray negative, and one study was observing a cohort who, although already x-ray positive, were not expected to progress to infectious TB disease. Others were excluded for too much uncertainty within the start and end states, or changes only within states, such as change in x-ray severity, but no change in bacteriological or symptom status. These reasonings are laid out in table B2

Table B2: A table on all the data excluded,and reasons why they have not been included

First Author	Year	Continent	Start state	End state	Number transitioned	Cohort size	Months of follow-up	Notes
Beeuwkes[15]	1938	North America	cxr neg micro neg sympt neg	cxr unk micro pos sympt pos	1	784	33	initial state x-ray negative
Beeuwkes[15]	1938	North America	cxr neg micro neg sympt neg	cxr unk micro neg sympt pos	3	784	33	initial state x-ray negative
National Tuberculosis Institute[33–41]	1968	Asia	cxr neg micro neg sympt unk	cxr pos micro pos sympt unk	44	31490	18	initial state x-ray negative
Okada[13]	2004	Asia	cxr neg micro neg sympt neg	cxr pos micro pos sympt unk	32	21580	24	initial state x-ray negative
Okada[13]	2004	Asia	cxr pos	cxr neg	26	309	24	ends outside disease

First Author	Year	Continent	Start state	End state	Number transitioned	Cohort size	Months of follow-up	Notes
			micro neg sympt neg	micro neg sympt unk				
International Union Against Tuberculosis[50]	1982	Europe	cxr pos micro neg sympt unk	cxr pos micro pos sympt unk	97	6990	60	cohort not expected to progress - effectively initial state x-ray negative
Styblo[51]	1965	Europe	cxr neg micro neg sympt neg	cxr unk micro pos sympt unk	241	73000	30	initial state x-ray negative

B5 Fitting process

The equations to define the model system are:

$$\frac{dM}{dt} = -m\_s * M + s\_m * S - r * M$$
$$\frac{dS}{dt} = m\_s * M - s\_m * S - s\_c * S + c\_s * C$$
$$\frac{dC}{dt} = s\_c * S - c\_s * C - d * C$$

where:

- M, S, and C are the states for minimal, subclinical, and clinical respectively
- m_s, s_m, s_c, and c_s are transitions between the states, where the first letter is the start state and the second letter is the end state
- r is recovery from minimal disease
- d is death from clinical disease (there is no other death included in the model)

As the data described how a cohort changed over time, and described only one outcome, the fitting process used a model system for each of the transitions and and data types, totalling 16 different versions of the model system. Full code is available on GitHub.

We used uniform priors for the four estimated parameters, all with a range of 0 to 12, where 12 would be equivalent to changing state once a month. During the fitting process, potential parameters are trialled within this range. figure B3 shows the different parameter values that were accepted over the 10,000 iterations of the model fit.

These accepted parameters in turn, inform figure B4, which shows the correlation between two parameters. It also shows the overall distribution of the accepted parameters. We see a strong positive correlation between the parameters that control transition between minimal and subclinical; as one transition increases, the also has to increase to prevent there being excess people in one state and too few in another. We also see this with the parameters that control transition between subclinical and clinical.

The rest of the pairings have slightly weaker, negative correlations. This is clearest with the subclinical to minimal and subclinical to clinical pairing, as if one increases, the other has to decrease to make sure that there are still sufficient individuals in subclinical to fit the data.

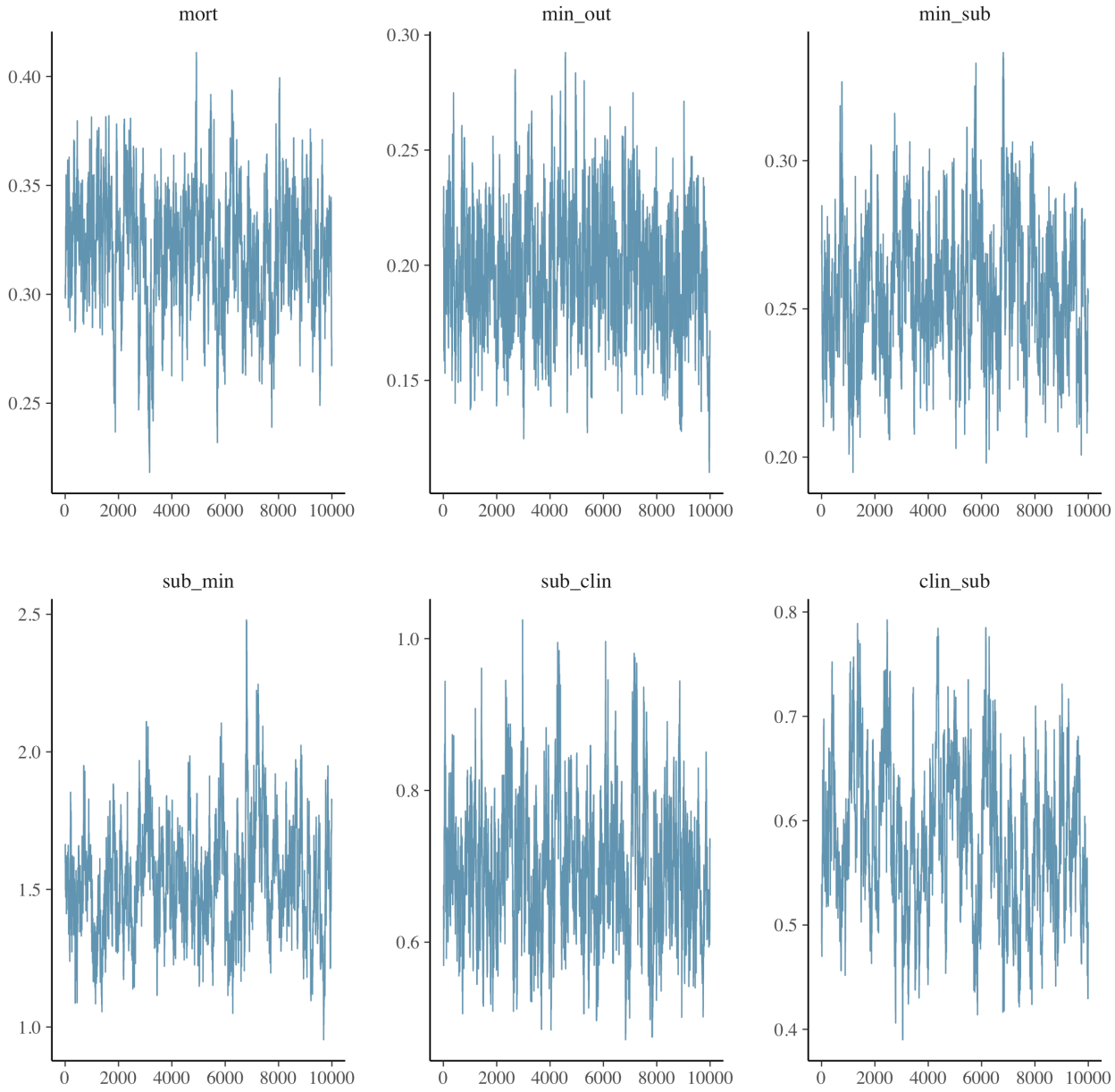


Figure B3: The traces of the accepted parameter range from the fitting process

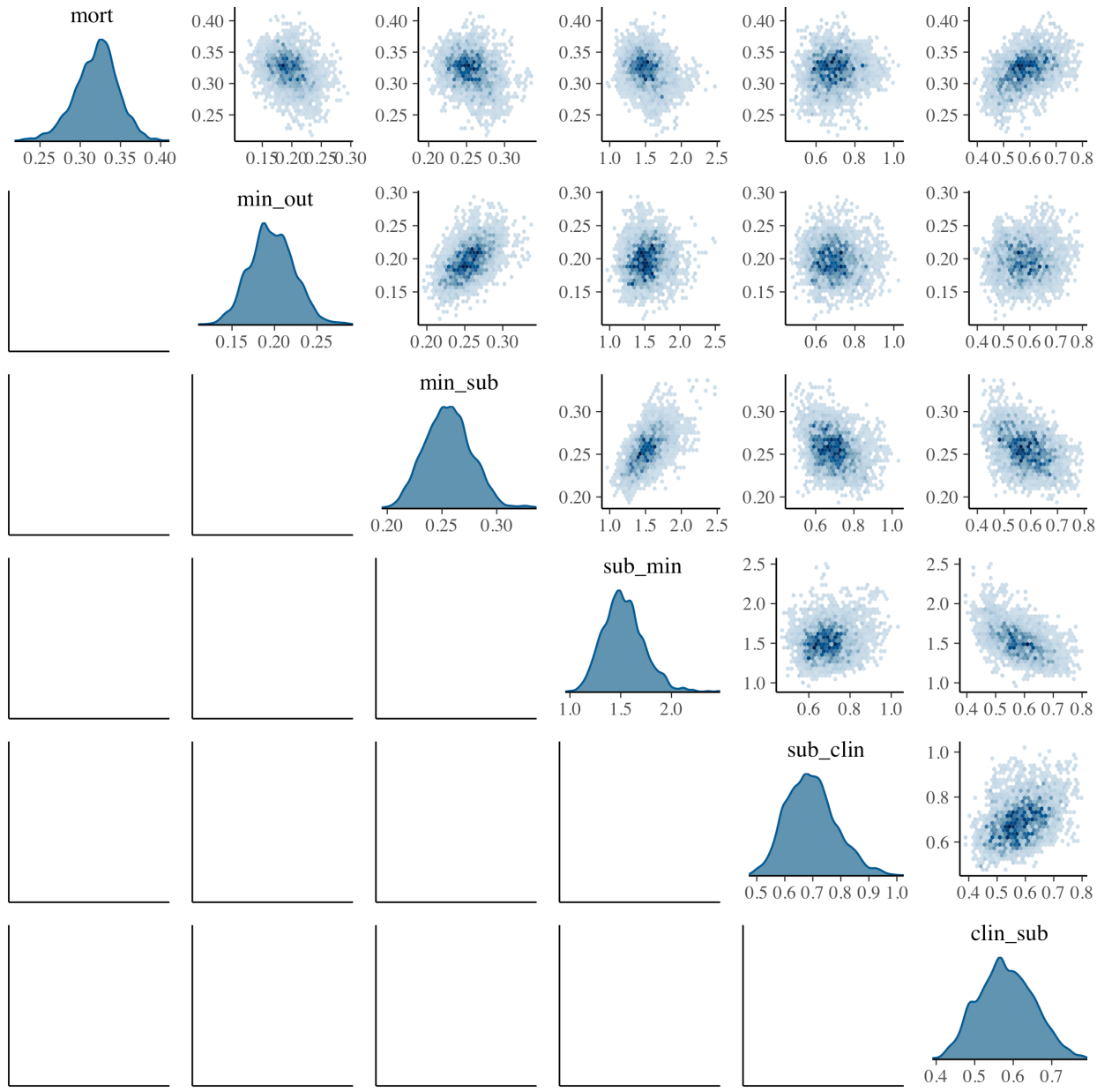


Figure B4: The distributions of and correlations between the accepted parameters from the fitting process

B5.1 Weighting

When calculating the likelihood, larger studies were weighted by the original cohort size to reflect the increased confidence that such studies provide. Thus larger cohorts have a heavier weighting and constrain the model more.

In order to prevent a single study with multiple observations being over-represented in the fit of a transition, we down-weighted both the sample size and the number of people transitioning by the number of repeated observations. This maintains the observed proportion to transition whilst reducing confidence and thus importance given to each individual data point within the

study. This is shown in table 1, with the number of repeats and the re-calculated effective cohort size and number transitioning. This ensures that the proportion remains the same, but less weight is given to each individual data point.

B5.2 Duration of disease

Tiemersma et al use an assumption of exponential duration of disease to quote an "average" duration of disease as three years and calculates this from the incident cases occuring between each survey.[6,33] They state that a δ of 0.3 fits the cumulative distribution for the number of observed cases and thus $\frac{1}{\delta} = 3.33$ years is the average duration given by the data, but that missed cases mean that is likely an over-estimate and so 3 years is the average duration of disease . What they are then quoting as average is the mean, so the median duration can be given by $\frac{ln(2)}{\delta}$. Taking $\delta = 0.33$ so that mean duration is 3, gives a median duration of 2.1 years.

The numbers quoted are incident cases between each survey, so we can use the duration of disease looking at a cohort that starts in subclinical disease. Therefore we need to find that at 2 years, 50% of the cohort that started in subclinical, is either still subclinical or is clinical.

To use this as a fitting point, we want to look at time 2 years, and see how close to 50% the number of people in subclinical + clinical from the subclinical cohort is.

An exponential function can be written as $p = a(b^t)$ where p is prevalence and t is time. We know that at $t = 0$, $p = 1$ so $a = 1$ and the equation simplifies to $p = b^t$.

We want to look at 2 years, find the prevalence, and then from that, calculate the time at which the prevalence would be 0.5. So to calculate b, we set $p = 0.5$ and $t = t_{med}$.

$$\begin{aligned} 0.5 &= b^{t_{med}} \\ b &= \left(\frac{1}{2}\right)^{\frac{1}{t_{med}}} \end{aligned}$$

So the full equation, at the fitting point of $t = 2$ and rearranging for t_{med} gives:

$$\begin{aligned} p &= 1\frac{2}{t_{med}} \\ ln(p) &= ln\left(\frac{1}{2}\right)\frac{2}{t_{med}} \\ t_{med} &= ln\left(\frac{1}{2}\right)\frac{2}{ln(p)} \\ t_{med} &= \big(ln(1) - ln(2)\big)\frac{2}{ln(p)} \\ t_{med} &= -2\frac{ln(2)}{ln(p)} \end{aligned}$$

This means, from fitting at a single time point, we can estimate the median duration of disease using the assumption of an exponential distribution of duration.

B5.3 Prevalence ratios

Prevalence surveys have found that approximately 50% of people with bacteriologically positive disease do not report experiencing symptoms,.[52] Whilst harder to ascertain, estimates of the proportion of people with bacteriologically negative disease range from two to three times the number of people with infectious disease. Both these have been included in the model fit as data points, calculated from the steady states of the system equations.

The subclinical to clinical ratio is calculated:

$$\begin{aligned} \frac{dC}{dt} &= s\_c * S - c\_s * C - d * C \\ 0 &= s\_c * S - c\_s * C - d * C \\ s\_c * S &= c\_s * C + d * C \\ \frac{S}{C} &= \frac{c\_s + d}{s\_c} \end{aligned}$$

For simplicity, the parameters representing transitions have been simplified to single letters:

- m_s -> e
- s_m -> f
- s_c -> g
- c_s -> h
- r -> j
- d -> k

and to create a non-zero steady state, an unknown α is the flow of new disease.

The system of equations then becomes:

$$\begin{aligned} \dot{M} &= \alpha - (e + j)M + fS \\ \dot{S} &= eM - (g + f)S + hC \\ \dot{C} &= gS - (h + k)C \end{aligned}$$

Assuming a steady state and using the equation for $\dot{S}$ we can get an equation for M in terms of S and C:

$$\begin{aligned} 0 &= eM - (g + f)S + hC \\ eM &= (g + f)S - hC \\ M &= \frac{(g + f)S - hC}{e} \end{aligned}$$

Substituting C in terms of S:

$$M = \frac{(g+f)S - h\frac{g}{h+k}S}{e}$$
$$M = \frac{(g+f)(h+k) - hg}{e(h+k)}S$$

Then to calculate $\frac{M}{S+C}$:

$$\frac{M}{S+C} = \frac{\frac{(g+f)(h+k) - hg}{e(h+k)}S}{S+C}$$
$$\frac{M}{S+C} = \frac{\frac{(g+f)(h+k) - hg}{e(h+k)}S}{S+\frac{g}{h+k}S}$$
$$\frac{M}{S+C} = \frac{\frac{(g+f)(h+k) - hg}{e(h+k)}}{1+\frac{g}{h+k}}$$
$$\frac{M}{S+C} = \frac{\frac{(g+f)(h+k) - hg}{e(h+k)}}{\frac{h+k+g}{h+k}}$$
$$\frac{M}{S+C} = \frac{(g+f)(h+k) - hg}{e(h+k+g)}$$

B5.4 True Minimals

In the systematic review preceding this work, x-ray positive, bacteriologically negative diseasee was analysed based on reporting of the presumed activity (whether the x-rays were classified as active or inactive). For modelling purposes, there was insufficient data to split groups starting in minimal disease beyond the symptoms at the end and the follow-up collection type, so the distinction between active and inactive x-rays has not been included. However, determining who truly has TB when the only test is an x-ray is difficult. To take this into account, we have used tuberculin skin test (TST) as a proxy for determining whether a positive x-ray is a result of TB infection progressing to disease, and so we can estimate the proportion of people classed as minimal that are actually minimal. These papers were not selected systematically but span a range of time and location. Table B3 shows each of the studies, the number of people who were found to be x-ray positive in the study, and then the number of those who were also TST negative.

Table B3: The different studies that contributed towards the decision to reduce the proportion of people with positive x-rays that were considered to be truly minimal

Author	xray_pos	tst_neg
Groth-Petersen, 1959	37494	6097
Roelsgaard, 1964	2857	772
Roelsgaard, 1961	559	169
Scheel, 1937	255	54
National Tuberculosis Institute, 1974	3761	1848

Applying a meta analysis to this, we find that the fixed effects result is 20% of x-ray positives are TST negative and so unlikely to be caused by TB, and the random effects suggests 28%, as can be seen in figure B5. Therefore throughout this paper we have assumed that 25% of all x-ray positives are non-TB, and thus reduced every cohort that starts in minimal accordingly.

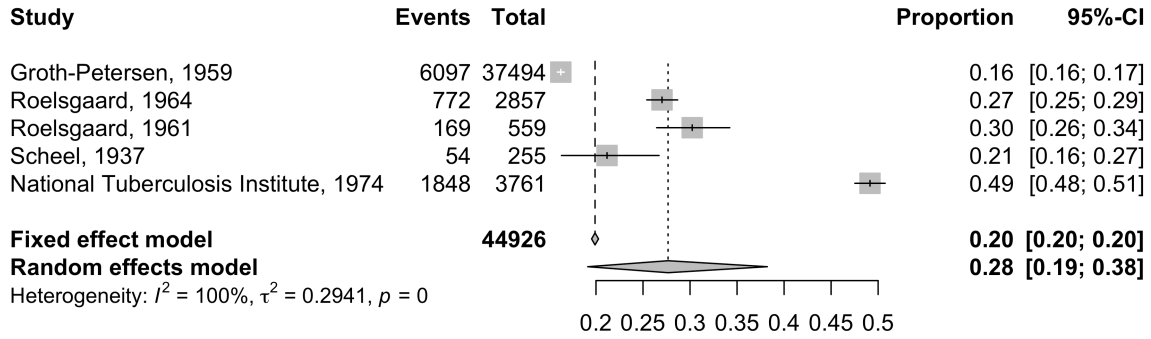


Figure B5: The results of a meta-analysis on the proportion of those with positive x-rays who also test tst negative, as a proxy for the proportion of positive x-rays that are not caused by TB

To check this assumption, we have tested this with 0% and 40% of x-ray positives being non-TB, as can be seen in table 6.

B6 Minimal disease

Figures B6 to B8 show the disease trajectories for simulated cohorts with minimal disease. Figure B6 shows a cohort with an approximate spread of all three disease states, based on the priors informing the model fit, with a ratio of clinical, subclinical, and minimal of 1:1:5. Over 5 years the trajectories are reported in the same way as **Error! Reference source not found.**. Figure B7 shows the the same trajectories, but for a cohort with just minimal disease. Figure B8 is a Sankey plot, like **Error! Reference source not found.**, but with the initial cohort split 1:1:5 between clinical, subclinical, and minimal, like Figure B6.

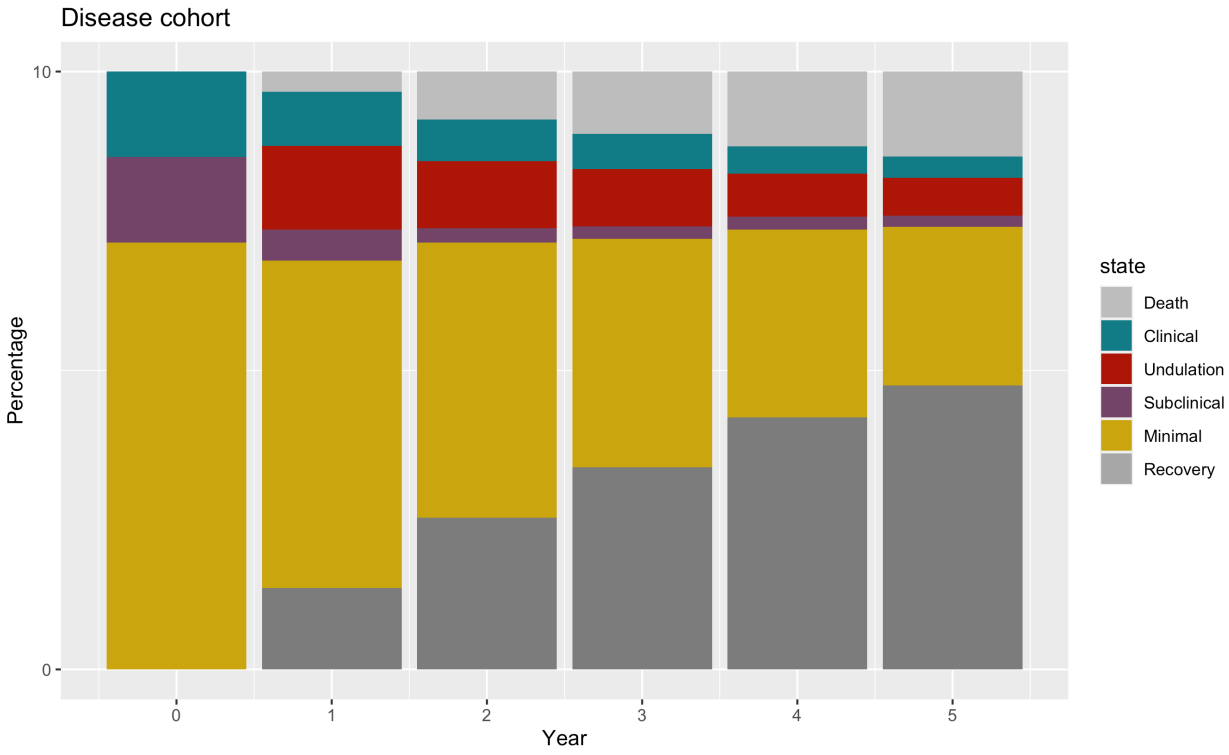


Figure B6: Trajectories of disease over time given a cohort starting with a mix of clinical, subclinical, and minimal disease in the ratio 1:1:5

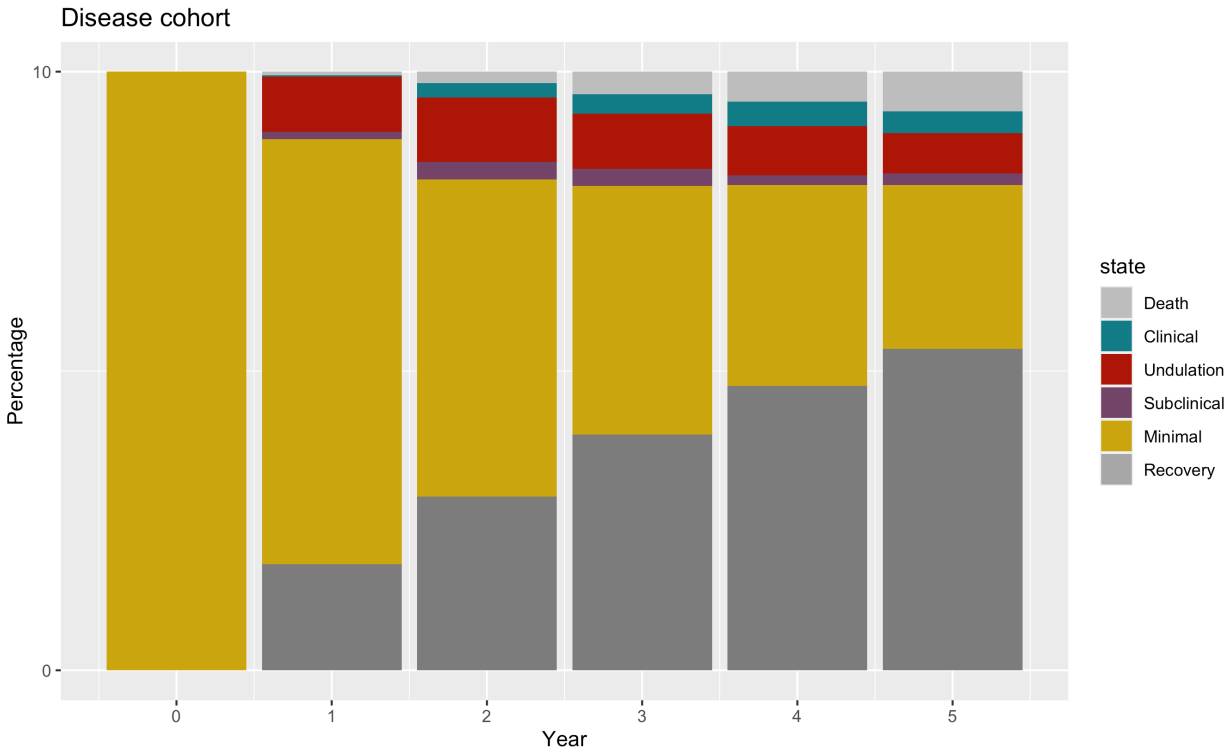


Figure B7: Trajectories of disease over time given a cohort starting with minimal disease

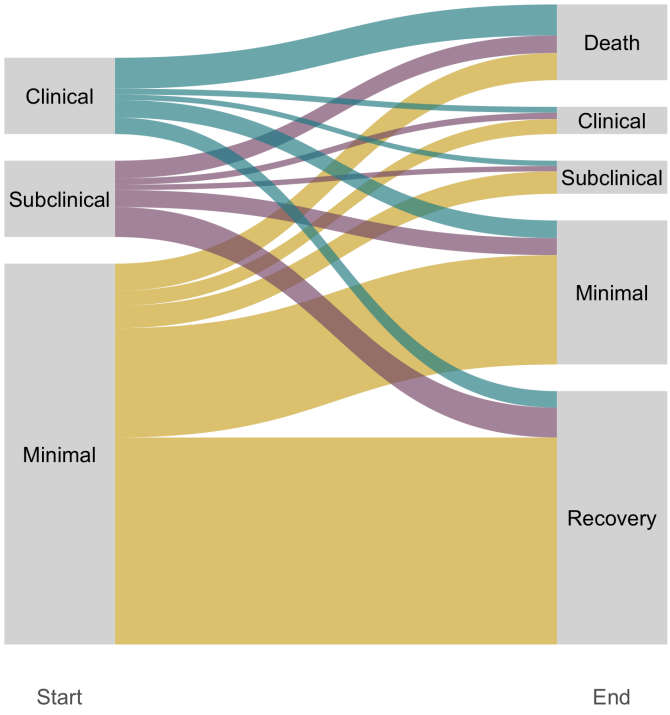


Figure B8: Final state after five years of people starting in minimal, subclinical and clinical disease

B7 Disease pathways

Figures B9 – B15 show potential individual trajectories of disease and how they would be classified within the trajectory plots of **Error! Reference source not found.** and Figures B6 and B7. Figure B9 shows a trajectory of subclinical disease in year 1 followed by death in year 2. Figure B10 shows a trajectory of subclinical disease for two years, followed by a year of subclinical disease. However, because recovery occurs before the end of the third year, this is classified as recovery in year three. Figure B11 highlights a year of subclinical disease and Figure B12 highlights a year of clinical disease both in the fourth year of the respective trajectories. Figure B13 shows a trajectory of minimal disease, highlighted for year three, but applicable to years 2 to 5. Figures B14 and B15 show undulating disease, B14 between minimal and subclinical in year five and B15 between subclinical and clinical in year two. These trajectories are for the purpose of description only, and do not represent the frequency of any equivalent trajectory occurring.

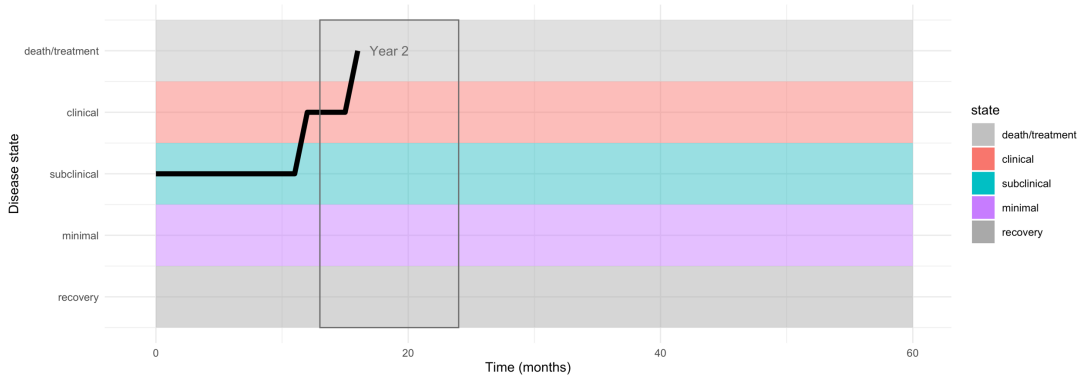


Figure B9: An example trajectory where the individual dies from TB. In this case, the death happens in year 2

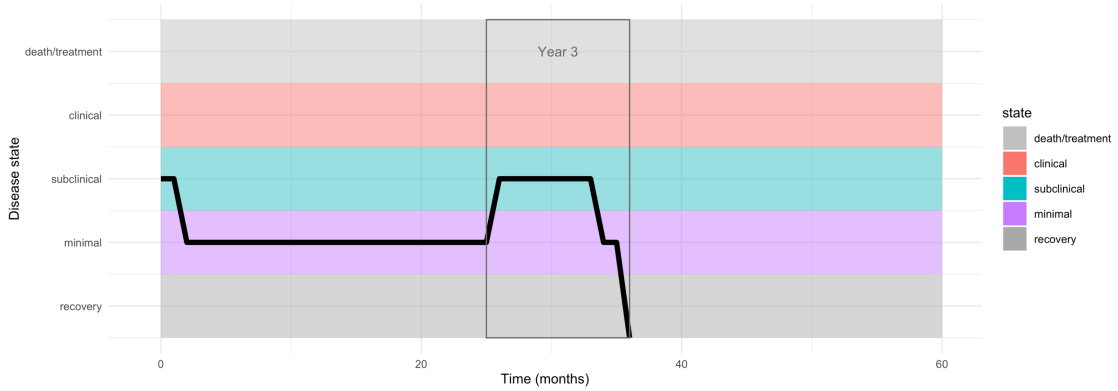


Figure B10: An example trajectory where the individual recovers from TB. In this case, the individual spent the first 2 years having regressed to minimal, then progresses to subclinical in the third year before regressing quickly to recovery

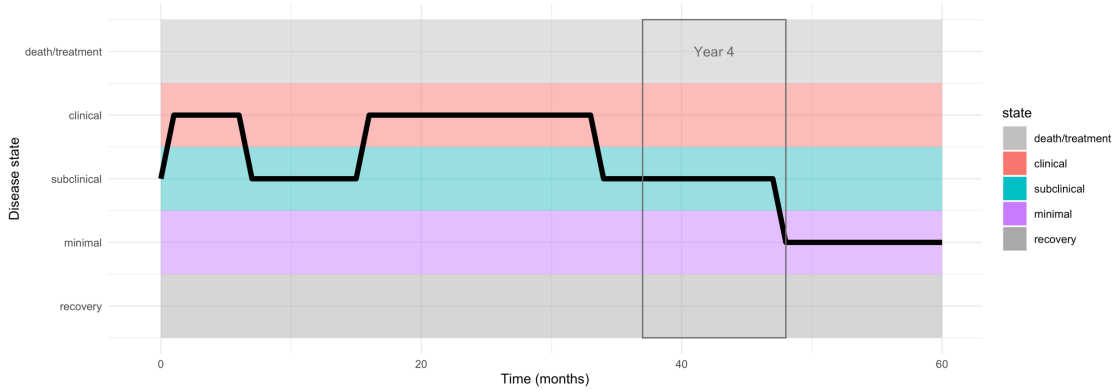


Figure B11: An example trajectory of subclinical in the 4th year. The previous years have time in both clinical and subclinical and year 5 is entirely in minimal, however, as the majority of time (>=9 months) in year 4 and there are fewer than three state changes, the 4th year is defined as subclinical

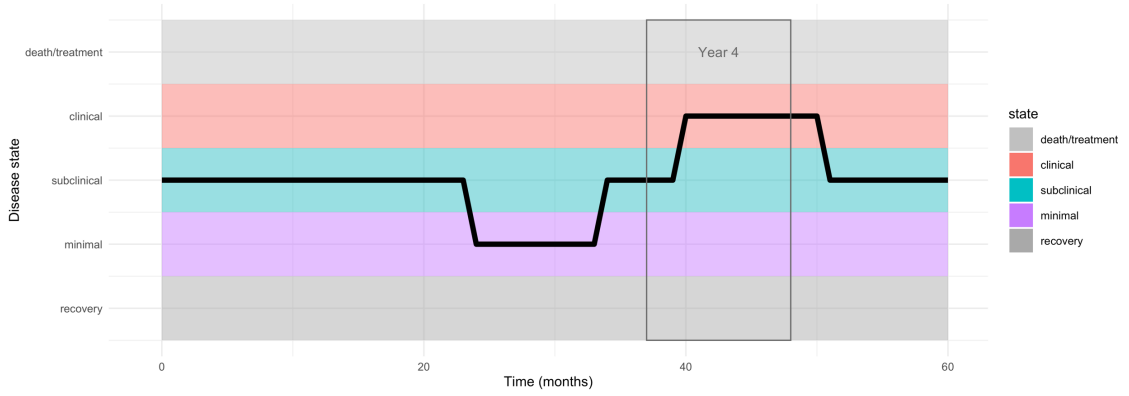


Figure 12: An example trajectory of clinical in the 4th year. The previous years have time in both minimal and subclinical and year 5 is mainly subclinical. As the majority of time (>=9 months) in year 4 is clinical and there are fewer than three state changes, the 4th year is defined as clinical

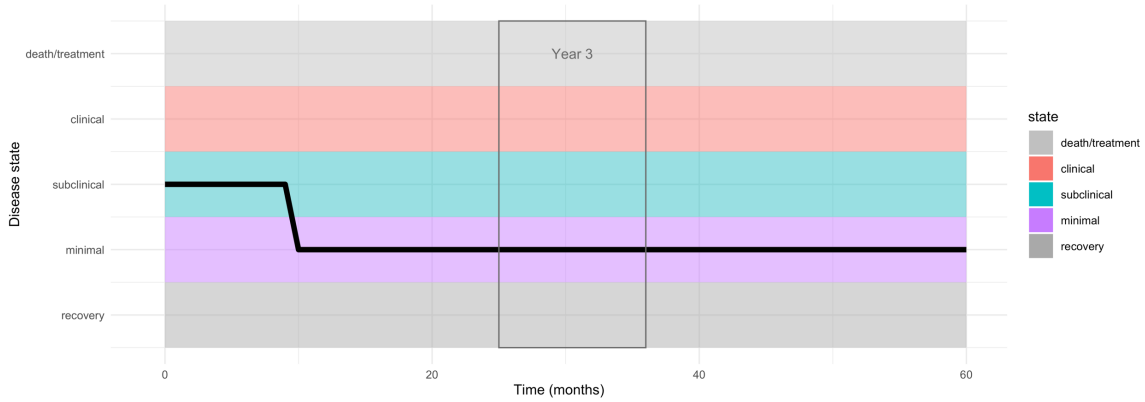


Figure 13: An example trajectory of minimal in the 3rd year. Other than the first year, with time in subclinical, the remaining years are also all minimal

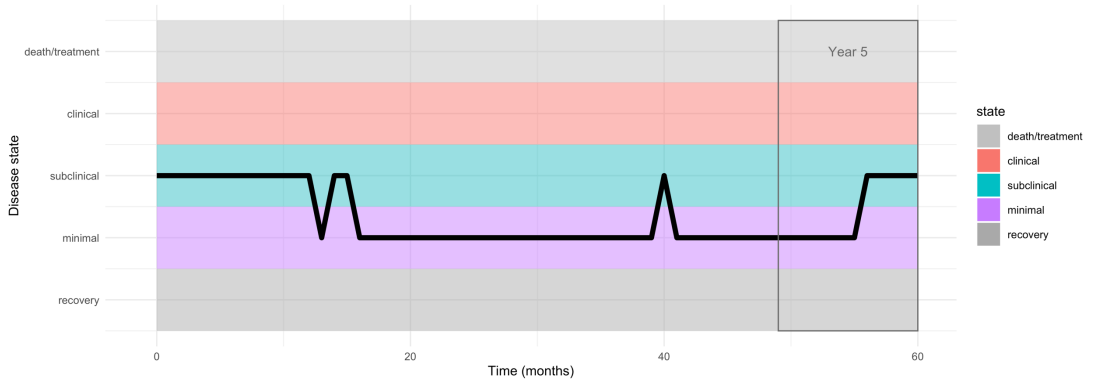


Figure 14: An example trajectory of undulating disease in the 5th year. There are 2 disease states in the 5th year, with neither lasting for 9 months. The majority of the remainder of this trajectory is in minimal

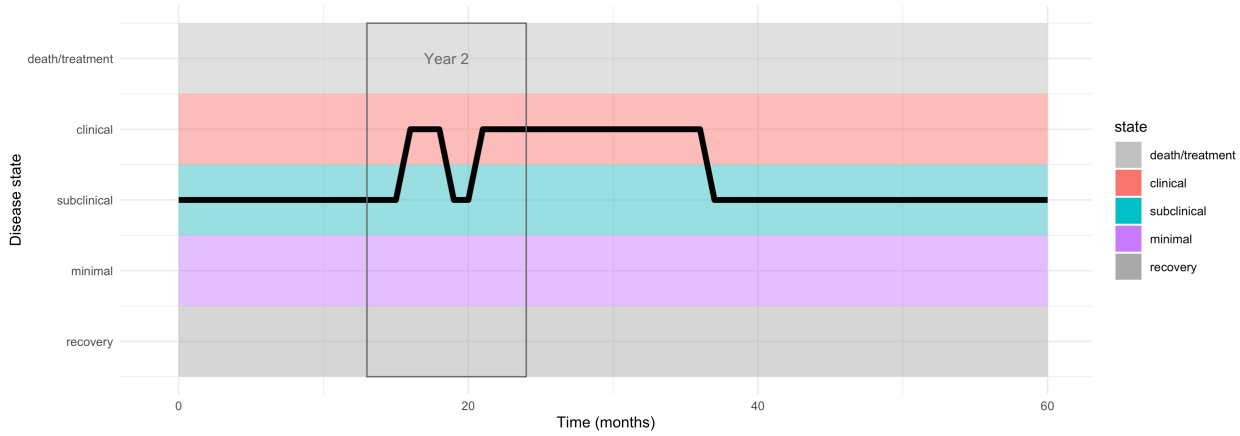
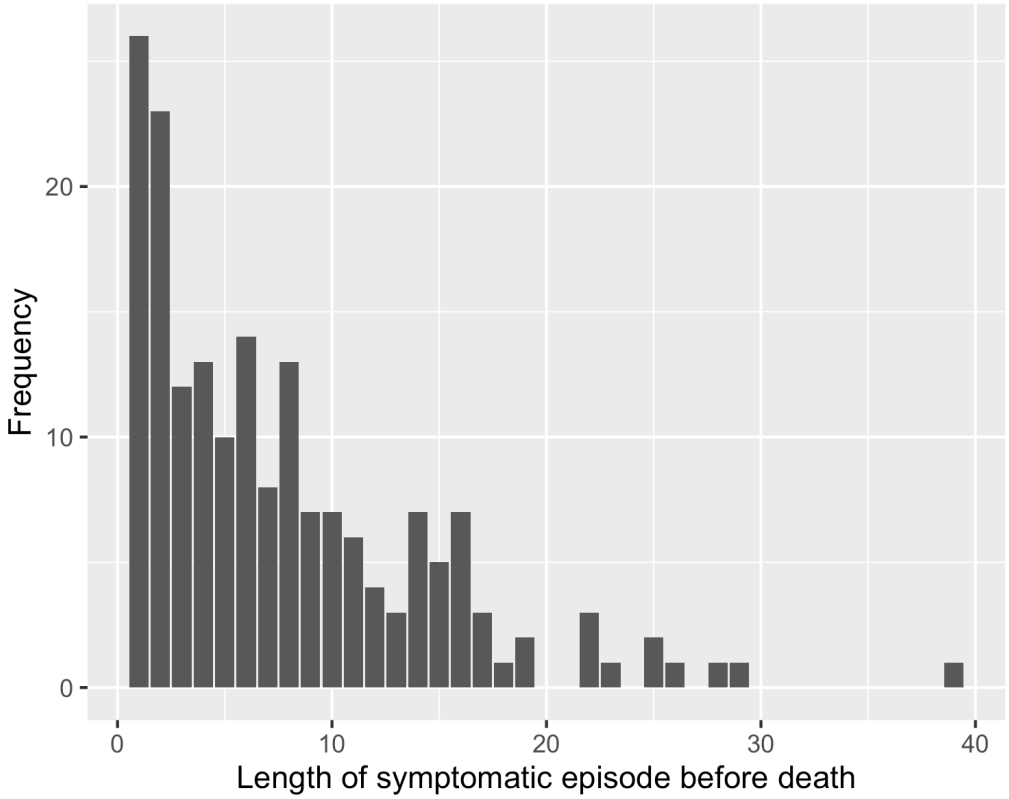
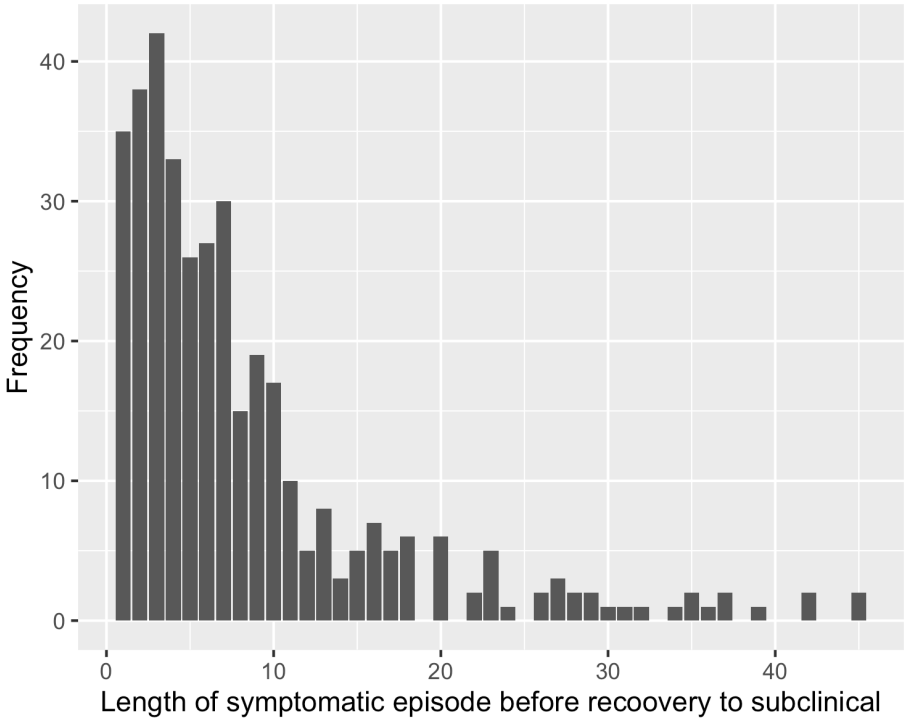


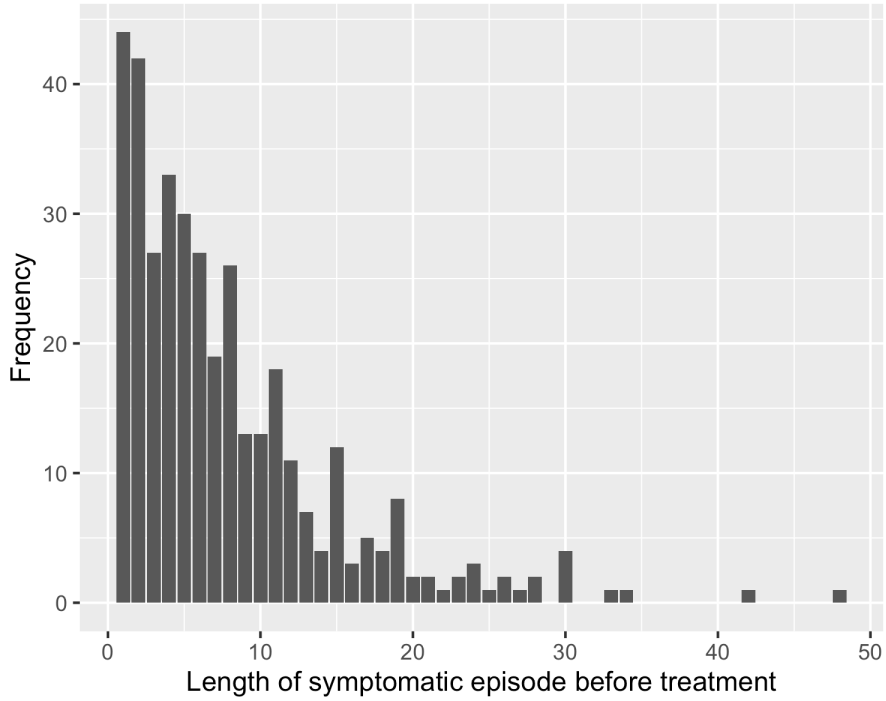
Figure 15: An example trajectory of undulating disease in the second year, with three state changes and neither state lasting for a total of 9 months.

B8 Duration of symptoms

Duration of symptoms can be split into three categories; duration before death, duration before regression to subclinical, and when applicable, duration before treatment. These three have not shown a significant difference in our analysis, but of note is the highly skewed distribution that we observe. Of those who become clinical, the minimum time spent clinical is one month (as that is the time step in the model), but a small proportion of individuals have persistant symptoms for a long time


Frequency
20
10
0
0
10
20
30
40
Length of symptomatic episode before death


Frequency
40
30
20
10
0
0
10
20
30
40
Length of symptomatic episode before recoovery to subclinical



B9 Additional results

Here we consider the median duration of disease and the proportion of people in each state at a given time. We can see that including treatment decreases the duration of disease and decreases the proportion clinical. When including minimal disease in the duration, we see that duration increases significantly, showing the importance of considering all those who are at risk of progressing to infectious disease. In the following figures, the top row is the number of people in all disease states (minimal, subclinical and clinical) over time, and the bottom row is the number of people in infectious disease states (subclinical and clinical).

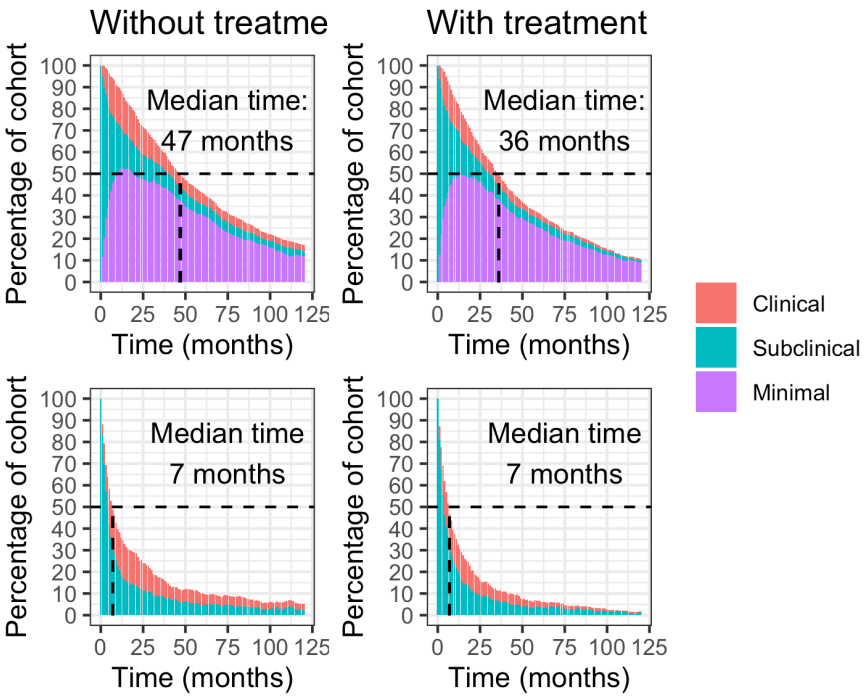


Figure 16: Median duration of infectious and all disease with and without treatment, starting with a cohort of subclinical individuals

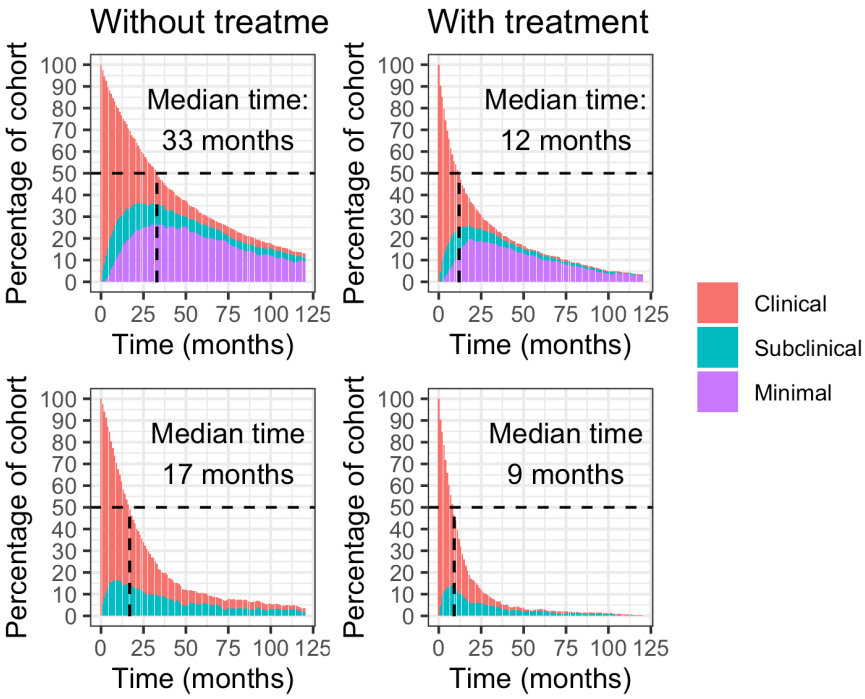


Figure 17: Median duration of infectious and all disease with and without treatment, starting with a cohort of clinical individuals

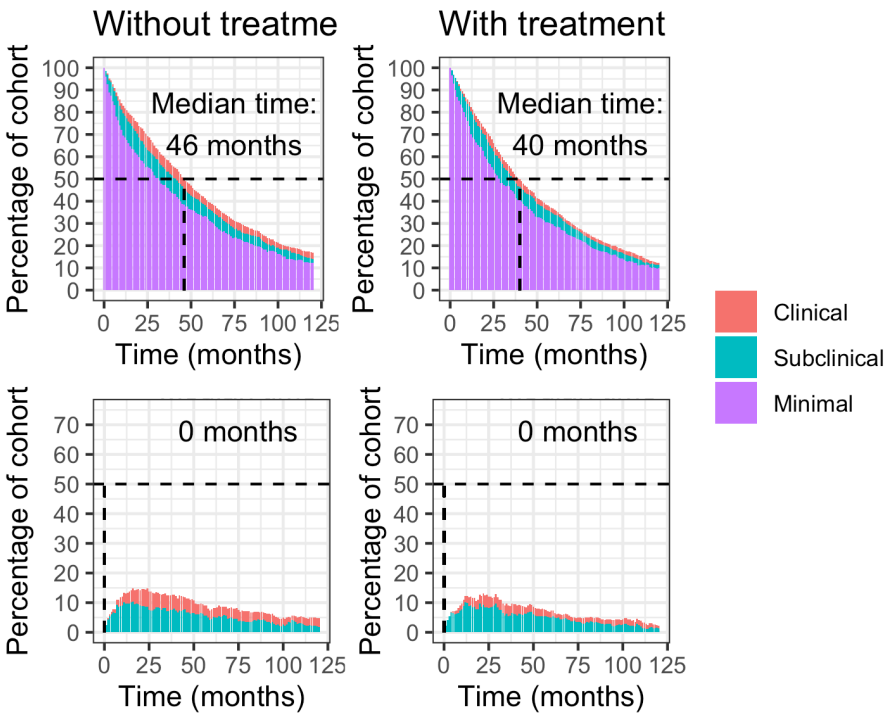


Figure 18: Median duration of infectious and all disease with and without treatment, starting with a cohort of minimal individuals

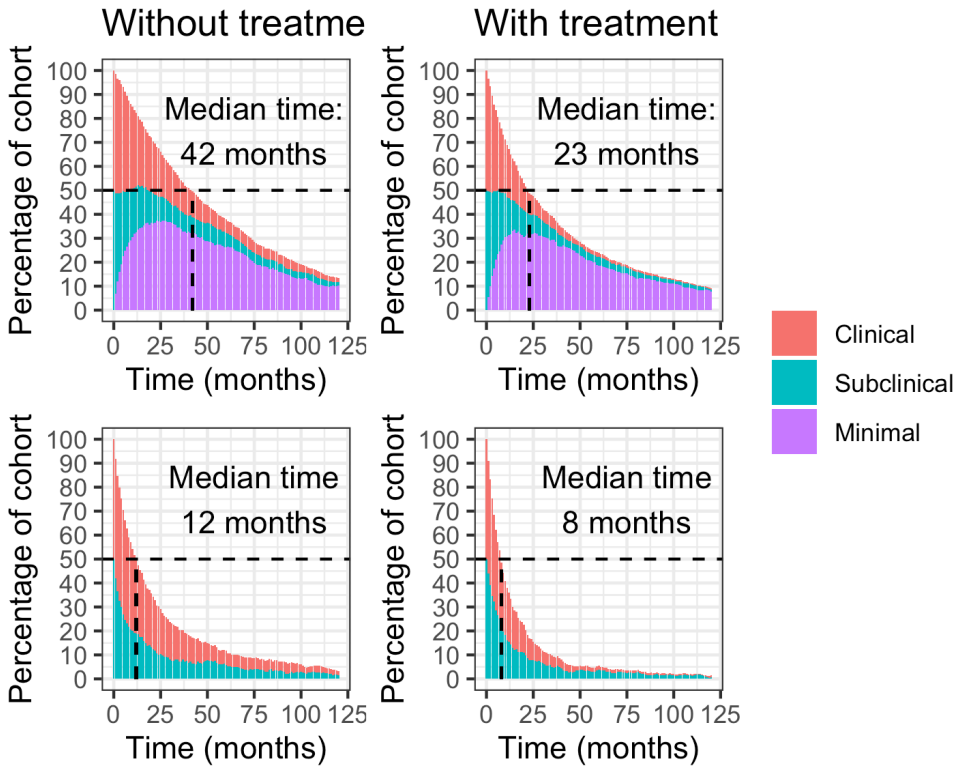


Figure 19: Median duration of infectious and all disease with and without treatment, starting with a cohort of half clinical and half subclinical individuals

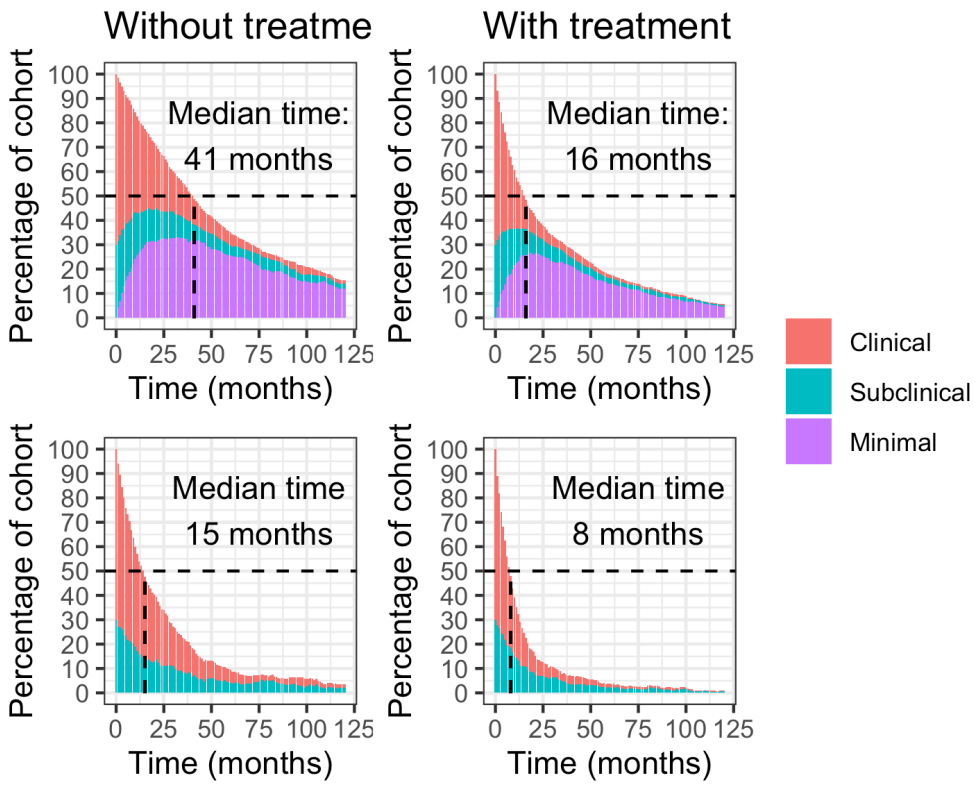


Figure 20: Median duration of infectious and all disease with and without treatment, starting with a cohort of 70% clinical and 30% subclinical individuals

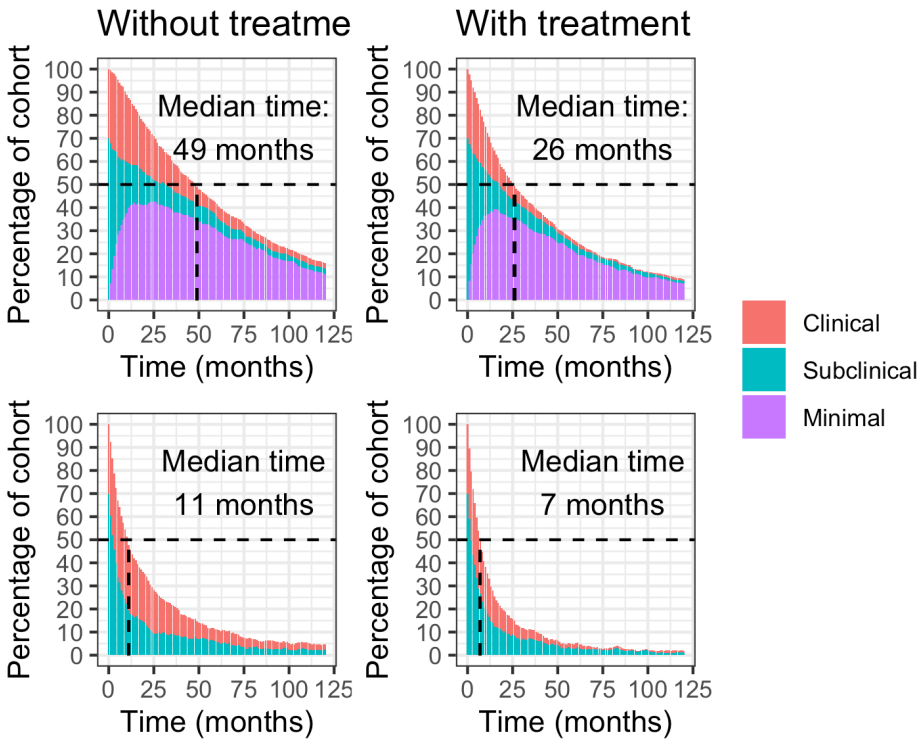


Figure 21: Median duration of infectious and all disease with and without treatment, starting with a cohort of 30% clinical and 70% subclinical individuals

B10 Sensitivity Analyses

We have run sensitivity analyses on different areas of the main analysis. For the purposes of comparison, we have included the median parameter estimates, and key outputs; median duration of disease, percentage of people clinical, undulating, subclinical, and minimal after 5 years, and the number of people who have died from TB over 10 years.

B10.1 Fitting

B10.1.1 Bootstrap studies

We ran the fit process removing the data from one study at a time. Table 4 shows the key outputs, showing that no one study is driving the fit, with similar results when each study is removed.

Table 4: A summary if the differences in fit when removing one study at a time

study	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.20	0.26	1.51	0.69	0.58	0.32	12	4.66	7.01	2.26	24.90	36.75	0	29.82
Hong Kong Chest Service[43–46]	0.21	0.14	0.69	1.14	1.23	0.32	18	4.25	9.08	4.05	23.30	38.84	0	26.78
Tuberculosis Society of Scotland[29,30]	0.20	0.14	0.69	1.18	1.22	0.32	18	4.44	8.71	3.86	24.27	40.12	0	25.22

study	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Puelma[17]	0.20	0.13	0.67	1.23	1.32	0.33	18	4.18	9.04	4.14	23.23	39.95	0	25.94
Beeuwkes[15]	0.21	0.13	0.64	1.10	1.33	0.33	19	3.83	9.19	5.11	23.63	38.95	0	25.74
National Tuberculosis Institute[33–41]	0.23	0.14	0.65	1.34	1.46	0.32	18	3.47	9.80	4.15	21.99	38.99	0	28.14
Aneja[42]	0.19	0.13	0.65	0.89	1.11	0.33	19	4.46	8.34	5.08	24.44	38.36	0	25.35
Bobrowitz[18,19]	0.22	0.14	0.73	1.17	1.20	0.33	17	4.06	8.70	3.56	22.66	38.52	0	28.77
Borgen[24,25]	0.21	0.14	0.70	1.11	1.17	0.32	18	4.54	8.97	4.00	23.47	38.90	0	26.97
Norregaard[48]	0.20	0.14	0.68	1.21	1.28	0.33	18	4.07	8.65	3.99	23.72	39.19	0	26.84
Cowie[47]	0.20	0.14	0.69	1.13	1.21	0.32	19	4.75	8.55	3.90	23.89	39.26	0	26.03

study	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Puffer[16]	0.23	0.15	0.74	1.07	1.11	0.33	18	4.79	8.40	4.29	21.18	39.10	0	28.33
Marshall[23]	0.21	0.14	0.70	1.16	1.24	0.32	18	3.98	9.01	3.78	22.93	39.90	0	27.12
Anastasatu[49]	0.22	0.17	0.87	0.93	0.94	0.33	17	4.57	7.58	3.14	22.74	38.87	0	29.31
Downes[14]	0.22	0.15	0.76	1.22	1.35	0.32	17	3.83	8.94	3.50	23.40	37.49	0	28.88
Pamra[32]	0.18	0.12	0.56	2.26	2.56	0.33	20	2.78	13.69	3.31	24.35	42.32	0	21.66
Okada[13]	0.21	0.13	0.70	1.48	1.43	0.32	19	4.26	9.69	3.54	23.06	40.65	0	25.50
Sikand[28]	0.20	0.14	0.69	1.12	1.22	0.33	18	4.21	8.47	4.05	23.33	40.47	0	26.07
Breu[27]	0.21	0.16	0.70	0.92	1.05	0.32	18	4.55	8.60	5.20	22.53	38.65	0	27.08
Manser[26]	0.20	0.14	0.70	0.90	0.93	0.31	19	5.13	8.36	5.15	23.93	39.90	0	24.61

study	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Bosworth[20,21]	0.19	0.13	0.62	1.39	1.43	0.34	18	3.96	10.30	3.82	23.58	42.54	0	23.57
Bosworth[20,22]	0.20	0.14	0.66	1.24	1.35	0.33	19	4.16	9.42	4.40	24.32	39.08	0	25.77
Frimodt-Moller[31]	0.20	0.14	0.68	1.14	1.23	0.32	18	4.38	8.86	4.59	22.86	38.71	0	26.82

B10.1.2 Change duration of infectiousness

The duration of infectiousness holds a fixed value of 2 years in the main analysis. In table 5 we change this fixed value to 18 months and 3 years.

Table 5: A summary of the differences in fit when testing different durations of infectiousness

duration	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.19	0.26	1.54	0.67	0.57	0.32	12	4.87	6.70	2.30	24.65	37.54	0	29.57
18 months	0.21	0.14	0.73	1.06	1.11	0.32	18	4.19	8.28	4.09	24.07	38.96	0	26.76
3 years	0.20	0.13	0.63	1.24	1.39	0.32	19	4.34	9.09	4.69	23.69	40.17	0	24.72

B10.1.3 Change proportion of minimal "true minimals"

Table 6: A summary of the differences in fit when changing the proportion of people that have x-ray changes due to TB disease

prop_min	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.19	0.26	1.54	0.67	0.57	0.32	12	4.87	6.70	2.30	24.65	37.54	0	29.57
100%	0.25	0.11	0.61	1.16	1.37	0.31	19	4.18	8.96	4.57	20.71	37.66	0	29.72
60%	0.25	0.11	0.63	1.08	1.22	0.30	19	4.40	8.81	4.63	21.27	36.12	0	30.33

B10.1.4 Change assummption on persistent symptoms

The studies where there was only information about the symptom state at the start, in the main analysis, we have assumed this symptom state persists. These data were from minimal to either subclinical or clinical, and so for this sensitivity, we have changed them all to minimal to infectious. The results are in table 7.

Table 7: A summary of the difference in fit, and the subsequent analyses, when removing the assumption that symptoms persist

symptoms assumption	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.19	0.26	1.54	0.67	0.57	0.32	12	4.87	6.70	2.30	24.65	37.54	0	29.57

symptoms assumption	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
no	0.20	0.17	0.85	0.77	0.81	0.32	16	4.66	7.18	4.14	23.75	38.04	0	27.90

B10.2 Cohort model

B10.2.1 Parameter values

The method of parameter choice for the simulation was one of three. In the main analysis, each step of the model, for each individual, the relevant parameters were chosen randomly from the posterior distribution. For the other two alternatives, we randomly sampled the parameters at the start of the simulation for each individual and fixed them for the whole run, and the other used the median parameters for each person. Table 8 shows that there is very little difference between either method overall.

Table 8: A summary of the differences in analyis when testing different method of parameter choice for the cohort model

method	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.19	0.26	1.54	0.67	0.57	0.32	12	4.87	6.70	2.30	24.65	37.54	0	29.57
fixed	0.19	0.26	1.54	0.67	0.57	0.32	12	4.75	7.20	2.03	24.67	37.35	0	29.37
median	0.19	0.26	1.54	0.67	0.57	0.32	12	4.58	6.63	2.14	24.16	38.74	0	29.64

B10.2.2 Treatment

Treatment was added to the model to simulate a case detection rate for a care pathway initiated by self reported symptoms. When considered in the main analysis, the case detection rate was implemented at 70%. In table 9 we compare the difference between case detection rates at 50%, 70%, and 90%.

Table 9: A summary of the differences in analysis when testing different passive case detection rates

treatment	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.19	0.26	1.54	0.67	0.57	0.32	12	4.87	6.70	2.30	24.65	37.54	0.00	29.57
0.5	0.19	0.26	1.54	0.67	0.57	0.32	8	1.21	4.06	1.24	18.44	21.76	29.16	26.11
0.7	0.19	0.26	1.54	0.67	0.57	0.32	7	0.79	3.64	1.01	17.40	19.03	33.89	25.54
0.9	0.19	0.26	1.54	0.67	0.57	0.32	7	0.56	2.91	1.11	16.77	16.65	39.32	23.81

B10.2.3 Trajectories

The trajectories are based on two variables, the proportion of time in a single state, and the number of state changes both over the previous 12 months. The main analysis defines undulating as less than nine months in a single state or 3 or more changes in state. In table 10 we compare the definition of undulation as less than 8 months or less than 10 months, whilst keeping the number of state changes fixed at 3 or more. In table 11 we compare the definition of undulation with less than 9 months fixed, and the state changes as either 2 or more, or 4 or more.

Table 10: A summary of the differences in analysis when varying the threshold for undulating trajectories

states	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.19	0.26	1.54	0.67	0.57	0.32	12	4.87	6.70	2.30	24.65	37.54	0	29.57
7	0.19	0.26	1.54	0.67	0.57	0.32	12	5.48	2.67	4.02	25.84	37.64	0	30.23
8	0.19	0.26	1.54	0.67	0.57	0.32	12	5.32	4.61	2.75	25.36	37.54	0	30.26
10	0.19	0.26	1.54	0.67	0.57	0.32	12	4.15	9.17	1.52	23.28	37.52	0	30.01
11	0.19	0.26	1.54	0.67	0.57	0.32	12	3.89	11.76	1.05	21.86	37.82	0	29.50

Table 11: A summary of the differences in analysis when varying the threshold for undulating trajectories

changes	min-out	min-sub	sub-min	sub-clin	clin-sub	mort	duration	clinical5	undulating5	subclinical5	minimal5	dead10	treated5	recovered5
Main	0.19	0.26	1.54	0.67	0.57	0.32	12	4.87	6.70	2.30	24.65	37.54	0	29.57
2	0.19	0.26	1.54	0.67	0.57	0.32	12	3.92	9.47	1.71	23.22	37.68	0	29.13
4	0.19	0.26	1.54	0.67	0.57	0.32	12	5.11	6.55	2.09	24.19	37.61	0	29.93

B11 References

1 GOTHI GD. NATURAL HISTORY OF TUBERCULOSIS. *Indian Journal of Tuberculosis* 1977; **25**.

2 Barry CE, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J *et al.* The spectrum of latent tuberculosis: Rethinking the biology and intervention strategies. *Nat Rev Microbiol* 2009; **7**: 845–855.

3 Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC *et al.* Tuberculosis. *Nat Rev Dis Primers* 2016; **2**: 16076.

4 Lin PL, Flynn JL. The End of the Binary Era: Revisiting the Spectrum of Tuberculosis. *The Journal of Immunology* 2018; **201**: 2541–2548.

5 Houben RMGJ, Esmail H, Emery JC, Joslyn LR, McQuaid CF, Menzies NA *et al.* Spotting the old foe—revisiting the case definition for TB. *The Lancet Respiratory Medicine* 2019; **7**: 199–201.

6 Tiemersma EW, Werf MJ van der, Borgdorff MW, Williams BG, Nagelkerke NJD. Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: A systematic review. *PLoS ONE* 2011; **6**: e17601.

7 Ragonnet R, Flegg JA, Brilleman SL, Tiemersma EW, Melsew YA, McBryde ES *et al.* Revisiting the natural history of pulmonary tuberculosis: A bayesian estimation of natural recovery and mortality rates. *Clin Infect Dis* 2020. doi:10.1093/cid/ciaa602.

8 Szucs E. Sputum examination in so-called closed tuberculosis. *JAMA* 1926; **86**: 946.

9 Tattersall WH. The survival of sputum-positive consumptives: A study of 1,192 cases in a county borough between 1914 and 1940. *Tubercle* 1947; **28**: 85–96.

10 Lindhardt M. *The statistics of pulmonary tuberculosis in denmark, 1925-1934; a statistical investigation on the occurrence of pulmonary tuberculosis in the period 1925-1934, worked out on the basis of the danish national health service file of notified cases and of deaths.* EMunksgaard: Copenhagen, 1939.

11 Blahd M, Leslie EI, Rosenthal SR. Infectiousness of the 'closed case' in tuberculosis. *Am J Public Health Nations Health* 1946; **36**: 723–726.

12 Association NT. *Diagnostic standards and classification of tuberculosis.* 1940 ed. New York, N.Y., 1940http://hdl.handle.net/2027/coo.31924089435949.

13 Okada K, Onozaki I, Yamada N, Yoshiyama T, Miura T, Saint S *et al.* Epidemiological impact of mass tuberculosis screening: A 2 year follow-up after a national prevalence survey. 2012; **NA**.

14 Downes J. The study of mortality among individuals with active pulmonary tuberculosis. 1938; **16**: 304–317.

15 Beeuwkes H, Hahn RG, Putnam P. A survey of persons exposed to tuberculosis in the household. *Am Rev Tuberc* 1942; **45**: 165–193.

16 Puffer RR, Stewart HC, Gass RS. Tuberculosis according to age, sex, family history, and contact. *Am Rev Tuberc* 1945; **51**: 295–311.

17 Puelma HO, Grebe G. Analysis of one hundred cases of minimal pulmonary tuberculosis. *Dis Chest* 1945; **11**: 375–379.

18 BOBROWITZ ID., HURST A., MARTIN M. Minimal tuberculosis; the prognosis and clinical significance of a sanatorium treated group. *American review of tuberculosis* 1947; **56**: 110–125.

19 BOBROWITZ ID., HURST A. Minimal tuberculosis; problems in roentgenologic interpretation. *Radiology* 1949; **52**: 519–32.

20 Bosworth E.B., Alling D.W. The after-history of pulmonary tuberculosis. I. Methods of evaluation. *American Review of Tuberculosis* 1954; **69**: 37–49.

21 Lincoln NS, Bosworth EB, Alling DW. The after-history of pulmonary tuberculosis. III. Minimal tuberculosis. *Am Rev Tuberc* 1954; **70**: 15–31.

22 Alling D.W., Bosworth E.B., Lincoln N.S. The after-history of pulmonary tuberculosis. V. Moderately advanced tuberculosis. *American Review of Tuberculosis* 1955; **71**: 519–528.

23 Streptomycin treatment of pulmonary tuberculosis: A medical research council investigation. *BMJ* 1948; **2**: 769–782.

24 Borgen L., Meyer SN., Refsum E. Mass photofluorography, tuberculin testing, and BCG vaccination in the district of aker (norway) 1947-49. *Acta tuberculosea Scandinavica* 1951; **25**: 327–55.

25 Refsum E. Mass investigation by photofluorography; an illustration of the value of the method in combating tuberculosis. *Acta tuberculosea Scandinavica* 1952; **27**: 288–302.

26 Manser H. [Tuberculosis in aged and its course during sanatorium treatment]. *Schweizerische Zeitschrift fur Tuberkulose Revue suisse de la tuberculose Rivista svizzera della tubercolosi* 1953; **10**: 65–82.

27 K. B. [Public health x-ray diagnosis of closed pulmonary tuberculosis later proved contagious]. *Beitrage zur Klinik der Tuberkulose und spezifischen Tuberkulose-Forschung* 1954; **111**: 437–44.

28 Sikand BK, Narain R, Mathur GP. Incidence of TB as judged by re-surveys. A study of delhi police. *Indian Journal of Tuberculosis* 1959; **6**: 73–83.

29 Tuberculosis Society of Scotland. A controlled trial of chemotherapy in pulmonary tuberculosis of doubtful activity. Report from the research committee of the tuberculosis society of scotland. *Tubercle* 1958; **39**: 129–137.

30 Society ST. A controlled trial of chemotherapy in pulmonary tuberculosis of doubtful activity. *Tubercle* 1963; **44**: 39–46.

31 Frimodt-Moller J. Results of treatment of non-bacillary tuberculosis in a domiciliary treatment programme, preliminary report. Proceedings of the 20th tuberculosis and chest diseases workers' conference. 1965, p 133.

32 Pamra SP, Mathur GP. Effects of chemoprophylaxis on minimal pulmonary tuberculosis lesions of doubtful activity. *Bull World Health Organ* 1971; **45**: 593–602.

33 National Tuberculosis Institute B. Tuberculosis in a rural population of south india: A five-year epidemiological study. *Bull World Health Organ* 1974; **51**: 473–88.

34 Chakraborty AK, Singh H, Srikantan K, Rangaswamy KR, Krishnamurthy MS, Stephen JA. Tuberculosis in a rural population of south india: Report on five surveys. *Indian Journal of Tuberculosis* 1982; **29**: 153–167.

35 Gothi GD, Chakraborty AK, Krishnamurthy VV, Banerjee GC. Prevalence and incidence of sputum negative active pulmonary tuberculosis and fate of pulmonary rradiological abnormalities found in rural population. *Indian Journal of Tuberculosis* 1978; **25**: 122–131.

36 Krishnamurthy VV, Nair SS, Gothi GD. A comparison of new cases (incidence cases) who had come from differeent epidemiological groups in the population. *Indian Journal of Tuberculosis* 1978; **25**: 144–146.

37 Gothi GD, Chakraborty AK, Jayalakshmi MJ. Incidence of sputum positive tuberculosis in different epidemiological grroups during five year follow up of a rural population in south india. *Indian Journal of Tuberculosis* 1978; **25**: 83–91.

38 Krishnamurthy VV, Nair SS, Gothi GD, Chakraborty AK. Incidence of tuberculosis among newly infected population and in relation to the duration of infected status. *Indian Journal of Tuberculosis* 1976; **23**.

39 Gothi GD, Nair SS, Chakraborty AK, Ganapathy KT. Five year incidencee of tuberculosis and crude mortality in relation to non-specific tuberculin senssitivity. *Indian Journal of Tuberculosis* 1976; **23**.

40 Chakraborty AK, Gothi GD. Relapses among naturally cureed cases of pulmmonary tuberculosis. *Indian Journal of Tuberculosis* 1976; **23**: 8–13.

41 Gothi GD, Chakraborty AK, Banerjee GC. Interpretation of photofluorograms of active pulmonary tuberculosis patients found in epidemiological survey and their five year fate. *Indian Journal of Tuberculosis* 1974; **11**: 90–97.

42 Aneja KS, Gothi GD, Samuel GER. Controlled study of the effect of spcific treatment on bacteriological status of 'suspect cases'. *Indian J Tuberc* 1979; **26**: 50–57.

43 Hong Kong Chest Service/Tuberculosis Research Centre MMRC. A controlled trial of 2-month, 3-month, and 12-month regimens of chemotherapy for sputum smear-negative pulmonary tuberculosis: The results at 30 months. *Am Rev Respir Dis* 1981; **124**: 138–142.

44 Hong Kong Chest Service/Tuberculosis Research Centre MMRC. A controlled trial of 2-month, 3-month, and 12-month regimens of chemotherapy for sputum-smear-negative pulmonary tuberculosis. *Am Rev Respir Dis* 1984; **130**: 23–28.

45 Hong Kong Chest Service/Tuberculosis Research Centre MMRC. Sputum smear negative pulmonary tuberculosis controlled trial of 3-month and 2-month regimend of chmotherapy: First report. *The Lancet* 1979; **313**: 1361–1363.

46 Hong Kong Chest Service/Tuberculosis Research Centre MMRC. A study of the characteristics and course of sputum smear-negative pulmonary tuberculosis. *Tubercle* 1981; **62**: 155–167.

47 Cowie RL. Diagnosis of sputum smear- and sputum culture-negative pulmonary tuberculosis. *SAMJ* 1985; **68**: 878.

48 Norregaard J. Abacillary pulmonary tuberculosis. *Tubercle* 1990; **71**: 35–38.

49 Anastasatu C, Berceea O, Corlan E. Controlled clinical trial on smear negatice, x-ray positive new cases, with the view to establishing if and how to treat them. *Bull Int Union Tub* 1985; **60**: 108–109.

50 Efficacy of various durations of isoniazid preventive therapy for tuberculosis: Five years of follow-up in the IUAT trial. *Bulletin of the World Health Organization* 1982; **60**: 555–564.

51 Styblo K, Dajkova D, Kubik A, Langerova M, Radkovsky J. Epidemiological and clinical study of tuberculosis in the district of kolin, czechoslovakia. ; : 56.

52 Frascella B, Richards AS, Sossen B, Emery JC, Odone A, Law I *et al.* Subclinical tuberculosis disease - a review and analysis of prevalence surveys to inform definitions, burden, associations and screening methodology. *Clin Infect Dis* 2020. doi:10.1093/cid/ciaa1402.