Precision targeting of preventative therapy for tuberculosis

Dr Rishi Kumar Gupta
University College London
PhD Thesis in Infectious Disease Epidemiology
March 2021

Supervisors:
Prof Ibrahim Abubakar
Prof Marc Lipman
Prof Mahdad Noursadeghi
Prof Andrew Copas

I, Rishi Kumar Gupta, 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.

Acknowledgements

I would first like to thank my supervisors (Ibrahim, Marc, Maddy and Andrew) for their unwavering support – from the point of preparing my Fellowship applications to the completion of this work. I am extremely grateful for their direction, technical advice and mentorship throughout the project. I would also like to pay tribute and express my gratitude to Steve Lawn for being my first mentor in TB research and for inspiring me to pursue this career path. I would have loved to have had the opportunity to discuss this work with him.

My project has been heavily dependent upon access to invaluable existing datasets. I am extremely grateful to my supervisors for facilitating this, along with the many contributors to the UK PREDICT study, and the external studies that contributed data to my individual participant data meta-analyses.

Many others at UCL have also helped me along the way. I am particularly grateful to Carolin Turner, Claire Calderwood and Maria Krutikov for being the second reviewers for my systematic reviews, Alexei Yavlinsky for all his work on the PERISKOPE-TB online tool, Matteo Quartagno for his patience while teaching me about handling missing data, and to Lele Rangaka for her ongoing advice and mentorship throughout.

Finally, and most importantly, I'd like to thank my parents for their love and guidance over the last 35 years, and Jo and Sunny for always being there for me. None of this work would have been possible without you.

Abstract

Background

Scale-up of preventative treatment for tuberculosis (TB) represents a cornerstone of global control efforts. I examined a range of approaches to enable more precise targeting of preventative treatment to people at highest risk.

Methods

I evaluated whether prognostic tests for TB (tuberculin skin test (TST), QuantiFERON Gold-in-tube (QFT-GIT) and T-SPOT.TB) may be optimised by implementing higher thresholds, or by a newer generation assay (QuantiFERON-TB Gold Plus; QFT-Plus). Next, I conducted a systematic review and individual participant data meta-analysis (IPD-MA) to examine TB risk among people tested for latent infection (LTBI) in settings with low TB transmission and to develop a multivariable prognostic model for incident TB. Finally, I performed a systematic review and IPD-MA of whole-blood RNA sequencing data to evaluate blood transcriptomic signatures as next-generation biomarkers.

Results

In a UK cohort of 9,610 adults, higher TST, QFT-GIT and T-SPOT.TB results were associated with increased incident TB risk. Implementing higher cut-offs led to a marginal improvement in positive predictive value, but at the cost of a marked loss in sensitivity. The newer generation QFT-Plus had similar predictive ability. In a pooled dataset of >80,000 participants from 18 cohort studies, TB risk was heterogeneous among people with LTBI, even after stratification by indication for testing. I developed and validated a multivariable prognostic model, which incorporates quantitative LTBI test results and clinical covariates, and demonstrated strong potential for clinical utility to inform provision of preventative treatment. Among 1,126 whole-blood RNA sequencing samples, eight transcriptomic signatures (comprising 1-25 transcripts) performed similarly for predicting incident TB, but only met global accuracy benchmarks over a 3-6 month time-horizon.

Conclusions

Personalised risk estimates integrating quantitative LTBI test results and clinical covariates may facilitate more precise targeting of preventative treatment. Blood transcriptomic biomarkers show promise, but only represent short-term TB risk. Future research priorities are highlighted.

Impact Statement

This thesis will have direct benefits to clinical care and healthcare policy, along with future research on tuberculosis (TB) and other diseases.

In Chapter 2, I demonstrate that, while higher latent TB test results are associated with increased TB risk, implementing higher thresholds to guide initiation of preventative treatment would risk missing most incident TB cases. These findings have important implications for policymakers by demonstrating that current diagnostic thresholds used for latent TB screening are appropriate, and should not be increased programmatically in settings aiming towards pre-elimination. Nonetheless, high quantitative test results may be considered by clinicians when assessing individual patients and could prompt provision of stronger encouragement to initiative preventative treatment. I have disseminated this work through publication in the *American Journal of Respiratory and Critical Care Medicine* and presentation at the European Respiratory Society Congress 2019.

In Chapter 3, I show that a newer generation latent TB assay (QuantiFERON TB Gold Plus) is likely to perform similarly to previous iterations and that the $CD8^+$-specific response is unlikely to have prognostic ability. My findings support clinical decision-making by demonstrating that newer generation results can be interpreted in line with previous versions and have been published in *Annals of the American Thoracic Society*.

In Chapter 4, I pooled data from 18 studies to develop the PERISKOPE-TB prognostic model, which estimates the risk of future TB for people tested for latent infection in settings with low incidence worldwide. The prognostic tool is available at http://periskope.org and facilitates shared decision-making between clinicians and patients when considering preventative treatment initiation, with direct clinical impact. PERISKOPE-TB could also be considered for inclusion in future clinical guidelines by policymakers. This work has been widely disseminated through publication in *Nature Medicine* and presentation at the British Thoracic Society Meeting 2021, and received national media coverage in *The Guardian*. The analysis methods also provide a framework for future studies seeking to develop prediction models using multi-source data. I have subsequently applied these methods to develop a prognostic model for adults hospitalised with COVID-19

across the UK, published in *The Lancet Respiratory Medicine* and implemented at https://isaric.net/risk for clinical use.

In Chapter 5, I examined the performance of blood transcriptomic biomarkers for prediction of future TB. I show that multiple biomarkers appear promising, but their performance is highly time-dependent. These findings have important implications for future studies evaluating the potential impact of blood transcriptomic biomarkers in TB screening, and suggest they are likely to be best applied among target populations at high short-term risk of disease, such as recent TB contacts. I published this work in *The Lancet Respiratory Medicine* and presented it at the TBScience 2019 conference in India. This work also provides a methodological framework for studies seeking to directly compare performance of multiple biomarkers or prediction tools for TB or other diseases. The transferability of these methods is demonstrated by my subsequent work comparing the diagnostic accuracy of candidate transcriptomic signatures for SARS-CoV-2 infection, and the performance of clinical prognostic models to predict COVID-19 outcomes.

Contents

List of Figures

List of Tables

List of Abbreviations

ACS	Adolescent cohort study
AIC	Akaike information criteria
AUROC	Area under the receiver operating characteristic curve
BCG	Bacillus Calmette–Guérin
CFP-10	Culture filtrate protein 10
CFU	Colony-forming units
CI	Confidence interval
CORTIS	Correlate of risk targeted intervention study
CPM	Counts per million
CT	Computed tomography
ELISA	Enzyme-linked immunosorbent assay
ELISPOT	Enzyme-linked immunosorbent spot
ESAT-6	Early secretory antigenic target 6
GC6-74	Grand challenges 6-74 household TB contacts study
HIV	Human immunodeficiency virus
IGRA	Interferon gamma release assay
IECV	Internal-external cross-validation
IPD-MA	Individual participant data meta-analysis
IQR	Interquartile range
IRR	Incidence rate ratio
LAM	Lipoarabinomannan
LTBI	Latent tuberculosis infection
M. tuberculosis	*Mycobacterium tuberculosis complex*
MRI	Magnetic resonance imaging
NPV	Negative predictive value
PCA	Principal component analysis
PCR	Polymerase chain reaction
PERISKOPE-TB	Personalised risk predictor for incident tuberculosis
PET	Positron emission tomography
PPV	Positive predictive value
PT	Preventative treatment
QFT	QuantiFERON
QFT-GIT	QuantiFERON Gold in-Tube
QFT-Plus	QuantiFERON TB Gold Plus

RCT	Randomised-controlled trial
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
SVM	Support vector machine
TB	Tuberculosis
TPP	Target product profile
TPM	Transcripts per million
TRIPOD	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
TST	Tuberculin skin test
UK PREDICT	UK prognostic evaluation of diagnostic IGRAs consortium
WHO	World Health Organization

1 Introduction

1.1 Natural history of *Mycobacterium tuberculosis* infection

1.1.1 Overview of natural history

Tuberculosis disease is caused by *Mycobacterium tuberculosis complex* (*M. tuberculosis*), which spreads from person-to-person via aerosolised particles. *M. tuberculosis* is an obligate pathogen and is transmitted from individuals with pulmonary disease[1]. Following infection, innate and adaptive host immune responses are sufficient to control *M. tuberculosis* in most people. A subset of individuals, however, progress to TB disease. The host immunological determinants of outcome following *M. tuberculosis* infection are imperfectly understood. Deficient responses relating to T cells (as occurs in the context of CD4 cell depletion in people living with HIV infection (PLHIV)[2]), tumour necrosis factor (as demonstrated in people treated with anti-tumour necrosis factor monoclonal antibody therapy[3]) and interferon-gamma responses (as in Mendelian susceptibility to mycobacterial disease[4]) have all been implicated as being associated with increased risk of mycobacterial disease.

1.1.2 The spectrum of M. tuberculosis infection

The natural history of *M. tuberculosis* infection may be considered as a spectrum, from “latent” infection to active disease. Latent TB infection (LTBI) can either be defined clinically or theoretically. The clinical definition usually refers to evidence of immunosensitisation to *M. tuberculosis* (Chapter 1.4), while the theoretical definition refers to the presence of viable *M. tuberculosis* bacilli that remain under immune control. Viability of quiescent *M. tuberculosis* bacilli is very rarely proven, for example by inoculation of guinea pigs with resected granulomas[5,6]. Both clinical and theoretical LTBI definitions are conditional upon the absence of concurrent evidence of TB disease. While the overall lifelong risk of TB among people with clinically defined LTBI has been estimated as 5-15%, this varies markedly between individuals and risk groups[7].

Active TB disease is defined by the presence of symptomatic, microbiological, radiological and/or pathological evidence of disease (Chapter 1.3). TB most commonly affects the lungs, though extra-pulmonary disease can occur at any

anatomical site, most frequently the lymph nodes and pleura. Common symptoms of pulmonary TB include cough and haemoptysis. Symptomatic presentations of extra-pulmonary TB can be highly variable according to the affected disease site. Both pulmonary and extra-pulmonary TB are typically associated with systemic upset, including symptoms of fever, night sweats, weight loss and general malaise.

The transition between latent infection and active disease is likely to be associated with increasing mycobacterial burden, in the absence of symptomatic disease. Recent analyses of TB prevalence surveys have suggested that 36.1-79.7% (median 50.4%) of people with prevalent bacteriologically-confirmed TB are asymptomatic, highlighting a need to support passive case-finding with active case-finding approaches in order to achieve early case detection[8]. This is not a new concept. The importance of early pre-symptomatic case detection has been historically described:

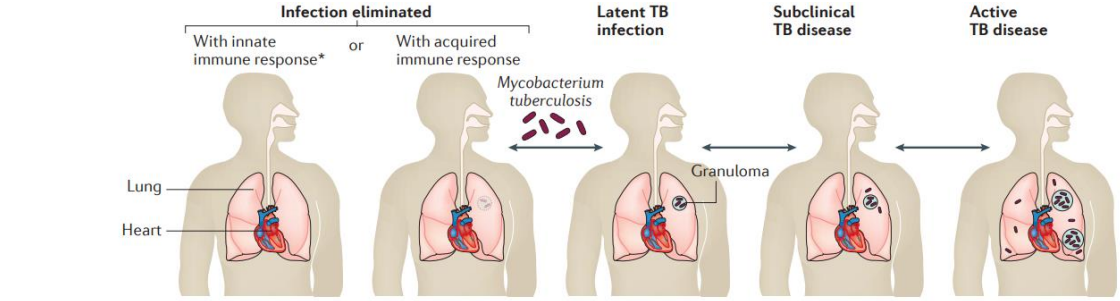
> *"Tuberculosis of the lungs begins without any warning to the patient. By the time patients voluntarily come for treatment, the disease is in an advanced stage, and such patients have perhaps already infected numerous other people. There are about a quarter of a million actual cases of pulmonary tuberculosis in the British Isles, and of these about 1,500 between the ages of 15 and 50 die each month. In spite of this, the total number is kept up by a steady flow of recruits. To prevent the disease in the future we must try more persistently to discover patients in the early stages, before they have become infectious, since this endeavour will ultimately mean the conquest of tuberculosis."*
>
> - **The Policy of the National Association for the Prevention of Tuberculosis, 1942**[9]

Recent definitions of the asymptomatic phase between latent infection and active disease have varied. Pai et al proposed a unifying "subclinical" phase, characterised by low (albeit increasing) bacillary burden and intermittent mycobacterial culture positivity (Figure 1-1)[1]. This interpretation was also reinforced by Esmail et al, with the subclinical phase accounting for increasing bacillary burden and associated pathology (Figure 1-2)[10]. A World Health Organization (WHO) target product profile (TPP) document, however, refers to this subclinical phase as "incipient TB", defined as "a prolonged asymptomatic phase of early disease during which pathology evolves"[11]. In contrast, Drain et al recently proposed distinct "incipient" and "subclinical" phases[12]. The former was

described as metabolic mycobacterial activity, indicative of imminent progression to disease, in the absence of microbiological or radiological evidence of disease. The more advanced “subclinical” phase was then described as microbiological or radiological evidence of disease in the absence of symptoms.

Figure 1-1: Spectrum of TB, as described by Pai et al.

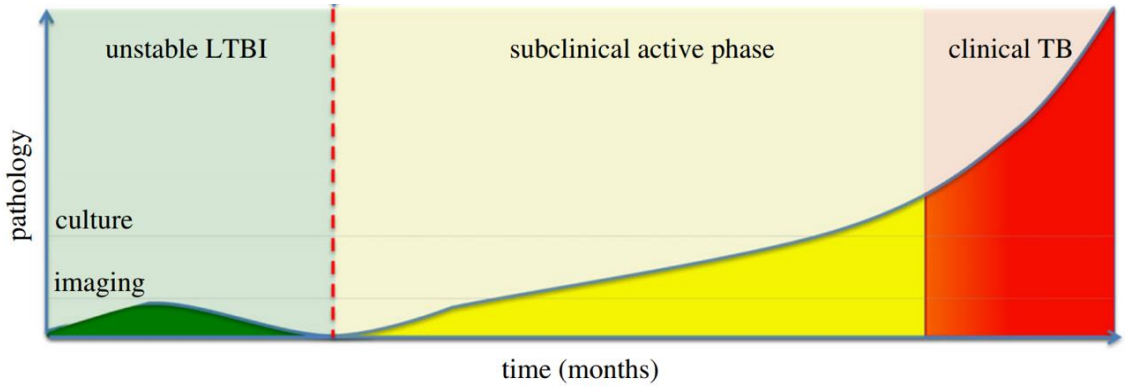
Reprinted by permission from Springer Nature: Pai et al (2016)[1]. Tuberculosis. Nature Reviews Disease Primers 2(1) http://doi.org/10.1038/nrdp.2016.76, Copyright 2016.



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

Figure 1-2: Spectrum of TB, as described by Esmail et al.

Reproduced from Esmail et al (2014)[10]. The ongoing challenge of latent tuberculosis. Phil. Trans. R. Soc. B369: 20130437. http://doi.org/10.1098/rstb.2013.0437 under CC-BY-3.0 license.



Discordant definitions are not a new phenomenon when describing the natural history of TB, as described by Ritter.

> *“Most European writers on medical topics, when they refer to primary or established pulmonary disease, usually designate by such descriptive terms as “early tuberculosis”, as “manifest tuberculosis” or*

> *as "first stage disorder". The expression "incipient tuberculosis" to denote a definite beginning involvement is never used. In speaking of early established pulmonary disease, no matter how long the disease has existed or how much primary involvement may be present, are not such descriptive terms more correct, more appropriate and less misleading than the term "incipient" now in use?"*
>
> - **John Ritter, 1916**[13]

Despite this inconsistency in definitions, it is clear that there is a global appetite to supplant the binary distinction between latent infection and active disease, in favour of a more continuous theoretical framework. In this thesis, the term "incipient" will hereafter be used to describe the asymptomatic phase between latent infection and active disease, as per the WHO definition[11].

1.1.3 Timing of progression from latent infection to TB disease

It is generally considered that progression to TB disease can occur quickly after initial *M. tuberculosis* infection, or at a later time point (so called "reactivation"). In the framework proposed by Drain et al, it is considered that individuals may progress through the spectrum from LTBI and TB disease at varying pace, or may have a more cyclical path with periods of regression, even in the absence of treatment[12]. However, it has long since been recognised that the risk of incident TB declines with increasing time since exposure, with a marked decline in TB incidence rate observed in the placebo arms of an early US trial among household TB contacts (Figure 1-3) [14], and the International Union Against Tuberculosis trial of 28,000 people with fibrotic pulmonary lesions (Figure 1-4)[15].

Figure 1-3: TB incidence in trial of isoniazid versus placebo among household contacts in USA.

Based on original data from Ferebee[14]. Reproduced from: Behr et al (2018)[16]. Revisiting the timetable of tuberculosis. BMJ; 362: k2738 https://doi.org/10.1136/bmj.k2738 under CC-BY-4.0 license.

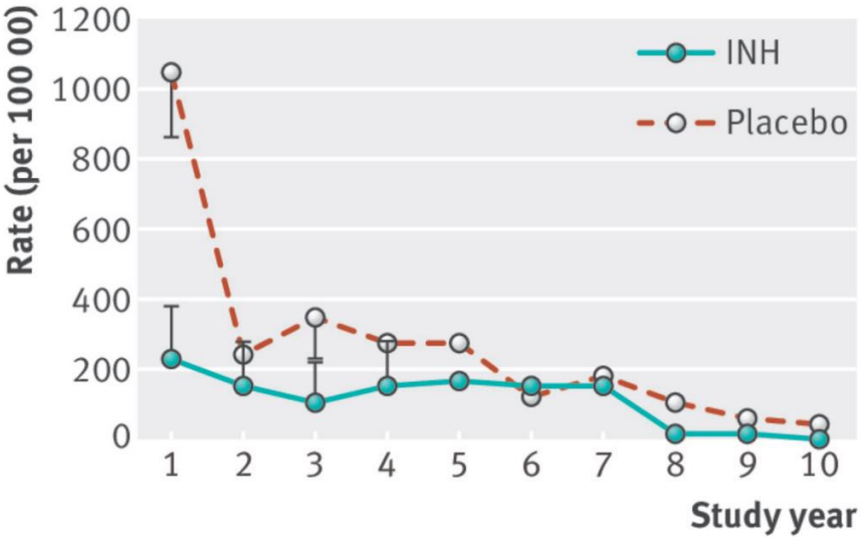
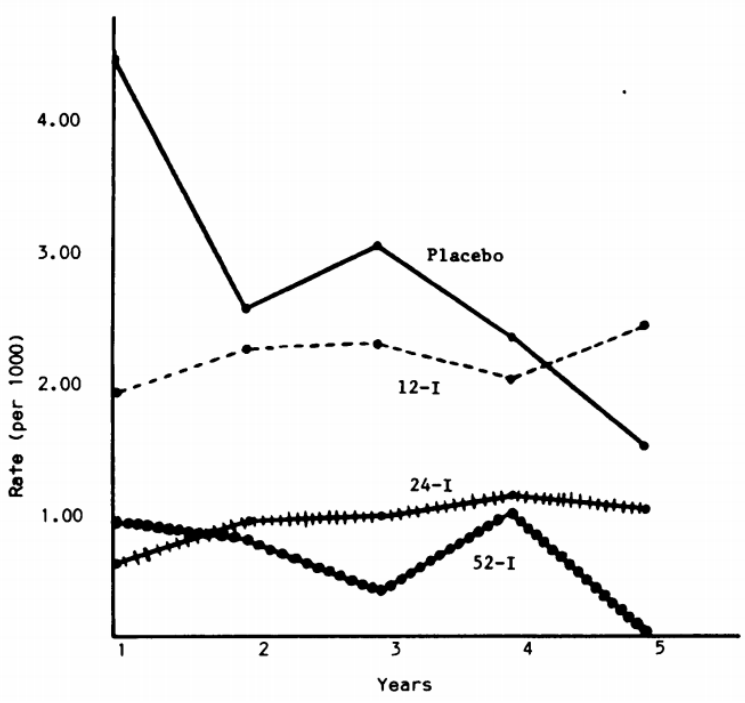


Figure 1-4: TB incidence in International Union Against Tuberculosis trial of people with fibrotic pulmonary lesions.

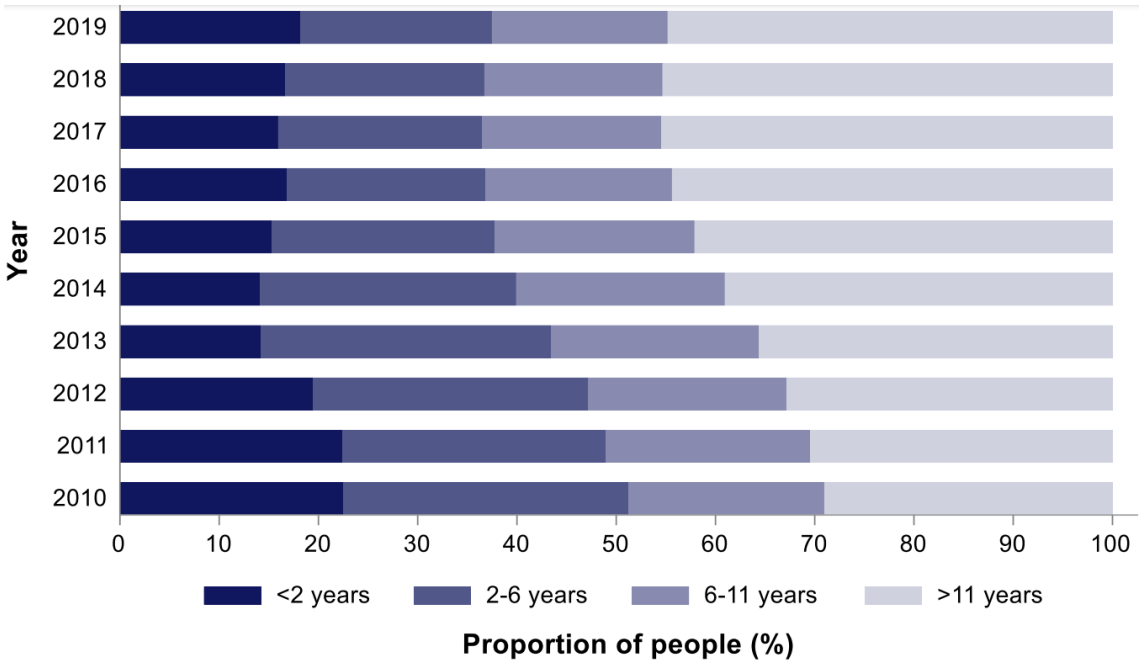
Reproduced from International Union Against Tuberculosis Committee on Prophylaxis[15]. Efficacy of various durations of isoniazid preventive therapy for tuberculosis: five years of follow-up in the IUAT trial. Bull World Health Organ. 1982; 6(4). Licence: CC-BY-3.0 IGO. 12-I, 24-I and 52-I indicate 12-52 weeks of isoniazid.



Reductions in TB incidence with increasing time since arrival have also been described among migrants from high to low TB incidence countries[17,18]. Together with post-mortem evidence that >95% of lung granuloma lesions are sterile (following guinea pig inoculation)[6], these data have been mooted to suggest that the vast majority of incident TB cases occur in the first two years following *M. tuberculosis* infection and that late reactivation after this interval may be a rare phenomenon[16]. However, 2019 surveillance data from England show that almost three quarters of cases are born outside the UK, with median time from entry to TB notification of 9 years (interquartile range (IQR) 3-18; Figure 1-5)[19]. Only 26% of foreign-born patients were known to have travelled abroad in the two years preceding diagnosis. While the relative contributions of local transmission and acquisition of infection during overseas travel are challenging to quantify, these data suggest that late reactivation among migrants may continue to have a significant role in the epidemiology of TB in low incidence countries (Chapter 1.2).

Figure 1-5: Time since entry for migrants notified with TB in England 2010-2019.

Reproduced from: Public Health England Tuberculosis in England 2020 report under Open Government Licence v3.0[19].



1.2 Tuberculosis epidemiology

1.2.1 Global trends in burden of disease

TB case fatality rates declined steeply during initiation of the combination anti-microbial chemotherapy era, which began in 1952 (Figure 1-6)[20]. While the 20th century saw marked declines in TB incidence and mortality rates in high-income countries, the onset of these reductions predated the advent of combination anti-microbials, as demonstrated in historical data from London and New York[20]. However, many low and middle-income countries have been left behind in TB control. For example, in Cape Town, declines in TB incidence and mortality were not evident in the pre-antimicrobial chemotherapy era, in contrast to London and New York[20]. Following a temporary reduction in Cape Town TB incidence in the 1950s-1970s, rates returned to pre-chemotherapy levels by 1980 and increased further over subsequent decades, in part driven by the HIV epidemic.

These trends have led to marked heterogeneity in the global burden of TB that persists to this day, fuelled by socioeconomic disparity and exacerbated further in regions with high HIV prevalence. TB continues be a leading cause of morbidity and mortality globally, with 10 million incident cases and 1.4 million deaths estimated in 2019[21]. County-level incidence rates vary markedly from <5/100,000 to >500/100,000, with two thirds of 2019 cases occurring in the eight worst affected countries.

Between 2015 and 2019, global estimated TB incidence rates fell by 9% (Figure 1-7), while mortality fell by 14% (Figure 1-8). However, these reductions fall short of the targets set out in the WHO End TB Strategy, which aims to reduce TB incidence rates and mortality by 20% and 35% by 2020, and 90% and 95% by 2035, respectively[22].

Figure 1-6: Changes in TB case-fatality, incidence and mortality from 1910-2010 in Cape Town, London and New York.

Adapted from original data from Hermans et al[20]. Dashed vertical line indicates start of combination anti-microbial therapy era

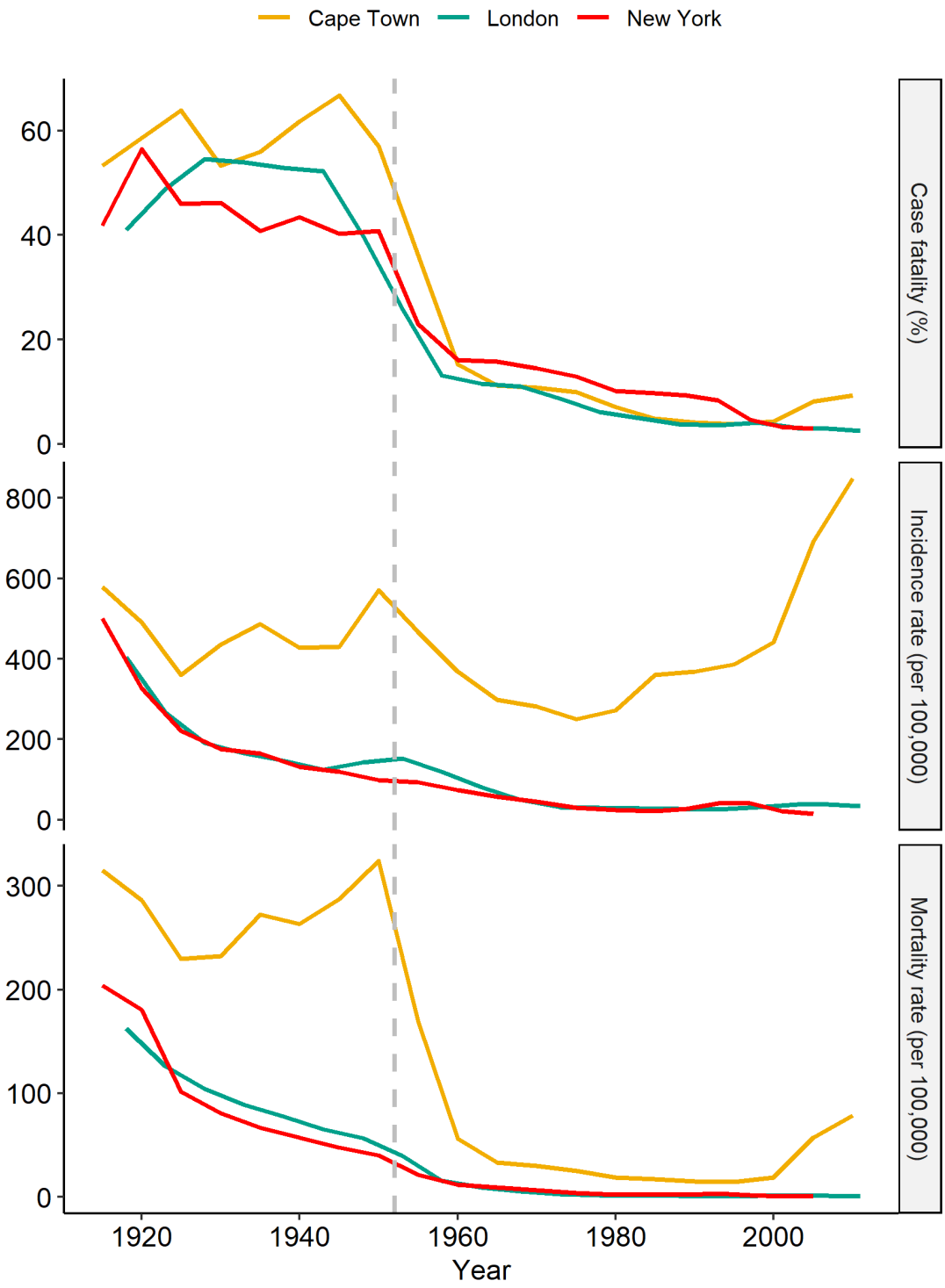


Figure 1-7: Global TB incidence rates 2000 to 2020.

Reproduced from Global TB Report, World Health Organization 2020[21]. Licence: CC BY-NC-SA 3.0 IGO.

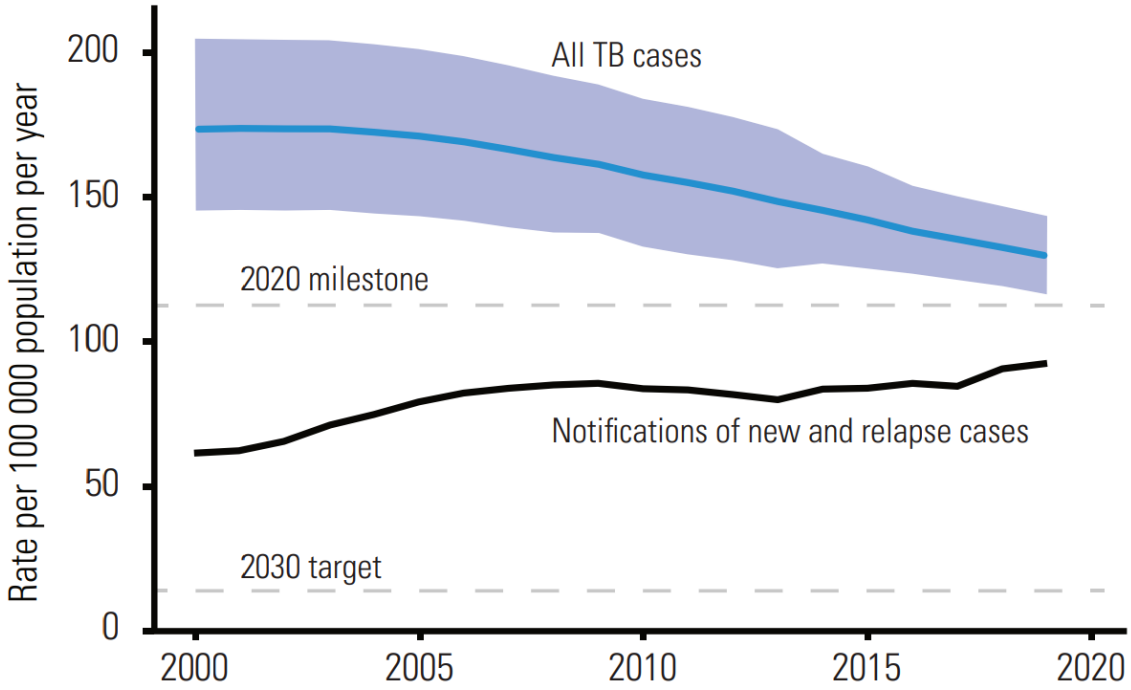
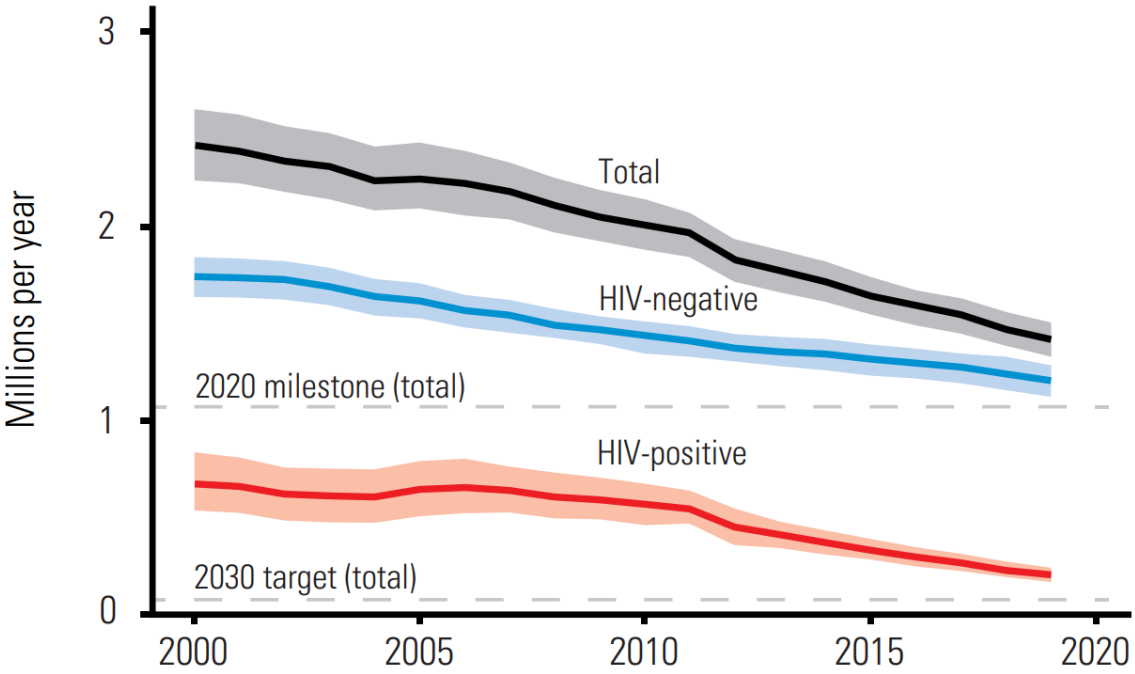


Figure 1-8: Global TB mortality 2000 to 2020.

Reproduced from Global TB Report, World Health Organization 2020[21]. Licence: CC BY-NC-SA 3.0 IGO.



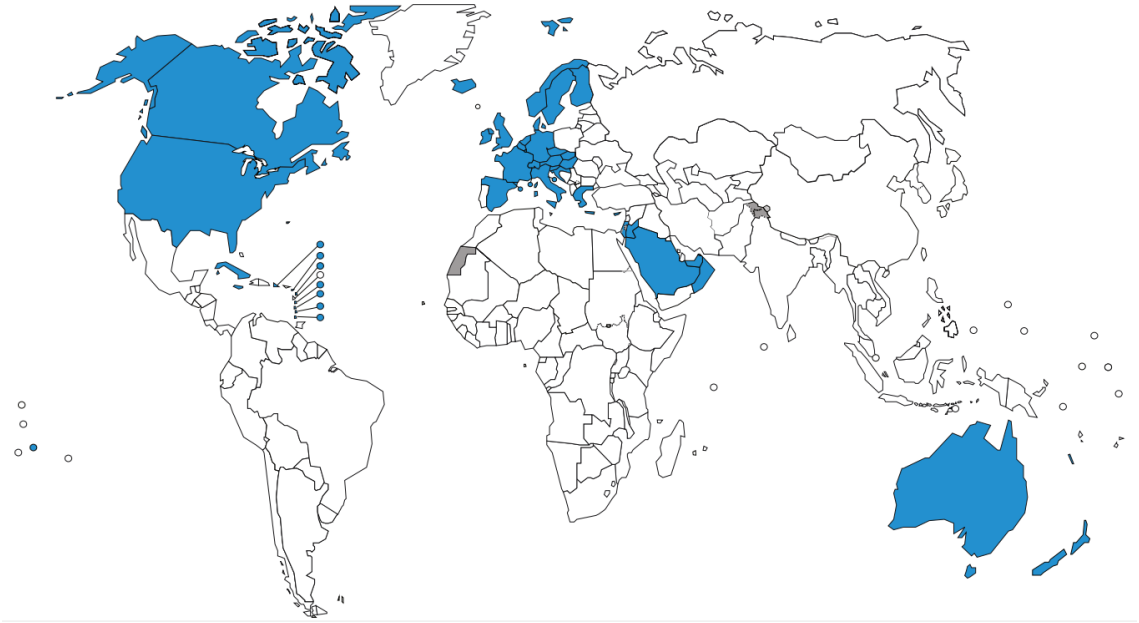
1.2.2 Trends in countries with low TB incidence

The WHO defines low TB incidence countries as those with annual incidence rates <10/100,000 persons[21]. A total of 54 countries, predominantly in North America and Western Europe, met this definition in 2019 (Figure 1-9).

Progress towards the WHO End TB Strategy targets has been more pronounced in the WHO European Region, with a 19% reduction in TB incidence rates and 31% reduction in mortality from 2015-2019. The End TB Strategy 2035 target for countries with low TB incidence is to aim towards pre-elimination (defined as annual incidence <1/100,000).

Figure 1-9: Countries with low TB incidence in 2019.

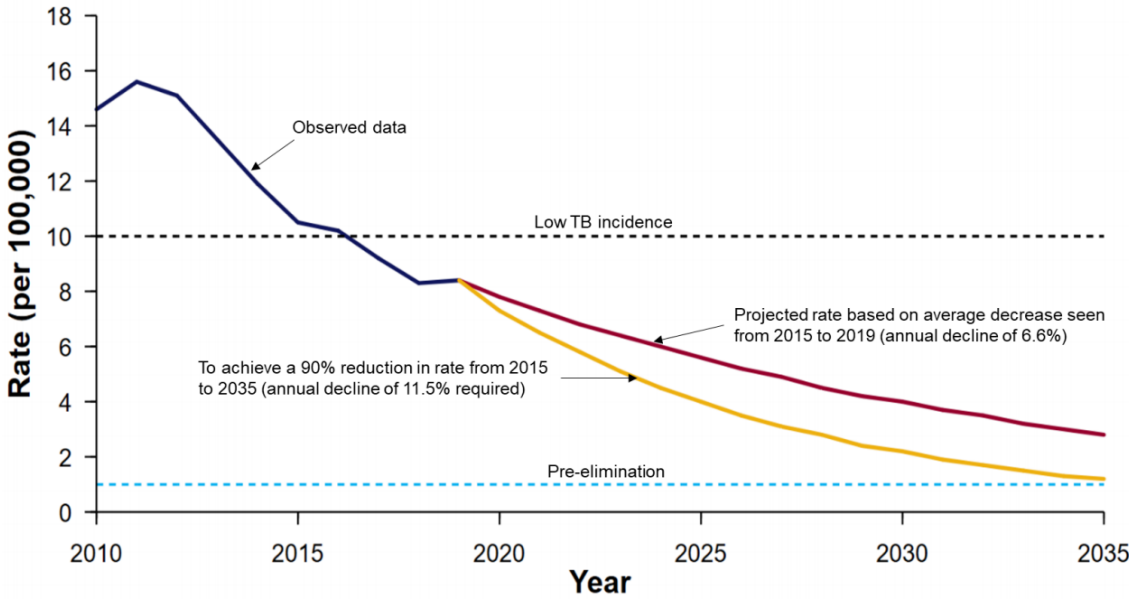
Reproduced from Global TB Report, World Health Organization 2020[21]. Licence: CC BY-NC-SA 3.0 IGO.



In England, TB incidence rates fell by an average annual decline of 6.6% from 2010 to 2019, but increased by 2.4% in 2019 compared to 2018[19]. Despite this progress, the average annual decrease is below the level required (11.5%) to achieve the WHO End TB target (Figure 1-10).

Figure 1-10: Observed and projected trends in TB notifications (2010-2035).

Reproduced from Public Health England Tuberculosis in England 2020 report under Open Government Licence v3.0[19].



1.2.3 The rationale for TB prevention

Early detection and prevention of TB disease reduces TB-associated morbidity and mortality. Since *M. tuberculosis* is an obligate pathogen, prevention of disease may also interrupt onward transmission. Early studies in Alaska demonstrated a dramatic reduction in TB incidence achieved through a combination of early case detection with TB treatment initiation, along with community-wide preventative treatment (Chapter 1.7.2)[23]. Scale-up of preventative treatment for people at high-risk of TB is therefore a key component of the WHO End TB Strategy[21,22], and was also a key target set in the UN high-level meeting on TB in 2018[24].

Alongside preventative treatment to people at high risk of disease, a range of other interventions are central to reducing the burden of TB disease, as reflected in WHO policy[21,22]. Following its discovery in the early 20th century, Bacillus Calmette–Guérin (BCG) vaccine was first evaluated in a placebo-controlled trial in the 1930s[25]. It was subsequently shown to have consistently high efficacy in protecting against childhood TB meningitis and miliary TB[26] and is therefore included in childhood vaccination programmes in many countries[27]. However, BCG efficacy for prevention of pulmonary TB has been heterogeneous. Meta-analyses have suggested that higher efficacy may be observed among people

who are not already infected with *M. tuberculosis* or other non-tuberculous mycobacteria[28], and that this may be partly explained by BCG protecting against initial *M. tuberculosis* infection as well as disease[29]. These data have suggested that BCG provides an important, but limited, contribution towards TB control, meaning that novel vaccines are required. Recent trials in high transmission settings have shown promise. BCG revaccination and a novel vaccine candidate (H4:IC31) among people who did not have evidence of LTBI at baseline (defined as negative interferon gamma release assay (IGRA); Chapter 1.4) reduced the risk of sustained IGRA positivity during follow-up[30], while the M72/AS01E candidate vaccine provided 49.7% (95% confidence interval (CI) 2.1-74.2) protection against TB disease over three years[31].

In addition to vaccines, infection control interventions to reduce transmission in healthcare facilities, congregate settings and households[32], along with adequate prevention and management of comorbidities associated with increased TB risk (including HIV, chronic lung disease, diabetes and alcohol dependence)[33] are integral to TB prevention efforts globally.

1.3 Diagnosis of TB disease

Microbiological investigations for TB disease represent the cornerstone for confirmatory diagnosis, supported by radiology. Microbiological sampling is dependent upon the site of disease, with respiratory sampling for suspected pulmonary TB done most frequently. Respiratory samples include spontaneously expectorated sputa, induced sputa (for example following inhalation of hypertonic nebulised saline), or samples acquired invasively via bronchoscopy. All microbiological diagnostics are highly dependent upon adequate sampling. This may be particularly challenging among children, people with suspected extra-pulmonary TB, and people with respiratory symptoms who cannot produce good quality sputum samples spontaneously, since access to sputum induction facilities is often limited, while bronchoscopy is usually confined to high- and middle-income settings[34,35].

1.3.1 Microbiology

1.3.1.1 Microscopy

Smear microscopy for acid-fast bacilli has been used for diagnosis of TB since the 19th century, when *M. tuberculosis* was first visualised by Robert Koch. It is limited by poor sensitivity, and limited specificity due to an inability to differentiate *M. tuberculosis* from other non-tuberculous mycobacteria. Despite these issues, microscopy is still in widespread use today – often as a surrogate measure of bacillary load to infer infectiousness in high-income settings, and also as a first line diagnostic in many low- and middle-income countries.

1.3.1.2 Culture

Mycobacterial culture represents the gold-standard TB diagnostic. Culture facilitates mycobacterial speciation along with drug susceptibility testing. Liquid culture has gradually replaced solid culture techniques in view of greater yield and faster results[36]. However, culture is limited by the requirement for specialist laboratory facilities and expertise, and is slow (with final results usually taking up to six weeks). These issues mean that global access to mycobacterial culture is limited, while the long interval from sampling to results can also risk losses to follow-up that undermine the cascade of TB care.

1.3.1.3 Molecular diagnostics

Molecular diagnostics, in the form of real-time polymerase chain reaction (PCR) assays, are now recommended as a first-line diagnostic test for adults and children under investigation for pulmonary and extra-pulmonary TB. The Xpert MTB/RIF assay (Cepheid, Sunnyvale, USA), which is a semi-automated cartridge-based system on the GeneXpert platform that detects *M. tuberculosis* and rifampicin-resistant mutations in the *rpoB* gene and provides results in under two hours, was first endorsed by the WHO in 2010[37]. An updated cartridge, called Xpert MTB/RIF Ultra (Xpert Ultra), was endorsed by the WHO in 2017[38]. This newer assay has a lower limit of detection than Xpert MTB/RIF, estimated as 15.6 colony-forming units (CFU) per mL of sputum compared to 112.6 CFU/mL for Xpert MTB/RIF[39].

The diagnostic accuracy of Xpert MTB/RIF and Xpert Ultra for pulmonary and extra-pulmonary TB among adults and children, compiled from Cochrane systematic reviews and meta-analyses, are shown in Table 1-1. Accuracy for extra-pulmonary TB is heterogeneous, dependent on sample type. While Xpert Ultra generally has higher sensitivity due to its lower limit of detection, this comes at the cost of a small reduction in specificity. Lower specificity has led to diagnostic uncertainty with Xpert Ultra, particularly relating to "trace" positive results, which are often considered to be false-positives and are particularly common in people with a history of previous TB, and in settings with high transmission[40].

Table 1-1: Diagnostic accuracy of Xpert MTB/RIF and Xpert MTB/RIF Ultra for diagnosis of TB.

Adapted from [41–43]. Compared to reference standard of culture unless stated otherwise. Presented as point estimate (95% CI).

	Xpert MTB/RIF			**Xpert MTB/RIF Ultra**		
	Studies	*Sensitivity*	*Specificity*	*Studies*	*Sensitivity*	*Specificity*
Adults						
Pulmonary	70	85% (82-88)	98% (97-98)	1	88% (85-91)	96% (94-97)
Cerebrospinal fluid	30	71.1% (62.8-79.1)	96.9% (95.4-98.0)	6	89.4% (79.1-95.6)	91.2% (83.2-95.7)
Pleural fluid	25	49.5% (39.8-59.9)	98.9% (97.6-99.7)	4	75.0% (58.0-86.4)	87.0% (63.1-97.9)
Lymph node aspirate*	4	81.6% (61.9-93.3)	96.4% (91.3-98.6)	1	70% (51-85)	100% (92-100)
Children						
Pulmonary (sputum)	23	64.6% (55.3-72.9)	99.0% (98.1-99.5)	3	72.8% (64.7-79.6)	97.5% (95.8-98.5)
Pulmonary (nasopharyngeal aspirate, stool)	14	45.7-73.0%	98.1-99.6%	1	45.7% (28.9-63.3)	97.5% (93.7-99.3)
Cerebrospinal fluid	6	54.0% (27.8-78.2)	93.8% (84.5-97.6)			
Lymph node aspirate or biopsy	6	90.4% (55.7-98.6)	89.8% (71.5-96.8)			

**Indicates composite reference standard.*

More recently, the Truenat platform (Molbio, Goa, India) has been endorsed by the WHO as an alternative to Xpert. Truenat uses a chip-based platform for *M. tuberculosis* detection and is widely implemented in India[44].

While molecular diagnostics were initially heralded as "game-changers" in TB[45], results of impact assessments on clinical outcomes have been disappointing, with no associated reduction in mortality in multiple randomised-controlled trials (RCTs), even when pooled in an individual participant data meta-analysis (IPD-MA) to improve power[46]. These findings may highlight a need for integration of rapid diagnostics into effective care cascades and wider health systems in order to achieve population-level impact.

1.3.1.4 Antigen-based diagnostics

Tests that detect urine mycobacterial lipoarabinomannan (LAM) antigen have emerged as point-of-care diagnostics for TB disease. The Alere Determine TB LAM test (AlereLAM; Alere, USA) was first endorsed for use in PLHIV by the WHO in 2015[47]. Guidance was updated in 2019 to now recommend using AlereLAM to assist in the diagnosis of active TB in HIV-positive adults, adolescents and children meeting any of the following criteria: signs and symptoms of pulmonary or extra-pulmonary TB; advanced HIV disease or seriously ill; or CD4 cell count <200 cells/mm^3 (<100 cells/mm^3 among outpatients) irrespective of signs and symptoms of TB[48]. These recommendations are based upon limited overall sensitivity among PLHIV, albeit higher among participants with CD4 <100 cells/mm^3 (Table 1-2). Two RCTs have evaluated the impact of AlereLAM testing on mortality among inpatients with advanced HIV disease, with pooled risk ratio for mortality 0.85 (0.76-0.94)[48–50].

A new urine-based LAM test has since been developed – the Fujifilm SILVAMP TB LAM (FujiLAM; Fujifilm, Tokyo, Japan). This showed promising performance using biobanked frozen urine samples among five cohorts of inpatient and outpatient adults with HIV, with better sensitivity than AlereLAM, when using a microbiological reference standard (Table 1-2)[51]. FujiLAM also shows promise among HIV-uninfected adult outpatients with symptoms suggestive of pulmonary TB (Table 1-2). FujiLAM is not yet endorsed by the WHO, pending additional validation and impact assessments. However, a major limitation of using LAM-

based diagnostics as confirmatory tests is that they do not provide information on *M. tuberculosis* drug susceptibilities.

Table 1-2: Diagnostic accuracy of urine LAM antigen tests among unselected participants.

Adapted from [48,51,52]. Compared to culture reference standard unless stated otherwise. Presented as point estimate (95% CI). PLHIV = people living with HIV.

	Studies	**Sensitivity**	**Specificity**
AlereLAM			
PLHIV (overall)	7	35% (22-50)	95% (89-98)
PLHIV (inpatients)	3	62% (41-83)	84% (48-96)
PLHIV (outpatients)			
PLHIV (CD4 ≤100 cells/mm^3)	3	47% (40-64)	90% (77-96)
FujiLAM			
PLHIV (inpatients and outpatients)	5	71% (59-81) vs. 35% (20-51) for AlereLAM	91% (87-94) vs. 95% (92-98) for AlereLAM
HIV-uninfected adult outpatients with presumptive pulmonary TB	1	53% (44-62) vs. 11% (6-18) for AlereLAM	99% (97-99.6) vs. 92% (89-95) for AlereLAM

1.3.1.5 Drug susceptibility testing

Phenotypic drug susceptibility testing is usually considered to be the gold-standard method of determining appropriate treatment regimens, but is dependent upon first culturing *M. tuberculosis*. Molecular approaches are being used increasingly, with the Xpert MTB/RIF system enabling rapid detection of rifampicin resistance in the first-line cartridge[43], and fluoroquinolones, aminoglycosides, and isoniazid in a recently developed second-line cartridge[53]. Line-probe assays are also commonly used to detect common resistance-conferring mutations. More recently, the scale-up of *M. tuberculosis* whole genome sequencing, launched for all positive *M. tuberculosis* cultures in England in 2017[54], alongside detailed curations of phenotypic-genotypic correlates[55], has led to genotypic drug susceptibility testing being used increasingly. However, despite these advances, access to phenotypic or genotypic drug susceptibility testing is limited in many low- and middle-income settings, where the TB burden

is often highest[21]. Multi-drug resistant (MDR) TB refers to resistance to both rifampicin and isoniazid, while extensively-drug resistant TB refers to additional resistance to fluoroquinolones and at least one injectable agent (amikacin, kanamycin, or capreomycin)[56].

1.3.2 Radiology

Chest radiography has historically played a central role in diagnosis and in mass population screening for pulmonary TB[57]. Typical parenchymal abnormalities include infiltrates, cavitation and fibrotic lesions, with a preponderance for the upper lobes. These parenchymal changes are often accompanied by evidence of mediastinal lymphadenopathy. Chest radiography has been recommended by the WHO to facilitate TB control through a range of applications[58]. These include supporting TB diagnostic evaluation, triaging risk of disease, and as a screening test during active case finding[58]. However, reliance on film-based radiographs and a shortage of trained chest radiograph readers have presented barriers to widespread use[59]. Increasing digital x-ray services with computer-assisted interpretation[60] may improve access to chest x-ray globally, facilitating increased use for active and passive case-finding. Despite these advances, chest radiography lacks specificity for TB and, in the absence of longitudinal imaging, cannot discriminate current TB from residual scarring caused by previous disease.

Cross-sectional imaging using computed tomography (CT) is also frequently used to support TB diagnosis in high-resource settings, while the addition of nuclear imaging using positron emission tomography (PET-CT) may improve sensitivity through detection of focal metabolic activity, represented by radionucleotide tracer uptake[61].

1.4 Latent TB diagnostics

There is no validated diagnostic test available to detect viable *M. tuberculosis* bacilli in the absence of disease. Instead, current LTBI diagnostics detect immunosensitisation to *M. tuberculosis* in the form of a cell-mediated T cell response. Consequently, the pragmatic clinical definition of LTBI is based upon (1) evidence of immunosensitisation; and (2) an absence of concurrent

disease[7,62]. Measurement of T cell responses to *M. tuberculosis*, however, cannot distinguish between LTBI, current or cured TB disease.

1.4.1 Tuberculin skin test

The tuberculin skin test (TST), involves intra-dermal injection of purified protein derivative (PPD) using the Mantoux technique, first used in the early 20th century. Following administration, the diameter of induration is read after a 48-72 hour interval. The test is relatively cheap, widely available and does not require specialist expertise or laboratory facilities. However, the TST requires two patient encounters (for administering and induration measurement), it lacks specificity as it can be affected by BCG vaccination and non-tuberculous mycobacteria exposure, and also has imperfect sensitivity, which may be partly explained by anergy in the context of current disease. TSTs provide quantitative values, which can then be categorised as binary results. A range of thresholds have been proposed, ranging from 5-15mm, with lower cut-offs often recommended for young children and immunocompromised people, among whom T cell responses may be attenuated. Hereafter, I will refer to TST cut-offs as 5mm (TST_5), 10mm (TST_{10}), 15mm (TST_{15}) and a BCG-stratified TST cut-off (5mm among unvaccinated; 15mm among vaccinated; $TST_{5/15}$). Sources of technical variation in TST may include boosting phenomena and inter- and intra-individual variation in injection placement and induration measurement[63].

1.4.2 Interferon gamma release assays

Interferon gamma release assays (IGRAs) have been commercially available since 2002. Similarly to the TST, they detect a T cell response to *M. tuberculosis*. They achieve this through in-vitro stimulation of blood using antigens from the region of difference 1 (RD1) area of the genome, namely early secretory antigenic target 6 (ESAT-6) and culture filtrate protein 10 (CFP-10). The most widely available commercial IGRAs are the QuantiFERON assays (Qiagen, Hilden, Germany), which use an enzyme-linked immunosorbent assay (ELISA), and T-SPOT.TB (Oxford Immunotec, Oxford, UK), using an enzyme-linked immunosorbent spot (ELISPOT) approach. For both QuantiFERON and T-SPOT.TB assays, mitogen and negative controls are run in parallel to TB antigenic stimulation. Quantitative results are obtained by subtracting negative

control responses from TB antigen responses, and can then be classified into binary results. Manufacturer-recommended cut-offs for QuantiFERON and T-SPOT.TB assays are shown in Table 1-3[64,65]. QuantiFERON assays have been regularly updated over time and including the QuantiFERON TB Gold, QuantiFERON TB Gold-in-tube (QFT-GIT) and QuantiFERON TB Gold Plus (QFT-Plus).

Table 1-3: Cut-offs used for QuantiFERON and T-SPOT.TB.

Quantitative values are calculated by subtracting negative control from maximal TB antigen response, assuming the test is valid. Based on manufacturer recommendations[64,65].

Test	Negative	Positive
QuantiFERON (IU/mL)	<0.35	≥0.35
T-SPOT.TB (spots)	<5	≥6

IGRAs have the advantage of being more specific for *M. tuberculosis* than TST as they are not affected by prior BCG vaccination or exposure to most non-tuberculous mycobacteria (with the exceptions of *M. marinum* and *M. kansasii*)[66]. However, IGRAs require specialist laboratory facilities and are relatively expensive, meaning that they are not routinely available in most low- and middle-income countries.

IGRAs are also subject to numerous sources of technical variability including: manufacturing differences between production lots; timing, volume and technique of blood collection; timing and duration of antigen stimulation; ELISA/ELISPOT imprecision; and immunological boosting phenomena (including due to prior TST)[66]. These sources of measurement error may contribute to the significant proportion of individuals who convert (transition from negative to positive results) or revert (positive to negative) during longitudinal sampling of healthcare workers[67]. The high frequency of dynamic fluctuations in IGRA values has led to proposals for a more rigorous definition for conversion, by inclusion of a 'borderline' zone in quantitative values, as opposed to using standard manufacturer cut-offs[68].

1.4.3 Diagnostic accuracy of TST and IGRAs

In the absence of a gold-standard, evaluation of the diagnostic accuracy of LTBI diagnostics is challenging. Sensitivity is frequently estimated among people with microbiologically-confirmed TB disease, while specificity for LTBI is estimated separately among people at low risk of exposure *M. tuberculosis* who reside in countries with low transmission. Using this approach, previous meta-analyses estimated QuantiFERON sensitivity (QFT Gold and QFT-GIT) as 76-80% and T-SPOT.TB sensitivity as 88-90%[69,70]. Specificity for QuantiFERON and T-SPOT.TB were estimated as 98% and 93%, respectively[69]. TST performance appeared more heterogeneous across studies, with pooled sensitivity 65-77% and specificity 97% among non-BCG vaccinated participants, but this was much lower and more heterogeneous among vaccinated participants[69,70]. However, interpretation of these performance characteristics is impaired by heterogeneity in TST cut-offs across studies, and by a lack of direct head-to-head assessments comparing performance of each index test among the same participants. The latter is an important consideration since population differences may contribute to differences in test performance across studies.

Test specificity can also be evaluated among people under investigation for TB disease, in order to assess whether TST and IGRAs could be part of the diagnostic evaluation for suspected TB. A previous systematic review and meta-analysis of data from low- and middle-income countries found a sensitivity of 83% (95% CI 63-94) and specificity of 61% (40-79) for T-SPOT.TB (6 studies) and a sensitivity of 69% (52-83) and specificity of 52% (41-63) for QFT-GIT (8 studies)[71].

A recent prospective observational cohort study further compared the diagnostic accuracy of QFT-GIT and T-SPOT.TB among symptomatic people presenting to TB services in the UK[72]. In this study, outcome measures were robust and pre-specified and the cohort design ensured a representative sample. Sensitivity and specificity of QFT-GIT and T-SPOT.TB for culture-confirmed and highly probable TB cases, were 67.3% (62.0-72.1) and 81.4% (76.6-85.3), and 80.4% (76.1-84.1) and 86.2% (82.3-89.4), respectively. A similar study design in China demonstrated similar findings with sensitivities of T-SPOT.TB and QFT-GIT of 85.2% and 84.8%, and specificities of 63.4% and 60.5%, respectively[73].

Taken together, these data consistently demonstrate imperfect sensitivity and moderate specificity among people with suspected TB, supporting the current recommendation that IGRAs should not be routinely used to support TB diagnostic evaluation[74]. Imperfect sensitivity may be partly explained by anergy in the context of concurrent disease[63,66]. Recent data have also detected humoral and interferon-gamma-independent T cell responses among people with high exposure to *M. tuberculosis* who are persistently TST and IGRA negative[75], though the diagnostic and prognostic value of these responses remains untested.

1.4.4 Prognostic ability of TST and IGRAs

The primary objective of LTBI testing is to identify people who may be at risk of future TB disease. Therefore, the prognostic ability of LTBI tests for incident TB is fundamental to their implementation in LTBI screening. Three systematic reviews and meta-analyses conducted over the last decade have sought to examine the prognostic ability of TST and/or IGRA across all populations and settings[76–78], and are summarised in Table 1-4.

Table 1-4: Prognostic performance of IGRAs and TST from previous systematic reviews and meta-analyses.

Adapted from[76–78]. Data are included where available from the systematic review and meta-analysis reports. Data shown as point estimates (95% Cis). Precision of estimates presented as originally reported.

	Campbell (2020)[76]			**Zhou (2020)**[77]		**Rangaka (2014)**[78]
	General population	*Case contacts*	*PLHIV*	*All*	*Head-to-head*	*All*
IGRA						
Studies		19	9	40	12	9
*Incidence rate (+)^*		17.0 (12.9-22.4)	16.9 (10.5-27.3)			Range 4-48
*Incidence rate (-)^*						Range 2-24
Incidence rate ratio		10.8 (6.1-19.0)	11.0 (4.6-26.2)			2.11 (1.29-3.46)
*Positive predictive value**		4.60%	5.10%	4.5% (3.3-5.8)	4.2% (2.5-6.3)	
Negative predictive value				99.7% (99.5-99.8)	99.4% (98.8-99.8)	
Risk ratio				9.35 (6.48-13.49)	7.12 (3.39-14.94)	2.22 (1.54-3.19)
TST						
Studies	3	27	9	40	12	5
Cut-off	10mm	10mm	5mm	10mm	10mm	10mm
*Incidence rate (+)^*	0.3 (0.1-1.1)	9.4 (6.3-14.1)	27.1 (15.0-49.0)			
*Incidence rate (-)^*						
Incidence rate ratio		4.1 (2.6-6.4)	11.1 (6.2-19.9)			1.60 (0.94-2.72)
Positive predictive value	0.2%	2.60%	7.10%	2.9% (2.1-3.7)	1.8% (0.8-3.1)	
Negative predictive value				99.2% (98.9-99.4)	99.2% (98.6-99.6)	
Risk ratio				4.28 (3.29-5.56)	4.30 (2.03-9.10)	

^Incidence rates shown per 1,000 person-years.

Rangaka et al evaluated the prognostic value of IGRA and reported heterogeneous incidence rates of 4-48/1,000 person years among people with positive tests[78]. In five studies classified as low risk of incorporation bias (where the test result has not contributed towards the outcome diagnosis), incidence rate ratios were 2.11 (1.29-3.46) for IGRA and 1.60 (0.94-2.72) for TST[78]. However, the number of studies included was low and heterogeneity in incidence rates precluded calculation of pooled estimates.

Campbell et al conducted a recent systematic review and meta-analysis to estimate the risk of TB among untreated people with positive TST or IGRAs, including 122 original studies[76]. They estimated pooled random-effects TB incidence rates of 0.3/1,000 person years (0.1-1.1; I^2 96%) among people in the general population with positive TST_{10}, 17.0/1,000 person-years (12.9-22.4; I^2 81%) among recent TB contacts with positive IGRA and 8.4/1,000 person-years (5.6-12.6; I^2 96%) among recent TB contacts with a positive TST_5. Elevated TB incidence rates were also evident among PLHIV, immigrants, people with silicosis or requiring dialysis, transplant recipients, and prisoners, when comparing those with positive to negative TST and IGRA test results using incidence rate ratios (IRRs). However, comparisons between TST and IGRA performance in this meta-analysis are precluded by a lack of direct head-to-head assessments of prognostic value among the same study participants. Moreover, there was evidence of marked heterogeneity in TB incidence rates across studies within each of the risk groups evaluated; the pooled estimates should therefore be interpreted with caution.

Zhou et al conducted a systematic review and meta-analysis of 40 studies to compare the prognostic ability of TST and IGRA[77]. In the primary analyses, which were not restricted to head-to-head analyses, pooled IGRA positive and negative predictive values were estimated as 4.5% (3.3-5.8) and 99.7% (99.5-99.8), with a risk ratio of 9.35 (6.48-13.49). For TST_{10}, positive and negative predictive values were estimated as 2.9% (2.1-3.7) and 99.2% (98.9-99.4), with a risk ratio of 4.28 (3.29-5.56). The authors concluded that IGRA had greater prognostic ability than TST. However, this meta-analysis had a number of important limitations. First, a sub-analysis restricted to head-to-head studies showed similar risk ratios of 7.12 (3.39-14.94) and 4.30 (2.03-9.10) for IGRA and TST_{10}, respectively. Person-time was also not accounted for with no IRRs estimated.

Moreover, a WHO systematic review and meta-analysis of five studies from settings with high TB incidence (defined as annual incidence ≥100/100,000) showed similar predictive ability for TST and IGRA, leading to the WHO recommendation that either TST or IGRA may be used for LTBI screening[62,79]. These five studies were not included in the meta-analysis by Zhou et al[79].

In addition, the UK Prognostic Evaluation of Diagnostic IGRAs Consortium (UK PREDICT) study, which represents the largest prospective study to date directly comparing TST and IGRA prognostic ability, was not included in the head-to-head sub-analysis by Zhou et al[77,80]. This prospective cohort study recruited 9,610 adults in the UK who were recent TB contacts and migrants from high TB burden countries. In the pre-specified head-to-head analysis among 6,380 participants with valid results for TST, QFT-GIT and T-SPOT, the three tests had similar prognostic ability when using a $TST_{5/15}$. TB IRRs among test positive compared to test negative participants ranged from 5.4-8.8, TB incidence rates among participants with positive results were 10.1-11.1/1,000 person-years, and positive predictive values (during all available follow-up) ranged from 3.3-4.2% (Table 1-5). Among participants with negative test results, incidence rates were low for all three tests (1.5-1.9/1,000 person-years), and negative predictive values were consistently high (99.4-99.5%).

Table 1-5: Prognostic performance of IGRAs and TST in UK PREDICT study.

Adapted from Abubakar et al[80]. Incidence rates shown per 1,000 person-years. Data presented as point estimates (95% CI).

	QFT-GIT	**T-SPOT.TB**	**TST_5**	**TST_{10}**	**$TST_{5/15}$**
Incidence rate (+)	10.1 (7.4-13.4)	13.2 (9.9-17.4)	6.8 (5.2-8.7)	8.5 (6.5-11.0)	11.1 (8.3-14.6)
Incidence rate (-)	1.9 (1.3-2.7)	1.5 (1.0-2.2)	1.2 (0.6-2.0)	1.4 (0.8-2.2)	1.6 (1.0-2.3)
Incidence rate ratio	5.4 (3.4-8.5)	8.8 (5.5-14.2)	5.8 (3.2-10.6)	6.2 (3.7-10.3)	7.1 (4.4-11.4)
Positive predictive value	3.3%	4.2%	2.2%	2.7%	3.5%
Negative predictive value	99.4%	99.5%	99.6%	99.6%	99.5%

In summary, while each of the above meta-analyses has limitations as discussed, the evidence collectively suggests three key messages. First, prognostic ability appears similar for QFT-GIT, T-SPOT.TB and TST, particularly when using TST_{15}. As a result, either TST or IGRAs are recommended in global, European and UK LTBI guidance[74,81,82]. Second, for all three tests, incidence rates and positive predictive values among people with positive results are relatively low, with incidence rates consistently <50/1,000 person-years and positive predictive values <5% across all studies. Third, there was significant heterogeneity in prognostic value across studies, suggesting that pooled estimates may be of limited value in clinical practice, since the risk of progression to incident TB among individuals with LTBI is likely to vary markedly between individuals.

1.4.5 Quantitative IGRAs

It has long been considered that stronger TST responses are more likely to represent true sensitisation to *M. tuberculosis*, as opposed to history of BCG vaccination or sensitisation to non-tuberculous mycobaceria[83]. This has led to interest in whether quantitative LTBI test results may be of prognostic value for incident TB. Recent data from studies in both adults and infants in low and high TB incidence settings have demonstrated that higher quantitative IGRA results are associated with increased risk of incident TB, thus raising hope that implementing higher cut-offs may improve prognostic value. Moreover, since current diagnostic thresholds for scoring a positive test are based on detecting sensitisation to *M. tuberculosis* rather than development of incident TB disease, optimising these thresholds might facilitate better implementation of existing LTBI diagnostics, while novel biomarkers with improved predictive value are awaited. However, these previous evaluations of quantitative IGRA results have been limited to the QFT-GIT assay only[84–88]. It remains unclear whether implementation of a higher threshold for positivity may actually be of use programmatically to improve their prognostic ability.

1.4.6 Newer generation skin tests and interferon-gamma release-assays

Next generation skin tests incorporating ESAT-6 and CFP-10 as *M. tuberculosis*-specific antigens are in development. One such candidate is C-Tb (Statens Serum Institute, Copenhagen, Denmark), which has demonstrated strong

concordance with QFT-GIT, without being affected by prior BCG vaccination[89]. However, there are no data to demonstrate superiority of C-Tb to TST when using higher cut-offs (i.e. TST_{15} or $TST_{5/15}$), and no studies have examined the prognostic value of C-Tb for incident TB.

A newer generation QuantiFERON (QFT-Plus) was also recently launched, adding a second TB antigen tube (TB2) that incorporates short peptides designed to stimulate a $CD8^+$ T cell response, in addition to the $CD4^+$-response tube (TB1) included in previous versions. The proposed rationale for this is that $CD8^+$-responses have been associated with mycobacterial load and recent TB exposure[90,91]. Initial evaluations have suggested QFT-Plus may have improved test sensitivity in active TB compared to QFT-GIT[92], and that the $CD8^+$-targeted antigen tube response may be associated with proxy measures of degree of TB exposure among contacts[93]. However, no studies have examined the prognostic value of QFT-Plus for predicting incident TB.

Next-generation IGRAs, with addition of newer antigens, are also under development in an attempt to improve diagnostic and prognostic performance[72].

1.4.7 Reversion following clearance of M. tuberculosis infection

As TST and IGRAs detect T cell responses to *M. tuberculosis*, rather than infection itself, it is widely recognised that positive tests frequently persist long after spontaneous or antimicrobial treatment-induced clearance of infection. This phenomenon was observed in early trials following effective treatment with preventative therapy regimens[14], and is reinforced by more recent evidence demonstrating that IGRAs are not useful for monitoring response to therapy among people with TB disease[94]. It has also been suggested that test reversion may be more likely following clearance if *M. tuberculosis* was recently acquired (in the preceding year)[95–97].

This important limitation of TST and IGRAs has major implications when considering the global burden of *M. tuberculosis* infection, recently estimated as 1.7 billion people, or 23% of the global population[98]. Such mathematical modelling estimates are heavily based upon LTBI prevalence surveys, using TSTs. Thus, the 23% estimate likely reflects immunoreactivity to *M. tuberculosis*, while the true underlying proportion who harbour viable bacilli remains unknown.

Persistent positivity of TST and IGRAs following clearance of *M. tuberculosis* infection also means that evidence of test conversion from negative to positive is likely to provide stronger evidence of recent infection. Recent conversion has therefore been associated with elevated risk of incident TB, compared to positive results from cross-sectional sampling only[99,100].

1.5 WHO target product profiles for novel TB diagnostics

The need for novel diagnostic tests to improve TB control led to the development of TPPs by the WHO. High-priority TPPs were first identified by surveying stakeholders; a Delphi-like process was then used to build consensus for each TPP, with characteristics with significant discordance discussed at a consensus meeting. The objective of the TPPs is to provide target specifications that product developers should aim to meet relating to both performance and operational characteristics. Initial high-priority diagnostics addressed by 2014 WHO TPPs[101] included a point-of-care non-sputum-based test capable of detecting all forms of TB (i.e. a confirmatory test), and a point-of-care test that can be used by first-contact health-care providers to identify those who need further testing (i.e. a triage test). The need to fulfil these TPPs has led to development of an interactive dashboard for active TB biomarkers[102], curated through systematic review[103].

More recently, a TPP and framework for evaluation for a test predicting progression from *M. tuberculosis* infection to TB disease was published, using similar methodology[11]. This consensus document differentiates “persistent infection tests” that are intended to detect *M. tuberculosis* infection, but revert to negative if the infection is spontaneously cleared, from “incipient TB tests” that are intended to detect people in whom progression to TB disease has commenced. Incipient TB test TPP specifications are shown in Table 1-6 and are intended to improve upon the prognostic ability of TST and IGRA. Minimum and optimum incipient TB test targets for development of active TB over a two year time horizon from testing are stated as sensitivity and specificity ≥75% and ≥90%, respectively.

Table 1-6: WHO TPP specifications for an incipient TB test.

Adapted from [11]. USD = US Dollars.

Characteristic	Optimal requirements	Minimal requirements
Intended use		
Goal of test / Intended use	Predicts risk of progression to active TB from TB infection within the next 2 years and provides a quantitative result that correlates with risk of progression. Result should decrease or revert to negative with treatment and thus allow an assessment of treatment success or cure and consequentially also reinfection.	Predicts risk of progression to active TB from TB infection within the next 2 years.
Type of specimen	Capillary whole blood (finger prick sample) / saliva / urine / stool / breath)	Whole blood by phlebotomy (or subpopulation of cells if simple processing included) / sputum
Performance characteristics		
Diagnostic sensitivity for progression to active TB	≥90% sensitivity	≥75% sensitivity
Diagnostic specificity for risk of progression to active TB	≥90% specificity	≥75% specificity
Operational characteristics		
Sample preparation	None or fully integrated	Allows for centrifugation/ incubation
Time to results	<24 hours	2.5 days
Pricing		
Cost of equipment	<500 USD	<5000 USD
Cost of consumables	<5 USD/test	10-100 USD/test

1.6 Incipient TB tests

A range of candidate biomarkers for incipient TB are currently under development and evaluation. These include transcriptomic, proteomic, metabolomic, microbiological and imaging measurements.

1.6.1 Whole blood transcriptomics

Since a seminal study describing whole blood transcriptomic perturbation during TB disease[104], there has been vast global interest in the pursuit of blood transcriptomic biomarkers for TB. In the original description by Berry et al, a 393 transcript signature for active TB, representing interferon signalling, was derived[104]. Over the subsequent decade, a multitude of blood transcriptomic biomarkers for TB have been described[105–112]. These biomarkers have become increasingly parsimonious, with reducing numbers of component transcripts, and single transcript biomarkers have now been identified[109,112]. Transcriptomic changes have also been shown to correlate with disease severity (as measured by PET-CT)[113–115] and to resolve during TB therapy[104,113,115,116]. Proposed applications of blood transcriptomic biomarkers have therefore included triaging risk of TB disease, supporting TB diagnosis, and monitoring response to treatment.

More recently, a landmark nested case-control study among a longitudinal cohort study among adolescents in South Africa (the “Adolescent Cohort Study”[117]) demonstrated that transcriptomic perturbation predates the clinical diagnosis of TB[118]. A 16-gene transcriptomic signature was derived that predicted progression to TB with area under the receiver operating characteristic (AUROC) curve 0.74 (95% CI 0.73-0.76) in the discovery set, giving sensitivity and specificity of 58.4% (56.1-60.7) and 80.0% (78.6-81.4), respectively. Performance of the signature was time-dependent, with greater discrimination for shorter time intervals from sampling. In an external validation cohort of recent TB case contacts in South Africa and the Gambia (Grand Challenges 6-74 study)[119], AUROCs were 0.72 (0.64-0.80) in the year following sampling, compared to 0.65 (0.53-0.76) 1-2 years after sampling, when measured using a real-time PCR assay. Three other studies have subsequently been published that demonstrate blood transcriptomic

perturbation predating TB diagnosis, raising hope that blood transcriptomic biomarkers may be promising candidates as incipient TB tests[119–121].

However, almost all early data describing discovery and validation of transcriptomic biomarkers were derived from case-control studies, which may be prone to spectrum bias due to inclusion of more extreme phenotypes, leading to overly optimistic discrimination metrics[122]. High quality independent validation studies of candidate biomarkers using robust outcome definitions among cohorts that are representative of target populations are therefore needed to better assess real-world performance.

The recently published Correlate of Risk Targeted Intervention Study (CORTIS) evaluated performance of a modified 11-gene version of the 16-gene signature described by Zak et al (RISK11) for predicting progression to TB, along with efficacy of signature-guided preventative therapy among HIV-uninfected adults in South Africa[123]. To do this, community-based participants were screened using RISK11, with a pre-specified biomarker cut-off. RISK11-positive participants were randomised to weekly open-label isoniazid and rifapentine for 12 weeks, versus no treatment. A subset of eligible RISK11-negative participants were randomly sampled and assigned to no treatment. Participants were screened at baseline for TB disease by testing two spontaneously produced sputum samples by Xpert MTB/RIF.

The study showed a cumulative probability of prevalent or incident tuberculosis disease of 6.6% (4.9-8.4) among untreated RISK11-positive participants and 1.8% in RISK11-negative participants, giving a risk ratio of 3.69 (2.25-6.05) for RISK11 over 15 months when using a pre-specified cut-off. Performance of RISK11 for diagnosis of baseline prevalent (84% of cases were asymptomatic) and incident TB is shown in Table 1-7. Even when using an optimised cut-off, RISK11 did not meet the WHO minimum TPP parameters for a triage test for prevalent TB or for an incipient TB test. Prognostic performance was highly time-dependent, with AUROC >0.80 for approximately 9 months, before falling to 0.58 between months 9 and 15. In addition, the preventative treatment regimen did not reduce TB incidence among RISK11+ participants over 15 months (treatment efficacy 7.0%, 95% CI −145-65).

The CORTIS trial represents a major contribution to the field. However, the lack of impact when used for general population screening means that optimal implementation of blood transcriptomic biomarkers for TB screening remains unclear. The lack of efficacy of preventative treatment to reduce TB incidence among RISK11+ participants also raises the question of whether preventative treatment regimens are sufficiently sterilising for people with incipient TB. Moreover, only one of the many candidate transcriptomic signature was evaluated in this study. A recent head-to-head evaluation of candidate signatures for TB disease was performed[124], but omitted promising signatures and compared diagnostic accuracy for incipient TB in only a single cohort over a 0-6 month time period only[125]. It is therefore unclear as to which of the many candidate transcriptomic signatures performs best for detection of incipient TB.

Table 1-7: Performance of RISK11 transcriptomic signature for prevalent and incident TB in the CORTIS trial.

Adapted from [123]. Data shown as point estimates (95% CI).

	Prevalent TB		**Incident TB (15 months)**	
	RISK11 (optimised cut-off)	*QFT-Plus*	*RISK11 (optimised cut-off)*	*QFT-Plus*
Risk ratio	7.39 (3.46-25.69)	4.43 (1.93-14.18)	2.67 (1.04-8.66)	2.83 (0.95-99.79)
Biomarker prevalence	25.8% (24.1-27.4)	63.4% (61.3-65.4)	25.3% (23.7-26.9)	63.2% (61.1-65.3)
AUROC	0.77 (0.68-0.86)	0.66 (0.58-0.73)	0.63 (0.47-0.80)	0.67 (0.54-0.79)
Sensitivity	72.1% (54.5-90.2)	88.7% (77.1-96.4)	47.5% (25.9-75.0)	83.2% (61.9-100.0)
Specificity	74.7% (73.1-76.4)	36.9% (34.9-39.0)	74.9% (73.2-76.5)	37.0% (34.9-39.1)
Positive predictive value	3.1% (2.0-4.3)	1.5% (1.0-2.2)	1.3% (0.6-2.1)	0.9% (0.5-1.4)
Negative predictive value	99.6% (99.2-99.9)	99.7% (99.3-99.9)	99.5% (99.1-99.9)	99.7% (99.2-100.0)

1.6.2 Proteomics and metabolomics

There has also been interest in development of proteomic and metabolomic biomarkers for incipient TB. A proteomic nested case-control analysis using

samples collected in the Adolescent Cohort Study led to discovery of 3-protein and 5-protein biomarkers[126]. During external validation in Gambian participants in the Grand Challenges 6-74 study, the 5-protein and 3-protein biomarkers had AUROCs of 0.66 (0.56-0.75) and 0.65 (0.55-0.75) within one year of TB diagnosis, and neither met the WHO TPP benchmarks for incipient TB tests. By comparison, C-reactive protein, a non-specific acute phase reactant marker, had AUROCs of 0.76 (0.69-0.83) and 0.62 (0.49-0.74) in the Adolescent Cohort Study discovery and Grand Challenges 6-74 validation cohorts, respectively.

A metabolomic signature, comprising 10 metabolites was also derived and validated in the Grand Challenges 6-74 study, again using a nested case-control approach[127]. Performance in the Grand Challenges 6-74 validation data showed an overall AUROC 0.67 (0.62-0.72), rising to 0.76 (0.68-0.84) in the 5 months following sampling. Once more, the WHO TPP parameters were not achieved over the longer time interval.

These proteomic and metabolomic analyses share commonality with the previously discussed transcriptomic data in demonstrating that the performance of incipient TB biomarkers is likely to be highly time-dependent, with better performance for shorter time horizons.

1.6.3 M. tuberculosis DNA detection

Early data have suggested that *M. tuberculosis* DNA can be identified from 2mL aliquots of peripheral blood mononuclear cells (PBMCs) from immunocompetent adults by PCR, following bacteriophage treatment using the Actiphage test (PBD Biotech Ltd, UK)[128]. In a recent report, 11/15 (73%) participants with microbiologically confirmed TB disease had positive Actiphage results, while all participants with non-TB respiratory illnesses (n=5) and healthy controls (n=28) had negative results. 3/18 recent TB contacts with positive IGRAs also had positive results, of whom two developed incident TB disease during follow-up.

A separate recent preliminary report described detection of *M. tuberculosis* DNA in large-volume PBMCs, particularly CD34+ haemopoietic stem cells, without bacteriophage treatment among recent TB contacts and PLHIV in Ethiopia[129]. *M. tuberculosis* DNA was detected using digital PCR in PBMC of 156/197 participants (79.2%) and was not associated with IGRA status. Among PLHIV

who received isoniazid preventative therapy 41/43 (95.3%) had detectable *M. tuberculosis* DNA at baseline, reducing to 23/43 (53.5%) following treatment.

Further validation of these preliminary findings is required in order to determine whether detection of *M. tuberculosis* DNA in blood may be a viable approach to facilitate diagnosis of true persistent infection and incipient disease and to monitor responses to preventative treatment.

1.6.4 Imaging

Chest radiography has been historically used to detect and define incipient TB. The National Association for the Study and Prevention of Tuberculosis defined incipient TB as:

> *"Slight or no constitutional symptoms (including particularly gastrointestinal disturbances, or rapid loss of weight). Slight or no elevation of temperature or acceleration of the pulse, at any time during the twenty four hours. Expectoration is usually small in amount or is absent. Tubercle bacilli may or may not be present. Slight infiltration limited to the apex of one or both lungs or a small part of one lobe. No tuberculosis complications."*
>
> - **From correspondence by John Ritter, 1916**[13]

Since radiographic fibrotic lesions are known to be associated with elevated risk of future TB disease[130], mass screening using chest x-ray could potentially play a role in incipient TB detection. However, chest radiographs lack specificity for TB and are also insensitive for early parenchymal changes. Through visualisation of foci of metabolic activity, PET-CT offers much higher sensitivity for detection of early infiltrates, nodular change and lymphadenopathy and may therefore be used to characterise incipient TB[131]. PET-magnetic resonance imaging (PET-MRI), which offers a lower dose of ionising radiation, has also been used to identify incipient changes among recent TB contacts[132]. However, both PET-CT and PET-MRI are highly expensive and are not scalable; their utility is therefore likely to be largely limited to research purposes, for example by providing radiological correlates to facilitate validation of other biomarkers that are more amenable to clinical translation.

1.7 Antimicrobial therapy for *M. tuberculosis*

1.7.1 Treatment for TB disease

For drug-susceptible pulmonary TB, quadruple antimicrobial therapy using rifampicin and isoniazid for six months, accompanied by ethambutol and pyrazinamide for an initial two-month 'intensive' phase has been the cornerstone of TB treatment since the 1980s[133,134]. This regimen continues to be recommended by the WHO and is highly effective when administered adequately[135].

Multiple trials have sought to establish shorter regimens, recently through use of fluoroquinolones, but many failed to achieve non-inferiority compared to standard six-month therapy[136–138]. A recent post-hoc individual participant meta-analysis of these trials showed that four-month regimens were non-inferior in participants with "minimal" disease, as defined by paucibacillary disease on sputum smear and absence of chest radiographic cavities, while suggesting that participants with high smear grades and cavitation may require prolonged treatment durations beyond six months[139]. These findings support a stratified approach to TB treatment for people with pulmonary TB. This is not a new concept, as studies from the 1980s in Hong Kong also suggested that pauci-bacillary (smear negative) disease might be adequately treated with four-month antimicrobial regimens[140]. Preliminary results from an RCT for drug-susceptible pulmonary TB suggest that eight weeks of daily treatment with high-dose rifapentine, isoniazid, pyrazinamide, and moxifloxacin followed by nine weeks of rifapentine, isoniazid, and moxifloxacin are non-inferior to standard six month treatment, though full results are awaited[141,142]. Another ongoing trial is assessing whether treatment could be reduced to as short as two months for non-severe pulmonary TB[143].

Longer treatment durations (usually 12 months) are indicated for patients with central nervous system involvement, while tailored therapies are required in the presence of resistance to first-line drugs[135]. WHO guidance updated in 2018 recommends oral therapy for drug-resistant TB, prioritising inclusion of fluoroquinolones (levofloxacin or moxifloxacin), bedaquiline and linezolid. Regimens are recommended to include at least four agents to which the *M. tuberculosis* isolate is likely to be susceptible in the first six months, and three

thereafter, for a total duration of 18-20 months. A shorter nine-month MDR-TB regimen can be given to selected patients - using moxifloxacin, clofazamine, pyrazinamide and ethambutol, with the addition of kanamycin, prothionamide, and high-dose isoniazid for an initial intensive phase[56]. Shorter, less toxic and all-oral regimens for drug-resistant TB are urgently required; a six-month regimen of bedaquiline, pretomanid, and linezolid has recently shown promise in an open-label, single arm study of 109 adults with multi- and extensively-drug resistant TB in South Africa, with 90% achieving favourable outcomes[144].

1.7.2 Preventative treatment

Preventative treatment for TB has been recognised as an important component of TB control since the 1950s (Figure 1-11). The efficacy of isoniazid as preventative treatment for TB was first established in a 1957 cluster RCT in the general community in Alaska, where TB incidence was approximately 1% annually. The trial demonstrated a >60% reduction in TB incidence with community-wide isoniazid for one year compared to placebo, which was maintained for at least five years (Figure 1-12)[23]. In parallel, an RCT in the United States of 12 months of isoniazid or placebo among 25,033 household TB contacts with positive TST_5 showed that isoniazid markedly reduced TB incidence in the first year after contact[14,145]. A later trial by the International Union Against Tuberculosis among 28,000 people with fibrotic pulmonary lesions showed that 24 weeks of isoniazid provided a 65% reduction in TB incidence compared to placebo. A 52-week course provided greater efficacy, but at the cost of greater hepatotoxicity[15]. Comstock subsequently used previous trial data to propose an optimal duration of isoniazid therapy of nine months[146].

Figure 1-11: History of TB preventative treatment.

Reproduced from Salazar-Austin et al (2019)[147] by permission of Oxford University Press. Seventy Years of Tuberculosis Prevention: Efficacy, Effectiveness, Toxicity, Durability, and Duration, American Journal of Epidemiology 188(12): 2078-2085. http://doi.org/10.1093/aje/kwz172

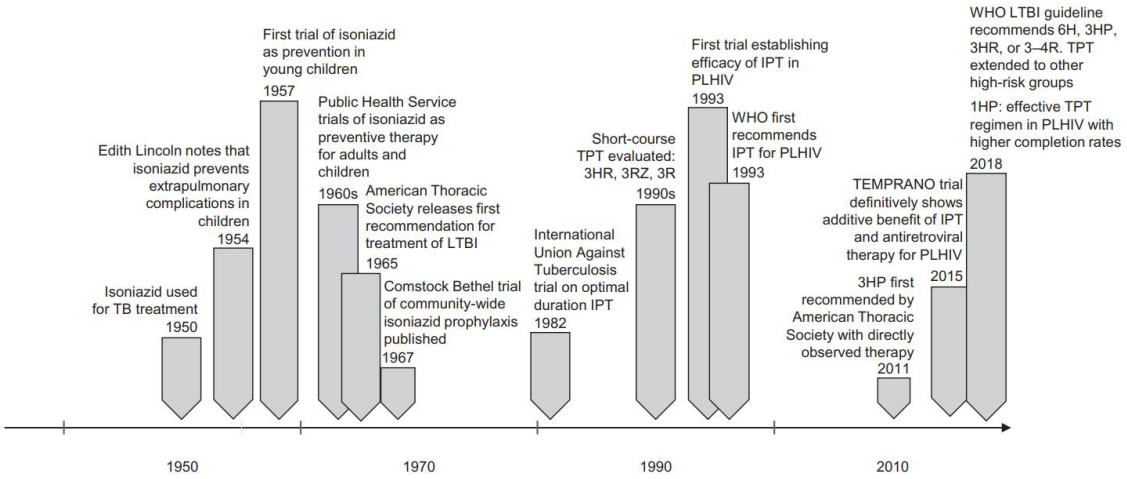
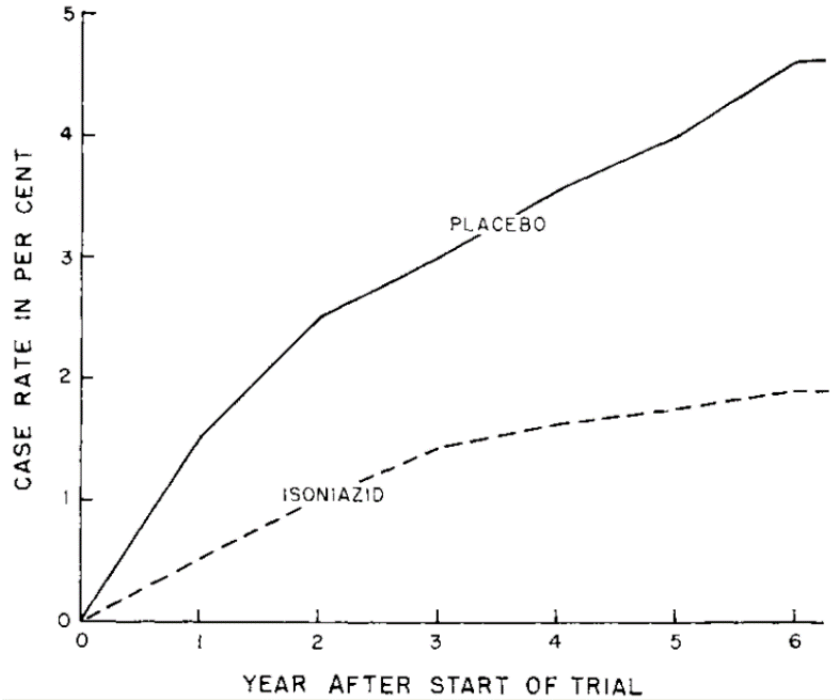


Figure 1-12: TB incidence in trial of community-wide isoniazid for one year compared to placebo in Alaska.

Reprinted from Comstock et al (1967). American Review of Respiratory Disease 95(6) with permission of the American Thoracic Society[a].



[a]*Copyright © 2021 American Thoracic Society. All rights reserved. The American Review of Respiratory Disease* is an official journal of the American Thoracic Society. *Now titled The American Journal of Respiratory and Critical Care Medicine*

Since these early trials, a range of alternative regimens have demonstrated equivalent efficacy to 6-9 months of isoniazid monotherapy and have been incorporated into WHO guidance[81]. The currently endorsed regimens for adults and children with LTBI are summarised in Table 1-8, along with the supporting evidence. Importantly, treatment completion has been consistently observed to be higher with shorter regimens[148–150].

Table 1-8: Summary of WHO-approved preventative treatment regimens for TB.

Regimen	Description	Efficacy data	References
6-9H	6-9 months of daily isoniazid	Superior to placebo	[15,146]
4R	4 months of daily rifampicin	Non-inferior to 9H among adults with LTBI	[151]
3HR	3 months of daily rifampicin and isoniazid	Network meta-analysis	[152]
3HP	3 months of weekly rifapentine and isoniazid	Non-inferior to 9H in HIV-uninfected people and PLHIV	[149,153]
1HP	1 month of daily rifapentine and isoniazid	Non-inferior to 9H among PLHIV	[150]

Of note, the 1HP regimen has only been evaluated among PLHIV aged ≥13 years (median follow-up 3.3 years); additional data are required among children and HIV-uninfected populations[150].

Among PLHIV, three RCTs from sub-Saharan Africa now collated in an IPD-MA have demonstrated that isoniazid offers incremental benefit over and above anti-retroviral therapy alone to prevent TB regardless of TST and IGRA results, with a hazard ratio 0.68 (95% CI 0.49-0.95)[154].

While early data from the Alaskan studies suggested good durability of protection by isoniazid preventative therapy[23], more recent data from settings with high transmission showed a rapid resurgence of incident TB risk following treatment completion among South African gold miners and PLHIV, respectively[155,156]. A meta-analysis of three RCTs done in countries with high TB incidence among PLHIV showed that continuous isoniazid reduces risk of TB more than six months' isoniazid (relative risk 0.62; 95% CI 0.42-0.89), and with greater efficacy among people with positive TST_5 (relative risk 0.51; 95% CI 0.30-0.86)[157]. This led to the

WHO recommending at least 36 months of daily isoniazid monotherapy in PLHIV in high TB transmission settings[81]. The differing durability of preventative treatment between the Alaskan and more recent studies is likely to be linked to ongoing levels of community transmission. This highlights a need to reduce transmission in parallel to preventative treatment administration in order to achieve durable effects in settings with high force of infection.

For people at risk of MDR-TB due to contact with an MDR index case, the WHO recommends consideration of preventative treatment (such as six months of levofloxacin daily), based on individualised risk assessment. There are currently no RCT data to support this, though a number of studies are ongoing[158–160].

1.7.3 Risks of preventative treatment

The most frequent serious complication of TB preventative treatment is hepatotoxicity, which can rarely be fatal[161]. Frequency of grade 3 or more hepatotoxicity (defined as alanine aminotransferase or aspartate aminotransferase 3-10 times the upper limit of normal with accompanying symptoms of nausea, anorexia, vomiting, fatigue or abdominal pain; or alanine aminotransferase or aspartate aminotransferase 5-10 times the upper limit of normal in the absence of symptoms) with 9H has been estimated as approximately 2.0%[148,149,151,162], though need for hospitalisation is rare (0.1%)[162]. Risk of peripheral neuropathy with isoniazid is reduced markedly by co-administration of vitamin B6 supplementation (pyridoxine), while rashes and grade 3-4 haematological abnormalities (defined as neutropenia <1 x 10^9 cells/L or platelets <50 x 10^9 cells/L) are rare (incidence <1%)[162]. Increasing age is associated with greater risk of isoniazid-associated hepatotoxicity[161], estimated as 5.5% for people aged 65-90 versus 1% for those aged 18-34 (adjusted odds ratio 5.3 (95% CI 1.9-13.3)[162].

In a pooled analysis of two RCTs comparing 4R vs 9H (3,205 individuals receiving isoniazid and 3,280 receiving rifampicin), adverse events (including hepatotoxicity) with the 4R regimen were not associated with age, and were less frequent compared to 9H (absolute risk difference −1.2%; 95% CI −1.9 to −0.5)[162]. When compared with 9H in another large RCT, the 3HP regimen has also been associated with lower risk of hepatotoxicity (0.3% vs 2.0%) but higher

risk of hypersensitivity reactions when using a broad definition (3.8 vs 0.5%), though there was no difference in overall risk of serious adverse events and no medication-associated deaths[149]. A network meta-analysis of RCTs found that 3HR regimen also had lower hepatotoxicity than isoniazid monotherapy, but only when compared with isoniazid regimens of >12 months[152]. An RCT comparing 1HP versus 9H among PLHIV found similar risks of serious adverse events overall (5 vs 6%) and slightly lower grade 3 or higher hepatotoxicity with 1HP (2 vs 3%), while no hypersensitivity events were reported.

In summary, rifamycin-containing regimens appear to have lower risks of hepatotoxicity than longer isoniazid monotherapy regimens, though rifapentine may be associated with a higher risk of hypersensitivity reactions. These small but considerable risks of adverse events must be weighed against the risk of TB disease on an individual level when considering initiation of preventative treatment.

1.7.4 Treatment for incipient TB

Treatment for TB has historically been dichotomised as either full treatment for TB disease, or preventative treatment for LTBI. This dichotomy means that optimal therapies for people in the incipient phase, between these two extremes, are yet to be defined. As discussed in 1.7.1, both historic and recent analyses have demonstrated that treatment may be tailored according to bacillary burden, with truncated regimens appearing effective for people with low-positive or negative sputum smears[139,140]. However, in the recently reported CORTIS trial, 3HP was ineffective in reducing TB incidence among HIV-uninfected adults with a positive blood transcriptomic biomarker over 15 months, suggesting that a preventative treatment regimen may be insufficiently sterilising among people with incipient TB[123]. However, a post-hoc analysis showed that, among people who took at least 11/12 doses, there were no incident TB cases until 9 months, suggesting that suboptimal adherence and ongoing community transmission may also be contributing factors to these findings.

1.8 Targeting LTBI screening

1.8.1 Target populations

The overarching aim of LTBI screening is to identify people at high risk of TB who may benefit from preventative treatment and among whom the benefits outweigh the risks (Chapter 1.7.3). Moreover, as discussed in sections 1.4.4 and 1.4.7, the positive predictive values of TST and IGRAs for incident TB are consistently <5% over a 2-year interval among adult risk groups, and these tests cannot discriminate persistent from cleared *M. tuberculosis* infection. There is therefore a clear need to target LTBI screening towards people at higher risk of disease, in order to increase prior probability and maximise potential benefits of preventative treatment. WHO guidance on target groups for LTBI screening and preventative treatment are summarised in Table 1-9. Among all target risk groups, it is imperative to evaluate people for evidence of TB disease prior to initiating preventative treatment, as individuals with disease require higher intensity therapy (Chapter 1.7.1). This evaluation may include symptom screening, along with chest radiography when available and microbiological tests when indicated (Chapter 1.3)[81].

Table 1-9: Recommended screening groups in WHO LTBI guidance.

Adapted from WHO LTBI guidance[81].

People living with HIV
Adults and adolescents living with HIV who are unlikely to have active TB should receive TB preventative treatment, including those on antiretroviral treatment, pregnant women and those who have previously been treated for TB, irrespective of the degree of immunosuppression and even if LTBI testing is unavailable.
Infants aged < 12 months living with HIV who are in contact with a person with TB should receive TB preventative treatment.
Children aged ≥ 12 months living with HIV should be offered TB preventative treatment if they live in a setting with high TB transmission, regardless of contact with TB.
All children living with HIV who have successfully completed treatment for TB disease may receive TB preventative treatment.
Household contacts (regardless of HIV status)
Children aged < 5 years who are household contacts of people with bacteriologically confirmed pulmonary TB should be given TB preventative treatment even if LTBI testing is unavailable
Children aged ≥ 5 years, adolescents and adults who are household contacts of people with bacteriologically confirmed pulmonary TB may be given TB preventative treatment
In selected high-risk household contacts of patients with multidrug-resistant tuberculosis, preventative treatment may be considered based on individualized risk assessment and a sound clinical justification.
Other people at risk
People who are initiating anti-TNF treatment, or receiving dialysis, or preparing for an organ or haematological transplant, or who have silicosis should be systematically tested and treated for LTBI.
Systematic LTBI testing and treatment may be considered for prisoners, health workers, immigrants from countries with a high TB burden, homeless people and people who use drugs
Systematic LTBI testing and treatment is not recommended for people with diabetes, people who engage in the harmful use of alcohol, tobacco smokers and underweight people unless they also belong to other risk groups included in the above recommendations.

WHO guidance prioritises recent household contacts of people with TB, PLHIV (irrespective of CD4 count) and other immunosuppressed groups (including people initiating anti-TNF treatment, receiving dialysis, preparing for an organ or haematological transplant, or with silicosis) in particular for LTBI screening and preventative treatment[81]. European guidance also recommends prioritised LTBI screening among these risk groups, with the addition of people with pulmonary fibrotic lesions[82].

While UK guidelines also recommend screening contacts and immunosuppressed groups, specific recommendations contrast with WHO guidance in a number of ways[74]. For example, UK guidance recommends

screening ‘close’ contacts (defined as people who have had prolonged, frequent or intense contact with a person with infectious TB), as opposed to household contacts only in WHO guidance. Among immunocompromised adults (including PLHIV), UK guidance recommends an individualised risk assessment to establish whether testing should be offered (based on degree of immunocompromise and exposure to *M. tuberculosis*). In addition, since 2015, systematic LTBI testing and treatment has been implemented for people aged 16-35 who arrived in the UK in the previous five years from countries with annual TB incidence rates >150/100,000[163].

1.8.2 Risk factors for progression from latent infection to TB disease

WHO, European and UK guidelines are based upon observations of increased TB risk among the populations specified. Table 1-10 summarises relative risk estimates for TB disease among people with LTBI, compared to people with no risk factors and without evidence of TB disease, based on a previous literature review[164]. Increased risk is associated with demographics (with particularly high risk among young children)[165], recent infection with *M. tuberculosis* (for example household and close contacts), comorbidities, immunocompromise, and radiographic abnormalities. However, such relative risk estimates and are often imprecise with wide point estimate ranges. They also assume multiplicative risk, without accounting for potential interactions between risk factors. Moreover, these estimates are frequently derived from cohort and case-control studies in which there is a risk of residual confounding, including due to differential risk of infection with *M. tuberculosis*, and without adjustment for TST and IGRA results. Nonetheless, these data can provide a framework to facilitate targeting of LTBI screening towards higher risk populations.

Table 1-10: Risk factors for progression from LTBI to TB disease from Canadian national guidance.

Adapted from [164].

Risk factor	Estimated relative risk for TB, compared to people with no known risk factor
High risk	
AIDS	110-170
HIV	50-110
Transplantation	20-74
Silicosis	30
Chronic renal failure requiring haemodialysis	7-50
Carcinoma of head and neck	11.6
Recent TB infection (<2 years)	15
Fibronodular radiographic disease	6-19
Moderate risk	
Tumour necrosis factor alpha inhibitors	1.5-5.8
Diabetes mellitus	2-3.6
Treatment with glucocorticoids (≥15mg/d prednisolone)	4.9
Age <4 years when infected	2.2-5
Slightly increased risk	
Heavy alcohol consumption (≥3 drinks/day)	3-4
Underweight (<90% ideal body weight, e.g. body mass index ≤20)	2-3.6
Cigarette smoker (20 cigarettes/day)	1.8-3.5
Chest x-ray granuloma	2

1.8.3 Interpreting LTBI tests in context

Since the risk of progression to TB is heterogeneous among people with LTBI, a further approach that may improve prediction of incident TB is to include LTBI test results along with other clinical covariates (including demographics, exposure to TB, and immune function) in a multivariable prognostic model, in order to interpret the result in context of the individual tested. An existing tool ("TSTin3D") aims to achieve this by calculating personalised risk of TB disease among people with latent infection[166]. These estimates are calculated using an equation, which is parameterised mathematically from multiple sources, as opposed to being based on individual-level data. The algorithm works by first

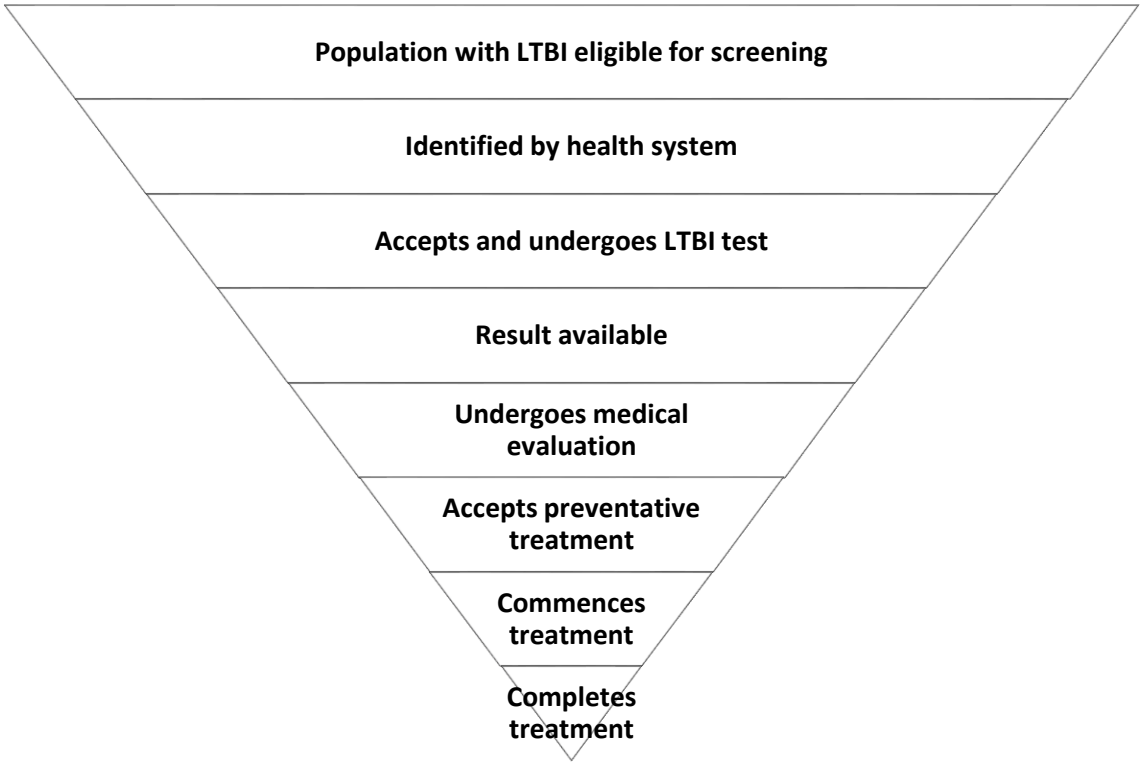
estimating the positive predictive value of the LTBI test as the patient's probability of having true latent *M. tuberculosis* infection, as determined by the likelihood of TB infection and the probability of a false positive test[167]. This is then multiplied by 'Annual Risk of Disease' (estimated as 0.1% in healthy unexposed people and 5% for the first two years from recent contact) and the relative risk estimates for specified risk factors taken from previous studies done anywhere in the world (summarised in Table 1-10). The performance of TSTin3D has recently been evaluated in a Canadian cohort, demonstrating moderate discrimination (C-statistics 0.66-0.68) and evidence of miscalibration with overestimation of risk[168]. Nonetheless, this provides a useful theoretical framework to conceptualise personalised risk beyond TST and IGRA results alone.

1.8.4 Cascade of LTBI care

A previous systematic review and meta-analysis has also highlighted major weaknesses in the cascade of LTBI care that are likely to undermine the impact and cost-effectiveness of screening programmes[169]. Losses occurred at multiple steps in the cascade, with only 10-50% of people who may be eligible for preventative treatment completing therapy. This evidence is supported by national data from the NHS England migrant LTBI screening programme, where the proportion of migrants with positive LTBI test result who are known to have accessed treatment was only 39% in 2019[19]. These data highlight a need for TB prevention services to strengthen each stage of the cascade in order to achieve population-level impact (Figure 1-13).

Figure 1-13: Steps in the cascade of LTBI care.

Adapted from Alsdurf et al[169].



1.9 Summary of motivation for thesis

Scale up of preventative treatment for TB represents a cornerstone of global TB control efforts as part of the WHO End TB Strategy[22]. However, both TST and IGRAs have low positive predictive value for incident TB. When implemented as part of LTBI screening programmes, this may lead to a significant burden of unnecessary preventative treatment, with associated risks of drug toxicity to patients, and economic costs to health services. Furthermore, limited positive predictive value may also undermine the uptake of LTBI treatment among target groups, due to the low perceived risk of TB[170].

Improvements in the risk-stratification of individuals with LTBI are therefore needed from both policymaker and patient perspectives. On a population level, these data would inform international policy regarding the populations that should be targeted for screening and treatment, enabling more efficient allocation of healthcare resources. On an individual patient level, better risk-stratification would facilitate a more personalised approach to TB prevention. Providing more accurate prognostication to clinicians and patients may also improve patient

engagement in care and therefore completion of preventative treatment when indicated, while concurrently reducing unnecessary treatment and the associated toxicity.

1.10 Research gaps

This thesis will address a number of research gaps.

First, optimising thresholds for existing LTBI tests might facilitate better implementation, while novel biomarkers are in development. Previous evaluations of the association between quantitative LTBI tests and incident TB risk have been limited to the QFT-GIT assay only[84–88]. It therefore remains unclear whether implementation of higher thresholds for positivity for QuantiFERON, T-SPOT.TB or TSTs may be of use programmatically to improve the risk-stratification of patients with LTBI.

Second, no studies have examined the prognostic value of the newer generation QuantiFERON assay (QFT-Plus) for predicting incident TB. The added value of the second TB antigen tube that aims to stimulate a $CD8^+$ T cell response is therefore unclear.

Third, LTBI tests may be further optimised by interpreting results in the context of the individual - accounting for demographics, exposure to *M. tuberculosis* and immune function. While an existing tool seeks to achieve this, it is not based on individual-level data and does not account for important covariates such as quantitative LTBI test results, proximity of exposure among contacts and infectiousness of index cases. Therefore, the potential added value of a directly data-driven multivariable prognostic model is currently untested.

Finally, there is also major interest in ‘next generation’ biomarkers targeting the incipient phase of TB, of which blood transcriptomic biomarkers are the most mature candidates. However, while a multitude of candidate transcriptomic biomarkers have been proposed, it remains unclear which of the many proposed candidate RNA signatures performs best, or whether any meets the WHO TPP benchmarks as incipient TB tests.

1.11 Aim and objectives

1.11.1 Aim

I aim to enable TB control strategies by facilitating precise targeting of preventative antimicrobial treatment.

1.11.2 Objectives

1. To test the hypothesis that higher quantitative QFT-GIT, T-SPOT.TB and TST results are associated with increased risk of incident TB, and assess the impact of implementing higher thresholds on test sensitivities, specificities and predictive values.
2. To assess the prognostic value of the newer generation QuantiFERON-TB Gold Plus test for predicting incident TB among recent TB contacts.
3. To examine the risk of TB among people tested for latent infection in low incidence settings and develop a prognostic model for incident TB using an individual participant data meta-analysis.
4. To systematically compare the diagnostic accuracy of candidate transcriptomic signatures for incipient TB, and to benchmark accuracy against the WHO target product profile parameters.

2 Objective 1: Quantitative interferon gamma release assay and tuberculin skin test results to predict incident TB: a prospective cohort study

2.1 Introduction

Previous studies have suggested that higher quantitative LTBI test results are associated with increased TB incidence rates[84,85,88]. However, these data have been limited to the QFT-GIT test only, and it has been unclear whether implementing higher thresholds may be of value programmatically.

I sought to address these knowledge gaps through a secondary analysis of data from the previously reported UK PREDICT TB study[80]. Firstly, I aimed to test the hypothesis that higher quantitative QFT-GIT, T-SPOT.TB and TST results were associated with increased risk of incident TB. Secondly, I sought to evaluate the test sensitivities, specificities and predictive values when higher thresholds for a positive test than currently recommended are used over a fixed three-year follow-up period. Finally, I plotted ROC curves for all three tests, to compare performance across the full range of test cut-offs.

2.2 Methods

2.2.1 Population

The UK PREDICT study cohort has been described in detail previously[80]. Briefly, individuals aged ≥16 years were recruited between May 2010 and June 2015 from multiple UK centres in London, Birmingham and Leicester. Inclusion criteria were: recent contacts of patients with active TB; or recent migrants from, or prolonged travellers to, high TB burden countries (defined as countries in sub-Saharan Africa or Asia). Participants who received preventative TB treatment after recruitment were excluded, as were participants diagnosed with suspected baseline prevalent TB (evidence of TB within 21 days of enrolment), since the aim of the study was to evaluate the risk of incident TB in the absence of preventative treatment.

2.2.2 Study procedures

Participants were tested with QFT-GIT, T-SPOT.TB and then Mantoux TST (Statens Serum Institut, Denmark) using standardised protocols on the same day, at least six weeks from last TB exposure or migration. Indeterminate results were classified as recommended by the manufacturers. Incident TB cases were identified via telephone interview at 12 and 24 months, and by linkage to the national TB surveillance system held at Public Health England in the originally reported UK PREDICT study[80]. National TB surveillance includes all statutory TB notifications and all results of positive *M. tuberculosis* cultures. For the current analysis, all study participants were re-linked to national surveillance records to identify individuals notified with TB until 31/12/2017. Follow-up was censored on the earliest of date of TB diagnosis, death or 31/12/2017. The original study procedures and protocol were approved by the Brent NHS Research Ethics Committee (10/H0717/14).

2.2.3 Statistical analysis

Analyses were performed using Stata (version 15) and R (version 3.5.1). Incidence rates and ratios relative to the negative test category (with 95% CIs) for incident TB were calculated using Poisson models, according to ordinal strata for quantitative results of each LTBI test during the full duration of follow-up. While Poisson models assume constant risk throughout follow-up[171], the rationale for this approach was that I sought to examine the prognostic ability of the index tests throughout follow-up, reflecting the way in which these tests are implemented in clinical practice, and consistent with the primary UK PREDICT analyses and other evaluations of prognostic tests for TB. In addition, since the date of testing can be considered arbitrary, there is no natural time scale for this analysis.

For participants with previous BCG vaccination (defined by self-report and scar inspection), 10mm was subtracted from the quantitative TST result to adjust for the associated sensitisation to BCG ('BCG-adjusted TST'). The rationale for this was that, in the main UK PREDICT analysis, a BCG-stratified TST cut-off of 5mm in BCG-naïve or 15mm in vaccinated participants performed most similarly to IGRA[80]. For QFT-GIT and BCG-adjusted TST, test strata were based on previous data[84,85,172–174], as shown in Table 2-1.

Table 2-1: Cut-offs for LTBI tests included in current analysis of UK PREDICT study.

QFT-GIT (IU/mL)	T-SPOT.TB (spots)	BCG-adjusted TST* (mm)
<0.35	<5	<5
0.35-0.69	5-7	5-9
0.7-3.99	8-49	10-14
≥4	≥50	≥15

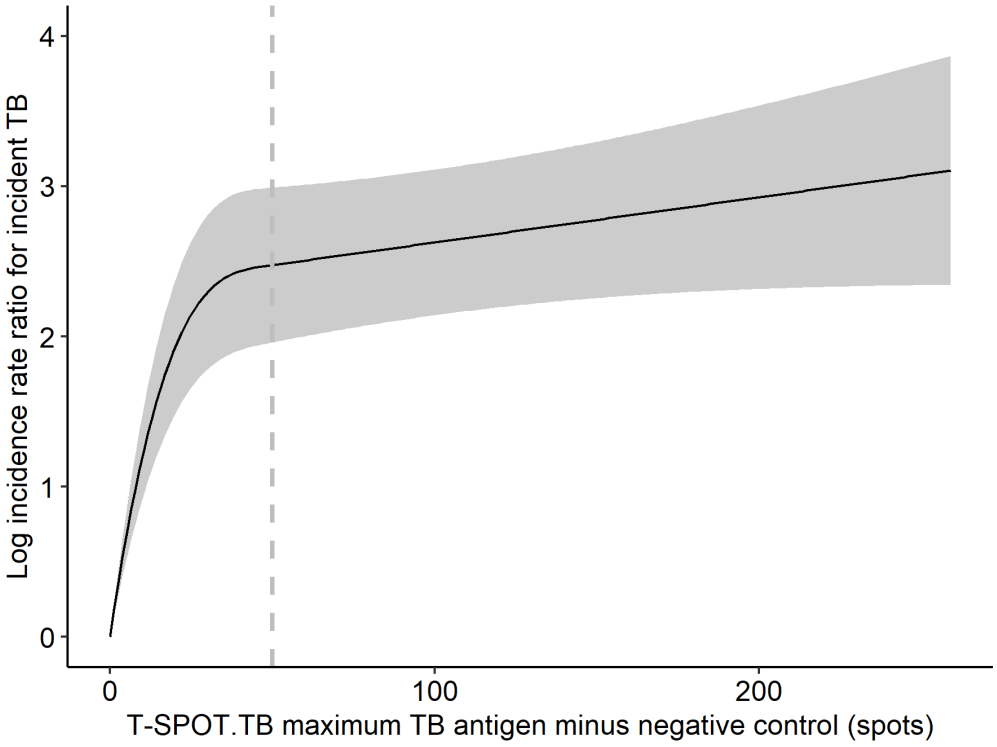
**Includes 10mm deduction for participants with prior BCG vaccination*

Since no previous data were available to inform the definition of test strata for T-SPOT.TB, I defined initial strata of spot counts in the maximal TB antigen panel minus the negative control of ≤4 spots and 5-7 spots, based on manufacturer test thresholds for borderline and positive results. I then modelled the non-linear relationship between the maximal T-SPOT.TB antigen minus negative control result and probability of incident TB using restricted cubic splines. Restricted cubic splines are a widely recommended approach to model associations between continuous predictors with non-linear associations and outcomes. In brief, this method applies piecewise cubic polynomials, which are forced to join at the junctions between the "pieces"[171]. The number of pieces and placement of the joining junctions are defined by "knot" placements. The predictor-outcome associations above and below the upper and lower knots are assumed to be linear in order to avoid overfitting to the tails of the distribution. Harrell recommends defining knot positions based on the distribution of the continuous predictor, using percentile placements[175].

In this analysis, I used restricted cubic splines with 4 knots, placed at recommended percentile locations[175], to examine the association between T-SPOT.TB results and incident TB risk in a Poisson regression model. Based on this, I defined ≥50 spots as the highest T-SPOT.TB stratum in the primary analysis (corresponding approximately to the top 5% of quantitative results in the cohort), since there was evidence of plateauing of risk above this point (Figure 2-1).

Figure 2-1: Poisson regression model using restricted cubic splines to examine association between quantitative T-SPOT.TB results and risk of incident TB.

Vertical dashed line indicates threshold of 50 spots used in primary analysis. Grey shaded area indicates 95% confidence intervals. Adapted from[176].



I then calculated sensitivity, specificity, positive and negative predictive values, plotted ROC curves and estimated the AUROCs for incident TB over a fixed three-year follow-up period at corresponding thresholds for each ordinal stratum. The AUROC metric can be interpreted as the probability that a randomly sampled person with the disease outcome has a higher score ranking than a randomly sampled disease-free person in the cohort[177]. An AUROC of 1 indicates perfect discrimination, while an AUROC of 0.5 is equivalent to an unbiased coin toss.

2.2.4 Sensitivity analyses

Six sensitivity analyses were performed:

- Prevalent TB was re-defined as a TB case diagnosed <42 days from enrolment (versus <21 days in the primary analysis).
- Subgroup analyses were performed restricted to only contacts and migrants, respectively.

- A fifth stratum was also included for each diagnostic test (≥8.00 IU/mL for QFT-GIT; ≥100 spots for T-SPOT.TB; ≥20mm for BCG-adjusted TST), due to the arbitrary nature of the defined thresholds.
- Quantitative TST results were analysed without adjustment for prior BCG vaccination, using unadjusted ordinal strata of <15mm, 15-19mm, 20-24mm and ≥25mm.
- I recalculated sensitivity, specificity and predictive values for a fixed follow-up period of six months to assess test performance for predicting short-term risk of progression.
- In the primary analysis, I did not perform a multivariable analysis to assess whether higher quantitative test results remained independently associated with risk of incident TB following adjustment for other covariates. The rationale for this was that my aim was to assess the potential programmatic impact of implementing higher diagnostic thresholds alone among target groups for LTBI screening programmes, rather than developing a multivariable risk prediction model. In a final sensitivity analysis, I examined associations between quantitative test results and incident TB in multivariable Poisson models adjusted for age, gender, ethnicity, country of birth and indication for screening (recent contact *vs.* migration), in order to examine whether associations between ordinal test strata and incident TB risk differed in multivariable, compared to primary univariable, analyses.

2.3 Results

2.3.1 Overview of study cohort

A total of 10,045 participants were recruited to the UK PREDICT study, of whom 175 had evidence of possible prevalent TB at enrolment and 260 received preventative treatment. The remaining 9,610 were therefore included in the final study cohort, with median follow-up 4.7 years (IQR 3.8-5.5).

Baseline characteristics of the cohort are summarised in Table 2-2. Median age was 33 years (IQR 26-47). Approximately half of participants were recruited due to being recent TB contacts (4,781/9,610; 49.8%) and migrants from high-incidence countries (4,729/9,610; 49.2%) respectively, while most (6,618/9,610;

68.9%) reported previous BCG vaccination. Where timing of previous BCG vaccination was available, this was reported to be >5 years prior to enrolment among 4,408/4,486 (98.3%) of participants. Quantitative results for QFT-GIT, T-SPOT.TB and TST were available for 8,562 (89.1%), 8,079 (84.1%) and 7,833 (81.5%) of participants, respectively.

There were 107 incident TB events during follow-up (median days to notification 188 days; IQR 76-488). Most incident TB notifications occurred in the first year following recruitment (71/107; 66.4%). Cumulative distribution plots for quantitative QFT-GIT, T-SPOT.TB, BCG-adjusted TST and unadjusted TST results, stratified by the incident TB outcome, are shown in Figure 2-2.

Table 2-2: Baseline characteristics of UK PREDICT TB cohort, stratified by whether or not participants progressed to incident TB during follow-up.

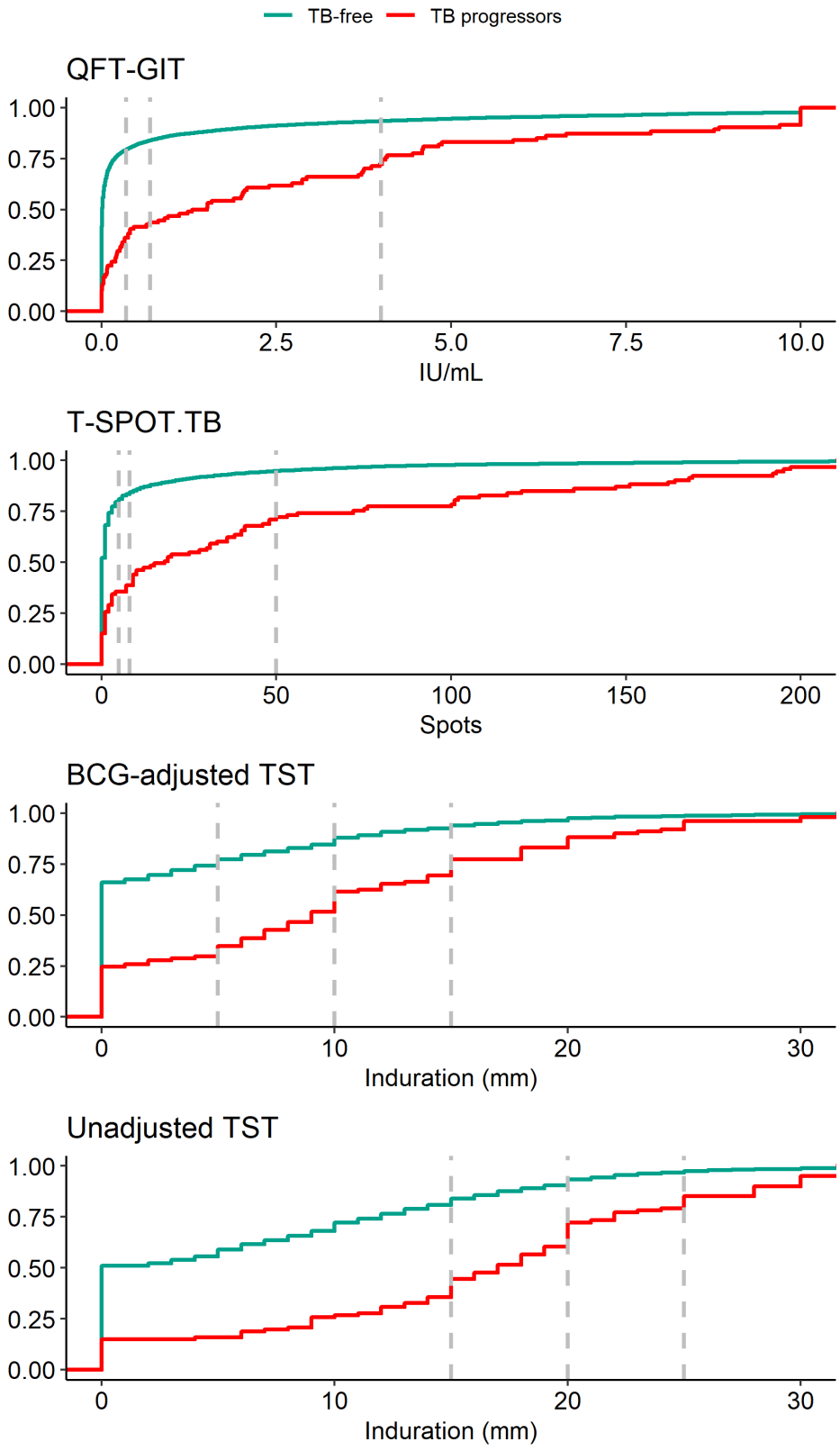
Data are presented as n (%) or median (IQR).

	TB-free	**TB progressors**	**Total**
Sex			
Male	4673 (49.2)	56 (52.3)	4729 (49.2)
Female	4758 (50.1)	51 (47.7)	4809 (50)
Missing	72 (0.8)	0 (0)	72 (0.7)
Age			
Median (IQR, range)	33 (26-47, 16-79)	30 (26-39, 16-65)	33 (26-47, 16-79)
Missing	22 (0.2)	0 (0)	22 (0.2)
Ethnicity			
Indian	3939 (41.5)	42 (39.3)	3981 (41.4)
White	1161 (12.2)	12 (11.2)	1173 (12.2)
Black African	1126 (11.8)	12 (11.2)	1138 (11.8)
Mixed	881 (9.3)	11 (10.3)	892 (9.3)
Pakistani	891 (9.4)	15 (14)	906 (9.4)
Bangladeshi	712 (7.5)	4 (3.7)	716 (7.5)
Black Caribbean	237 (2.5)	5 (4.7)	242 (2.5)
Other	315 (3.3)	5 (4.7)	320 (3.3)
Missing	241 (2.5)	1 (0.9)	242 (2.5)
UK Born			
No	7917 (83.3)	91 (85)	8008 (83.3)
Yes	1536 (16.2)	16 (15)	1552 (16.1)
Missing	50 (0.5)	0 (0)	50 (0.5)

Contact or migrant			
Contact	4711 (49.6)	70 (65.4)	4781 (49.8)
Migrant	4692 (49.4)	37 (34.6)	4729 (49.2)
Missing	100 (1.1)	0 (0)	100 (1)
QFT-GIT (IU/mL)			
<0.35	6603 (69.5)	34 (31.8)	6637 (69.1)
0.35-0.69	398 (4.2)	7 (6.5)	405 (4.2)
0.7-3.99	793 (8.3)	27 (25.2)	820 (8.5)
≥4	552 (5.8)	26 (24.3)	578 (6)
Indeterminate	119 (1.3)	3 (2.8)	122 (1.3)
Missing	1038 (10.9)	10 (9.3)	1048 (10.9)
T-SPOT.TB (spots)			
<5	6257 (65.8)	33 (30.8)	6290 (65.5)
5-7	316 (3.3)	3 (2.8)	319 (3.3)
8 to 49	876 (9.2)	30 (28)	906 (9.4)
≥50	416 (4.4)	27 (25.2)	443 (4.6)
Indeterminate	119 (1.3)	2 (1.9)	121 (1.3)
Missing	1519 (16)	12 (11.2)	1531 (15.9)
BCG-adjusted TST (mm)			
<5	5739 (60.4)	30 (28)	5769 (60)
5 to 9	805 (8.5)	22 (20.6)	827 (8.6)
10 to 14	612 (6.4)	18 (16.8)	630 (6.6)
≥15	576 (6.1)	31 (29)	607 (6.3)
Missing	1771 (18.6)	6 (5.6)	1777 (18.5)
Follow-up (years)			
Median (IQR)	4.69 (3.82-5.52)	0.51 (0.21-1.34)	4.68 (3.78-5.51)
Total	**9503**	**107**	**9610**

Figure 2-2: Cumulative distribution plots showing distributions of quantitative test results, stratified by outcome.

Dashed lines indicate thresholds used to define ordinal strata in this analysis. Adapted from [176].



2.3.2 TB incidence rates in ordinal test strata

For all tests, higher TB incidence rates and ratios were associated with higher test results (Figure 2-3). For QFT-GIT, TB incidence rates (per 1,000 person-years) increased from 1.10 (95% CI 0.78-1.53) in the <0.35 IU/mL stratum, to 10.02 (6.82-14.72) in the ≥4.00 IU/mL stratum (likelihood ratio test for trend p<0.0001). For T-SPOT.TB, TB incidence rates increased from 1.10 (0.78-1.54) in the ≤4 spots stratum to 12.73 (8.73-18.57) in the ≥50 spots stratum (p<0.0001). For the BCG-adjusted TST, TB incidence rates increased from 1.07 (0.75-1.54) in the <5mm stratum to 10.95 (7.70-15.57) in the ≥15mm stratum (p<0.0001).

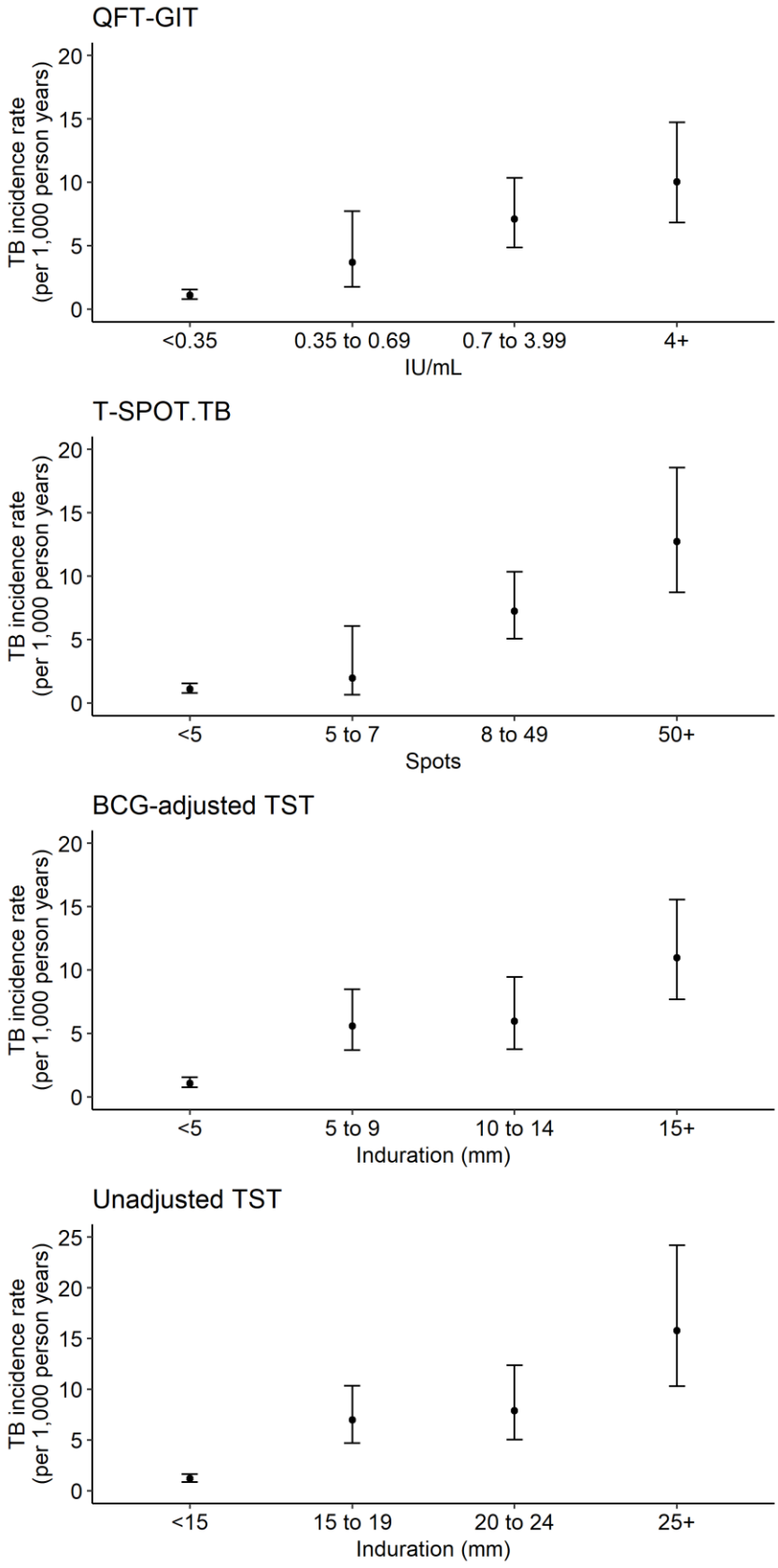
Table 2-3: TB incidence rates and ratios in ordinal strata for quantitative results of QFT-GIT, T-SPOT.TB and BCG-adjusted TST.

Data shown as point estimates (95% CIs).

	TB cases	**Person-years (1,000s)**	**Incidence rate (per 1,000 person years)**	**Incidence rate ratio**
QFT-GIT (IU/mL)				
<0.35	34	31.05	1.1 (0.78-1.53)	Reference
0.35-0.69	7	1.90	3.68 (1.75-7.71)	3.36 (1.49-7.57)
0.7-3.99	27	3.80	7.1 (4.87-10.35)	6.48 (3.91-10.74)
≥4	26	2.59	10.02 (6.82-14.72)	9.15 (5.49-15.25)
T-SPOT.TB (spots)				
<5	33	30.07	1.1 (0.78-1.54)	Reference
5 to 7	3	1.54	1.95 (0.63-6.06)	1.78 (0.55-5.8)
8 to 49	30	4.15	7.23 (5.06-10.35)	6.59 (4.02-10.81)
≥50	27	2.12	12.73 (8.73-18.57)	11.6 (6.98-19.3)
BCG-adjusted TST (mm)				
<5	30	27.91	1.07 (0.75-1.54)	Reference
5 to 9	22	3.94	5.58 (3.67-8.47)	5.19 (3-9)
10 to 14	18	3.03	5.95 (3.75-9.44)	5.54 (3.09-9.93)
≥15	31	2.83	10.95 (7.7-15.57)	10.19 (6.17-16.84)

Figure 2-3: TB incidence rates in ordinal strata for quantitative results of QFT-GIT, T-SPOT.TB and TST.

Adapted from [176].



2.3.3 Sensitivities, specificities and predictive values over 3 years' follow-up

During the first three years of follow-up, positive predictive values were uniformly low but increased modestly with the higher thresholds for all three tests (Table 2-4). For example, for QFT-GIT, positive predictive values were 3.0% (2.2-3.9) for values ≥0.35 IU/mL *vs.* 3.6% (2.2-5.5) for ≥4.00 IU/m. However, as thresholds for test positivity increased, sensitivity declined for all tests (Table 2-4). For the QFT-GIT, sensitivity decreased from 61.0% (49.6-71.6) with a threshold ≥0.35 to 23.2% (14.6-33.8) with a threshold ≥4.00 IU/mL. A similar pattern was seen for T-SPOT.TB and BCG-adjusted TST.

Table 2-4: Sensitivity, specificity, positive predictive values and negative predictive values during three years' follow-up with pre-specified test thresholds.

n = numerator; N = denominator.

		QFT-GIT (IU/mL)			T-SPOT.TB (spots)			BCG-adjusted TST (mm)		
		≥0.35	**≥0.7**	**≥4**	**≥5**	**≥8**	**≥50**	**≥5**	**≥10**	**≥15**
Sensitivity	*n*	50	44	19	53	50	22	62	42	25
	N	82	82	82	81	81	81	89	89	89
	Estimate	61.0%	53.7%	23.2%	65.4%	61.7%	27.2%	69.7%	47.2%	28.1%
	95% CI	49.6-71.6	42.3-64.7	14.6-33.8	54-75.7	50.3-72.3	17.9-38.2	59-79	36.5-58.1	19.1-38.6
Specificity	*n*	6134	6511	7242	5856	6155	6948	5520	6295	6882
	N	7755	7755	7755	7363	7363	7363	7445	7445	7445
	Estimate	79.1%	84.0%	93.4%	79.5%	83.6%	94.4%	74.1%	84.6%	92.4%
	95% CI	78.2-80	83.1-84.8	92.8-93.9	78.6-80.4	82.7-84.4	93.8-94.9	73.1-75.1	83.7-85.4	91.8-93
Positive predictive value	*n*	50	44	19	53	50	22	62	42	25
	N	1671	1288	532	1560	1258	437	1987	1192	588
	Estimate	3.0%	3.4%	3.6%	3.4%	4.0%	5.0%	3.1%	3.5%	4.3%
	95% CI	2.2-3.9	2.5-4.6	2.2-5.5	2.6-4.4	3-5.2	3.2-7.5	2.4-4	2.6-4.7	2.8-6.2
Negative predictive value	*n*	6134	6511	7242	5856	6155	6948	5520	6295	6882
	N	6166	6549	7305	5884	6186	7007	5547	6342	6946
	Estimate	99.5%	99.4%	99.1%	99.5%	99.5%	99.2%	99.5%	99.3%	99.1%
	95% CI	99.3-99.6	99.2-99.6	98.9-99.3	99.3-99.7	99.3-99.7	98.9-99.4	99.3-99.7	99-99.5	98.8-99.3

ROC curve analysis revealed similar AUROCs for quantitative QFT-GIT, T-SPOT.TB and BCG-adjusted TST for predicting incident TB over three years' follow-up, respectively (Table 2-5; Figure 2-4). Paired DeLong tests[178] revealed no statistical differences between AUROCs for QFT-GIT (p=0.21) or BCG-adjusted TST (p=0.14), when compared to T-SPOT.TB. The ROC curve points with maximal accuracy, as defined by the maximal Youden index[179], were similar to those currently used in practice (Table 2-5).

Figure 2-4: ROC curves for prediction of incident TB during three years' follow-up for quantitative QFT-GIT, T-SPOT.TB and TST.

Adapted from [176].

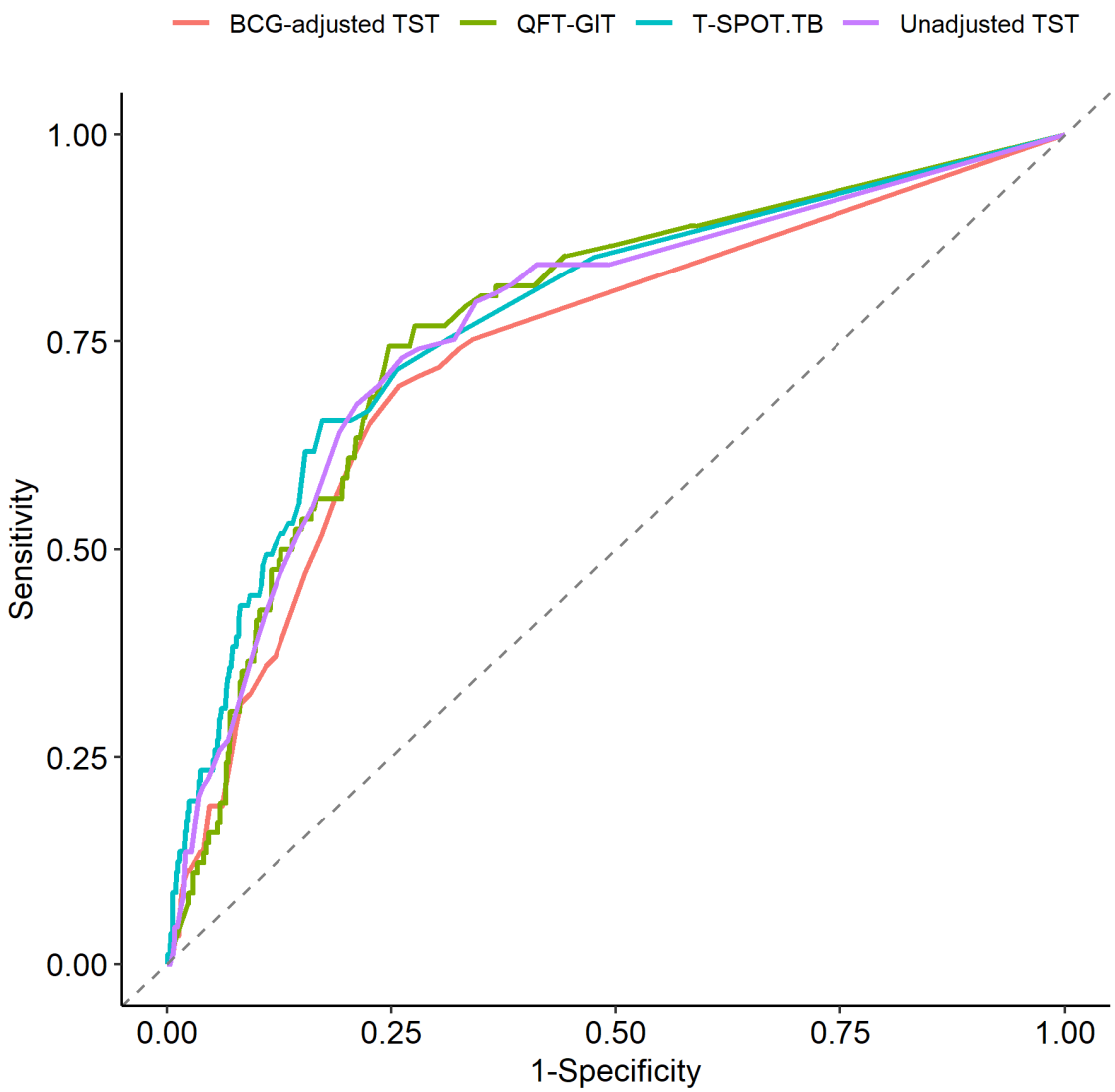


Table 2-5: AUROCs and thresholds with maximal Youden indices, with corresponding sensitivities and specificities, for index tests.

Data shown as point estimates (95% CI).

AUROC	Commonly used threshold	Maximal Youden index threshold	Sensitivity	Specificity
QFT-GIT				
0.77 (0.72-0.82)	≥0.35 IU/mL	0.2 IU/mL	0.74 (0.64 -0.83)	0.75 (0.74-0.76)
T-SPOT-TB				
0.78 (0.73-0.83)	≥6 spots	6.5 spots	0.65 (0.55-0.75)	0.83 (0.82-0.84)
BCG-adjusted TST				
0.74 (0.69-0.79)	≥5 mm*	4.5 mm	0.7 (0.59-0.78)	0.74 (0.73-0.75)
Unadjusted TST				
0.77 (0.72-0.82)	≥10 mm	11.5 mm	0.73 (0.63-0.81)	0.74 (0.73-0.75)

** equivalent to $TST_{5/15}$*

2.3.4 Sensitivity analyses

In the sensitivity analyses, exclusion of incident TB cases <42 days from enrolment resulted in slightly lower TB incidence rates across all strata, but had little impact on IRRs between strata compared to the primary analysis (Table 2-6). Similarly, the subgroup analyses restricted to TB contacts and migrants revealed lower TB incidence rates overall among migrants compared to TB contacts, but little difference in IRRs between strata for each test compared to the primary analysis (Table 2-7, Table 2-8, Table 2-9).

Inclusion of a fifth stratum for QFT-GIT (≥8 IU/mL) had no effect in increasing incidence rates further but, for T-SPOT.TB and BCG-adjusted TST (≥100 spots for T-SPOT.TB; ≥20mm for BCG-adjusted TST), led to further increases in the incidence rates in these strata (Table 2-10). However, as for the main analysis, there was a further loss of sensitivity when implementing corresponding diagnostic thresholds (Table 2-11).

Analysis of quantitative TST results without adjustment for prior BCG vaccination resulted in similar findings to the primary analysis, suggesting that BCG adjustment had minimal effect (Figure 2-2, Figure 2-3, Figure 2-4, Table 2-5).

Limiting follow-up to six months produced similar sensitivity, specificity and predictive value results to the main analysis, suggesting that imperfect test sensitivity is unlikely to be explained by re-exposure events during follow-up (Table 2-12).

Finally, associations between quantitative test results and incident TB in multivariable Poisson models were similar to the primary univariable analyses, suggesting that higher quantitative test results were associated with increased incident TB risk after adjustment for baseline covariates (Table 2-13).

Table 2-6: TB incidence rates according to ordinal test strata - sensitivity analysis excluding prevalent TB, defined as cases <42 days from enrolment.

Primary analysis defined prevalent TB as cases notified <21 days from recruitment. Shown as point estimates (95% CI). IR = incidence rate; IRR = incidence rate ratio.

QFT-GIT (IU/mL)	**TB cases**	**Person-years (1,000s)**	**IR (per 1,000 person-years)**	**IRR**
<0.35	28	31.05	0.9 (0.62-1.31)	Reference
0.35-0.69	6	1.90	3.15 (1.42-7.02)	3.49 (1.45-8.44)
0.7-3.99	25	3.80	6.57 (4.44-9.72)	7.29 (4.25-12.5)
≥4	23	2.59	8.87 (5.89-13.34)	9.83 (5.66-17.07)
T-SPOT.TB (spots)				
<5	31	30.07	1.03 (0.72-1.47)	Reference
5 to 7	2	1.54	1.3 (0.33-5.21)	1.26 (0.3-5.28)
8 to 49	26	4.15	6.27 (4.27-9.21)	6.08 (3.61-10.24)
≥50	22	2.12	10.38 (6.83-15.76)	10.07 (5.83-17.38)
BCG-adjusted TST (mm)				
<5	28	27.91	1 (0.69-1.45)	Reference
5 to 9	19	3.94	4.82 (3.07-7.56)	4.8 (2.68-8.6)
10 to 14	13	3.02	4.3 (2.5-7.4)	4.28 (2.22-8.27)
≥15	26	2.83	9.19 (6.26-13.49)	9.16 (5.37-15.62)

Table 2-7: TB incidence rates in ordinal test strata stratified by risk group for screening (contacts vs. migrants).

Shown as point estimates (95% CI). IR = incidence rate; IRR = incidence rate ratio.

Contacts				
QFT-GIT (IU/mL)	**TB cases**	**Person-years (1,000s)**	**IR (per 1,000 person-years)**	**IRR**
<0.35	22	16.08	1.37 (0.9-2.08)	Reference
0.35-0.69	5	0.90	5.55 (2.31-13.34)	3.36 (1.49-7.57)
0.7-3.99	18	2.01	8.97 (5.65-14.23)	6.48 (3.91-10.74)
≥4	15	1.11	13.53 (8.16-22.45)	9.15 (5.49-15.25)
T-SPOT.TB (spots)				
<5	25	16.56	1.51 (1.02-2.23)	Reference
5 to 7	3	0.89	3.39 (1.09-10.5)	2.24 (0.68-7.43)
8 to 49	16	2.04	7.86 (4.81-12.82)	5.2 (2.78-9.74)
≥50	18	1.10	16.37 (10.31-25.98)	10.84 (5.91-19.87)
BCG-adjusted TST (mm)				
<5	22	15.15	1.45 (0.96-2.21)	Reference
5 to 9	11	2.21	4.99 (2.76-9.01)	3.43 (1.67-7.08)
10 to 14	12	1.80	6.66 (3.78-11.74)	4.59 (2.27-9.27)
≥15	21	1.76	11.96 (7.8-18.34)	8.24 (4.53-14.98)
Migrants				
QFT-GIT (IU/mL)	**TB cases**	**Person-years (1,000s)**	**IR (per 1,000 person-years)**	**IRR**
<0.35	12	14.59	0.82 (0.47-1.45)	Reference
0.35-0.69	2	0.99	2.02 (0.5-8.07)	2.45 (0.55-10.96)
0.7-3.99	9	1.76	5.12 (2.66-9.84)	6.22 (2.62-14.77)
≥4	11	1.47	7.5 (4.16-13.55)	9.12 (4.03-20.68)
T-SPOT.TB (spots)				
<5	8	13.10	0.61 (0.31-1.22)	Reference
5 to 7	0	0.64	-	-
8 to 49	14	2.08	6.74 (3.99-11.37)	11.03 (4.63-26.3)
≥50	9	0.99	9.12 (4.75-17.53)	14.94 (5.76-38.71)
BCG-adjusted TST (mm)				
<5	8	12.34	0.65 (0.32-1.3)	Reference
5 to 9	11	1.71	6.45 (3.57-11.65)	9.95 (4-24.73)
10 to 14	6	1.20	4.98 (2.24-11.09)	7.68 (2.67-22.15)
≥15	10	1.06	9.42 (5.07-17.51)	14.53 (5.74-36.82)

Table 2-8: Sensitivity, specificity, positive predictive values and negative predictive values for LTBI tests – sensitivity analysis among contacts.

Shown for three years' follow-up with pre-specified test thresholds. n = numerator; N = denominator; PPV = positive predictive value; NPV = negative predictive value; CI = confidence interval.

		QFT-GIT (IU/mL)			**T-SPOT.TB (spots)**			**BCG-adjusted TST (mm)**		
		≥0.35	*≥0.7*	*≥4*	*≥5*	*≥8*	*≥50*	*≥5*	*≥10*	*≥15*
Sensitivity	*n*	31	27	10	32	29	14	38	28	17
	N	51	51	51	53	53	53	57	57	57
	Estimate	60.8%	52.9%	19.6%	60.4%	54.7%	26.4%	66.7%	49.1%	29.8%
	95% CI	(46.1-74.2)	(38.5-67.1)	(9.8-33.1)	(46-73.5)	(40.4-68.4)	(15.3-40.3)	(52.9-78.6)	(35.6-62.7)	(18.4-43.4)
Specificity	*n*	2999	3169	3533	3074	3239	3610	2847	3270	3606
	N	3740	3740	3740	3822	3822	3822	3945	3945	3945
	Estimate	80.2%	84.7%	94.5%	80.4%	84.7%	94.5%	72.2%	82.9%	91.4%
	95% CI	(78.9-81.5)	(83.5-85.9)	(93.7-95.2)	(79.1-81.7)	(83.6-85.9)	(93.7-95.2)	(70.7-73.6)	(81.7-84.1)	(90.5-92.3)
PPV	*n*	31	27	10	32	29	14	38	28	17
	N	772	598	217	780	612	226	1136	703	356
	Estimate	4.0%	4.5%	4.6%	4.1%	4.7%	6.2%	3.3%	4.0%	4.8%
	95% CI	(2.7-5.7)	(3-6.5)	(2.2-8.3)	(2.8-5.7)	(3.2-6.7)	(3.4-10.2)	(2.4-4.6)	(2.7-5.7)	(2.8-7.5)
NPV	*n*	2999	3169	3533	3074	3239	3610	2847	3270	3606
	N	3019	3193	3574	3095	3263	3649	2866	3299	3646
	Estimate	99.3%	99.2%	98.9%	99.3%	99.3%	98.9%	99.3%	99.1%	98.9%
	95% CI	(99-99.6)	(98.9-99.5)	(98.4-99.2)	(99-99.6)	(98.9-99.5)	(98.5-99.2)	(99-99.6)	(98.7-99.4)	(98.5-99.2)

Table 2-9: Sensitivity, specificity, positive predictive values and negative predictive values for LTBI tests – sensitivity analysis among migrants.

Shown for three years' follow-up with pre-specified test thresholds. n = numerator; N = denominator; PPV = positive predictive value; NPV = negative predictive value; CI = confidence interval.

		QFT-GIT (IU/mL)			**T-SPOT.TB (spots)**			**BCG-adjusted TST (mm)**		
		≥0.35	*≥0.7*	*≥4*	*≥5*	*≥8*	*≥50*	*≥5*	*≥10*	*≥15*
Sensitivity	*n*	19	17	9	21	21	8	24	14	8
	N	31	31	31	28	28	28	32	32	32
	Estimate	61.3%	54.8%	29.0%	75.0%	75.0%	28.6%	75.0%	43.8%	25.0%
	95% CI	(42.2-78.2)	(36-72.7)	(14.2-48)	(55.1-89.3)	(55.1-89.3)	(13.2-48.7)	(56.6-88.5)	(26.4-62.3)	(11.5-43.4)
Specificity	*n*	3065	3270	3629	2706	2838	3254	2600	2946	3193
	N	3932	3932	3932	3451	3451	3451	3414	3414	3414
	Estimate	78.0%	83.2%	92.3%	78.4%	82.2%	94.3%	76.2%	86.3%	93.5%
	95% CI	(76.6-79.2)	(82-84.3)	(91.4-93.1)	(77-79.8)	(80.9-83.5)	(93.5-95)	(74.7-77.6)	(85.1-87.4)	(92.6-94.3)
PPV	*n*	19	17	9	21	21	8	24	14	8
	N	886	679	312	766	634	205	838	482	229
	Estimate	2.1%	2.5%	2.9%	2.7%	3.3%	3.9%	2.9%	2.9%	3.5%
	95% CI	(1.3-3.3)	(1.5-4)	(1.3-5.4)	(1.7-4.2)	(2.1-5)	(1.7-7.5)	(1.8-4.2)	(1.6-4.8)	(1.5-6.8)
NPV	*n*	3065	3270	3629	2706	2838	3254	2600	2946	3193
	N	3077	3284	3651	2713	2845	3274	2608	2964	3217
	Estimate	99.6%	99.6%	99.4%	99.7%	99.8%	99.4%	99.7%	99.4%	99.3%
	95% CI	(99.3-99.8)	(99.3-99.8)	(99.1-99.6)	(99.5-99.9)	(99.5-99.9)	(99.1-99.6)	(99.4-99.9)	(99-99.6)	(98.9-99.5)

Table 2-10: TB incidence rates in ordinal test strata – sensitivity analysis including fifth stratum for quantitative results of each test.

Shown as point estimates (95% CI). IR = incidence rate.

QFT-GIT (IU/mL)	TB cases	Person-years (1,000s)	IR (per 1,000)
≥*8*	11	1.30	8.49 (4.70-15.33)
T-SPOT.TB (spots)			
≥*100*	21	0.94	22.42 (14.62-34.38)
BCG-adjusted TST (mm)			
≥*20*	17	1.36	12.46 (7.75-20.04)

Table 2-11: Sensitivity, specificity, positive predictive values and negative predictive values for LTBI tests – sensitivity analysis with additional test thresholds.

Shown during three years' follow-up. n = numerator; N = denominator.

		QFT-GIT ≥ 8 IU/mL	T-SPOT.TB ≥ 100 spots	BCG-adjusted TST ≥ 20mm
Sensitivity	*n*	9	16	12
	N	82	81	89
	Estimate	11.0%	19.8%	13.5%
	95% CI	(5.1-19.8)	(11.7-30.1)	(7.2-22.4)
Specificity	*n*	7501	7181	7172
	N	7755	7363	7445
	Estimate	96.7%	97.5%	96.3%
	95% CI	(96.3-97.1)	(97.1-97.9)	(95.9-96.7)
Positive predictive value	*n*	9	16	12
	N	263	198	285
	Estimate	3.4%	8.1%	4.2%
	95% CI	(1.6-6.4)	(4.7-12.8)	(2.2-7.2)
Negative predictive value	*n*	7501	7181	7172
	N	7574	7246	7249
	Estimate	99.0%	99.1%	98.9%
	95% CI	(98.8-99.2)	(98.9-99.3)	(98.7-99.2)

Table 2-12: Sensitivity, specificity, positive predictive values and negative predictive values in UK PREDICT TB cohort – sensitivity analysis, with follow-up limited to six months.

Calculated using pre-specified test thresholds. n = numerator; N = denominator; PPV = positive predictive value; NPV = negative predictive value.

		QFT-GIT (IU/mL)			**T-SPOT.TB (spots)**			**BCG-adjusted TST (mm)**		
		≥0.35	*≥0.7*	*≥4*	*≥5*	*≥8*	*≥50*	*≥5*	*≥10*	*≥15*
Sensitivity	*n*	26	22	13	33	31	17	39	29	18
	N	45	45	45	44	44	44	49	49	49
	Estimate	57.8%	48.9%	28.9%	75.0%	70.5%	38.6%	79.6%	59.2%	36.7%
	95% CI	(42.2-72.3)	(33.7-64.2)	(16.4-44.3)	(59.7-86.8)	(54.8-83.2)	(24.4-54.5)	(65.7-89.8)	(44.2-73)	(23.4-51.7)
Specificity	*n*	6618	7019	7829	6278	6595	7487	5758	6575	7194
	N	8394	8394	8394	7913	7913	7913	7783	7783	7783
	Estimate	78.8%	83.6%	93.3%	79.3%	83.3%	94.6%	74.0%	84.5%	92.4%
	95% CI	(78-79.7)	(82.8-84.4)	(92.7-93.8)	(78.4-80.2)	(82.5-84.2)	(94.1-95.1)	(73-75)	(83.7-85.3)	(91.8-93)
PPV	*n*	26	22	13	33	31	17	39	29	18
	N	1802	1397	578	1668	1349	443	2064	1237	607
	Estimate	1.4%	1.6%	2.2%	2.0%	2.3%	3.8%	1.9%	2.3%	3.0%
	95% CI	(0.9-2.1)	(1-2.4)	(1.2-3.8)	(1.4-2.8)	(1.6-3.2)	(2.3-6.1)	(1.3-2.6)	(1.6-3.3)	(1.8-4.6)
NPV	*n*	6618	7019	7829	6278	6595	7487	5758	6575	7194
	N	6637	7042	7861	6289	6608	7514	5768	6595	7225
	Estimate	99.7%	99.7%	99.6%	99.8%	99.8%	99.6%	99.8%	99.7%	99.6%
	95% CI	(99.6-99.8)	(99.5-99.8)	(99.4-99.7)	(99.7-99.9)	(99.7-99.9)	(99.5-99.8)	(99.7-99.9)	(99.5-99.8)	(99.4-99.7)

Table 2-13: Associations between quantitative test results and incident TB in multivariable Poisson models adjusted for age, gender, ethnicity, country of birth and indication for screening.

P values for categorical variables with multiple levels indicate likelihood ratio tests. Sample sizes for the multivariable models were 8081 (94 events) for QFT-GIT, 7616 (93 events) for T-SPOT.TB and 7506 (101 events) for BCG-adjusted TST. There was no evidence of multi-collinearity between the ordinal test strata and other variables (variance inflation factors 1.04-1.11). IRR = incidence rate ratio.

	IRR	95% CI	p		IRR	95% CI	p		IRR	95% CI	p
QFT-GIT (IU/mL)				**T-SPOT.TB (spots)**				**BCG-adjusted TST (mm)**			
<0.35	1 (ref)			*<5*	1 (ref)			*<5*	1 (ref)		
0.35-0.69	3.61	1.59-8.16	<0.0001	*5 to 7*	1.85	0.57-6.05	<0.0001	*5 to 9*	5.12	2.95-8.9	<0.0001
0.7-3.99	6.93	4.15-11.58		*8 to 49*	7.51	4.51-12.5		*10 to 14*	5.66	3.13-10.22	
≥4	10.36	6.15-17.44		*≥50*	13.41	7.91-22.73		*≥15*	10.30	6.19-17.12	
Age				**Age**				**Age**			
≤35	1 (ref)			*≤35*	1 (ref)			*≤35*	1 (ref)		
>35	0.65	0.42-0.99	0.046	*>35*	0.58	0.38-0.91	0.017	*>35*	0.71	0.47-1.07	0.10
Gender				**Gender**				**Gender**			
Male	1 (ref)			*Male*	1 (ref)			*Male*	1 (ref)		
Female	1.10	0.73-1.65	0.66	*Female*	1.07	0.71-1.62	0.75	*Female*	1.03	0.7-1.53	0.88
Ethnicity				**Ethnicity**				**Ethnicity**			
White	1 (ref)			*White*	1 (ref)			*White*	1 (ref)		
South Asian	0.98	0.5-1.94	0.68	*South Asian*	0.88	0.44-1.74	0.69	*South Asian*	1.03	0.52-2.06	0.95
Black African or Caribbean	0.68	0.3-1.54		*Black African or Caribbean*	0.66	0.3-1.47		*Black African or Caribbean*	0.87	0.4-1.91	
Other	0.89	0.4-1.97		*Other*	0.75	0.33-1.69		*Other*	1.00	0.45-2.21	
UK born				**UK born**				**UK born**			
No	1 (ref)			*No*	1 (ref)			*No*	1 (ref)		
Yes	0.96	0.52-1.78	0.90	*Yes*	0.96	0.51-1.82	0.91	*Yes*	0.78	0.42-1.44	0.43
Contact or migrant				**Contact or migrant**				**Contact or migrant**			
Migrant	1 (ref)			*Migrant*	1 (ref)			*Migrant*	1 (ref)		
Contact	1.90	1.2-3.02	0.007	*Contact*	1.98	1.24-3.18	0.004	*Contact*	1.47	0.94-2.3	0.09

2.4 Discussion

2.4.1 Summary of key findings in context of previous literature

In this study, I have demonstrated that higher quantitative results for the QFT-GIT, T-SPOT.TB and TST were strongly associated with higher TB incidence rates, supporting existing data derived among adult and paediatric populations, from low and high TB incidence settings respectively[84,85,88]. However, I found that implementing higher thresholds would lead to a marked loss of sensitivity for all three tests, whereby the majority of incident TB cases would test negative. A modest loss of test sensitivity may be acceptable in some circumstances, where the programmatic goal is to identify the subgroup with the highest risk of progression, if it is accompanied by a substantial improvement in positive predictive value. However, positive predictive values for all three tests remained modest, even in the highest ordinal strata and when limited to contacts, who generally have higher prior probability of developing TB disease. While this is likely partly a reflection of the low pre-test probability of incident TB in low TB incidence settings (such as the UK), even among risk groups for screening such as TB contacts and recent migrants, it also highlights the limitations of our existing LTBI tests for predicting incident TB when appropriately implemented in target populations.

A previous analysis showed that TST stratified by BCG status ($TST_{5/15}$) yielded comparable performance to both commercial IGRAs[80]. The ROC curve data in the current analysis reinforces this conclusion, as AUROCs were very similar for all three tests, with overlapping 95% confidence intervals. However, my sensitivity analysis using TST results without adjustment for prior BCG vaccination were very similar to the primary analysis, suggesting that BCG-adjustment was not required and that the $TST_{5/15}$ cut-off (based on previous UK guidance[180]) may have performed more similarly to IGRA in the original UK PREDICT analysis due to use of the TST_{15} cut-off among most participants, as opposed to being driven by BCG-stratification[80]. This interpretation is consistent with previous analyses showing that BCG vaccination during infancy is likely to have minimal impact on TST measurements later in life[181]. While UK vaccination policy included school age BCG vaccination prior to 2005, most participants in the UK PREDICT cohort were born abroad and were likely to have received vaccination in infancy[182].

Notably, my results demonstrated that higher quantitative TST results (with or without BCG-adjustment) were associated with higher TB incidence, in contrast to a previous cross-sectional report among 529 individuals suggesting that induration size above 5mm does not confer additional risk[183]. Taken together, these results can be explained by TST and IGRAs being differing ways of measuring a common underlying T cell mediated response to *M. tuberculosis* antigens, with stronger quantitative responses for each test being associated with higher incident TB risk.

2.4.2 Policy implications

These data demonstrate that, despite stronger responses being associated with higher TB risk, implementing higher diagnostic thresholds for QFT-GIT, T-SPOT.TB and TST is unlikely to be of use in settings aiming towards TB elimination, and fails to bridge the gap to the WHO TPP for incipient TB tests. One policy approach may be to offer preventative therapy to all patients with positive LTBI tests, using current thresholds, with the offer of additional support to complete treatment for those with higher quantitative results, who are at highest risk of disease. However, such an approach should ensure that current resources are not diverted away from supporting those with lower quantitative positive test results to commence and complete preventative therapy, since doing so may risk missing the majority of progressors.

Notably, the maximal Youden index values for each of the index tests were similar to currently implemented thresholds. The Youden index assumes equal weighting to sensitivity and specificity[179]. While the trade-off between sensitivity and specificity is likely to vary between settings, clinicians, and patients, the WHO TPP for incipient TB biomarkers gives equal weighting to each[11], suggesting that the maximal Youden Index approach may be considered as a reasonable approach towards defining an optimal threshold. Nonetheless, in view of the modest predictive value of all three LTBI tests, it remains imperative to ensure appropriate use in target risk groups who have higher prior probability of developing incident TB, in order to mitigate the risks of offering unnecessary preventative treatment to low risk individuals, with associated risks of drug toxicity to patients, economic costs to health services, and further undermining the cascade of LTBI care (Chapter 1.9).

2.4.3 Strengths of this study

This study has a number of strengths. Firstly, it is a large scale (n=9,610), prospective cohort study, with standardised clinical data capture and laboratory protocols. Extension of follow-up by re-linkage to national surveillance allowed lengthy median follow-up (approaching five years), and robust identification of incident TB cases through a validated linkage algorithm[184]. The study population, consisting of UK recent TB contacts and migrants from high TB incidence settings, is well characterised and highly representative of target groups for LTBI screening programmes in low TB incidence settings[62]. These findings are therefore likely to be generalisable to other low TB incidence countries, though additional data are required among adults in higher burden settings. Furthermore, availability of LTBI test results was high, with QFT-GIT, T-SPOT.TB and TST results available for 8,562 (89.1%), 8,079 (84.1%) and 7,833 (81.5%) of participants, respectively.

2.4.4 Study limitations

A limitation of this study was that an updated version of the QFT-GIT (QFT-Plus), which became available late in the follow-up period, was not assessed[92]. Prospective evaluations of the predictive value of this assay for incident TB are required.

Secondly, a positive LTBI test could potentially have led to differential work-up bias, since a positive result may have increased clinical suspicion of TB and therefore led to more TB diagnoses among participants with positive index prognostic tests and exaggerated incidence rate ratios. However, test results (for a single test done as part of routine care) were available to participants' clinicians only for TB contacts aged ≤35 years since, during the study period, these were the only participants who met national criteria for LTBI testing[172]. The magnitude of this bias is therefore likely to be small.

Thirdly, only baseline testing was performed. I was therefore unable to assess conversions or reversions during serial testing, which may be frequent[185]. Since the reality of both contact and migrant LTBI screening programmes is that serial testing (beyond 6 weeks post-contact) would have major cost implications, this limitation reflects the constraints of routine programmatic conditions; the ability of

the tests to identify progressors at the point of initial screening is therefore a key attribute.

I was also unable to account for any further TB exposure events between recruitment and diagnosis, which may have led to an underestimate of test sensitivity. However, 66.4% of progression events occurred in the first year (median time to disease 188 days), and there is a relatively low risk of TB transmission in the UK. Moreover, test sensitivity when using conventional thresholds remained 57.8-79.6% in my sensitivity analysis when follow-up was limited to six months; the impact of this bias is therefore likely small.

Exclusion of people who received preventative therapy may have introduced selection bias if these individuals were at higher risk of incident TB, resulting in systematic underestimation of incident TB risk in the study cohort. Since participants with positive LTBI test results (using current thresholds) are more likely to be offered preventative treatment, this may have led to underestimation of incidence rate ratios in ordinal test strata, compared to the test negative group. This limitation reflects a ubiquitous challenge for studies examining the prognostic ability of LTBI tests when conducted among populations where preventative treatment is the current standard of care for people with LTBI. However, only 260 people were excluded on this basis in the current study.

Finally, the primary analyses conducted in this chapter were univariable, with 107 outcome events. I undertook a multivariable sensitivity analysis to evaluate whether associations between LTBI test ordinal strata and incident TB risk changed following adjustment for baseline co-variates. While the multivariable analysis was limited by power due to a relatively small number of outcome events, this analysis showed little change in the observed incidence rate ratios, thus supporting the primary analysis findings.

2.4.5 Conclusion

Optimal implementation of existing LTBI diagnostic tests is critical while novel commercial assays with improved predictive value are developed. While higher quantitative QFT-GIT, T-SPOT.TB and TST results were associated with higher TB incidence rates in this study, the implementation of higher diagnostic thresholds for these tests comes at the cost of a marked loss of sensitivity, such

that only a minority of incident TB cases are detected. Moreover, the improvement in positive predictive value with higher test thresholds was modest. Incorporation of quantitative results into validated multivariable risk prediction models may be of use to further improve prediction of incident TB. However, a better biomarker may ultimately be required to transform risk-stratification of patients with LTBI.

2.5 Contribution statement

This study was a secondary analysis of the UK PREDICT prospective cohort study. I led the analyses presented in this chapter - from protocol development to completion and dissemination - and co-ordinated data linkage to the UK Enhanced TB Surveillance system, in order to extend follow-up of the full study cohort for the primary incident TB outcome.

2.6 Outputs relating to this chapter

This study is published in the *American Journal of Respiratory and Critical Care Medicine*:

Gupta RK, Lipman M, Jackson C, Sitch A, Southern J, Drobniewski F, Deeks JJ, Tsou C, Griffiths C, Davidson J, Campbell C, Stirrup O, Noursadeghi M, Kunst H, Haldar P, Lalvani A, Abubakar I (2020). Quantitative interferon gamma release assay and tuberculin skin test results to predict incident tuberculosis: a prospective cohort study. **Am J Resp Crit Care Med.** https://doi.org/10.1164/rccm.201905-0969OC

I also presented this work as an oral presentation at the European Respiratory Society Congress, Madrid in September 2019 (awarded a British Thoracic Society Travel Grant).

In addition, I led the analysis of a follow-up study to UK PREDICT which evaluated the yield of screening for LTBI and TB disease in London Emergency Departments and is published in the *European Respiratory Journal*:

Gupta RK, Lule SA, Krutikov M,…, Abubakar I (2021). Screening for tuberculosis among high-risk groups attending London Emergency

Departments: A prospective observational study. **European Respiratory Journal**; in press https://doi.org/10.1183/13993003.03831-2020

3 Objective 2: Evaluation of QuantiFERON-TB Gold Plus for predicting incident TB among recent contacts: a prospective cohort study

3.1 Introduction

A newer generation QuantiFERON assay (QFT-Plus) was recently launched, adding a second TB antigen tube (TB2). This incorporates short peptides designed to stimulate a $CD8^+$ T cell response, in addition to the $CD4^+$-response tube (TB1) included in previous versions. No previous studies have examined the predictive ability of QFT-Plus for incident TB, or assessed whether the addition of the CD8-targeted antigen tube adds any prognostic value to the assay. I aimed to address these knowledge gaps in a previously established cohort of UK TB contacts.

3.2 Methods

3.2.1 Population and recruitment procedures

Adult (≥16 years) contacts of pulmonary and extra-pulmonary TB index cases were recruited to the Next Generation Tests for Latent Tuberculosis Infection existing study from ten London TB clinics, when attending for routine contact screening, between July 2015 and November 2016. Following recruitment, participants completed a questionnaire and blood sampling was conducted for the QFT-Plus assay (at least six weeks from last known TB exposure). Contacts with evidence of prevalent TB disease (defined as TB diagnosed within 21 days of enrolment as per previous work[80]), and those accepting preventative therapy (offered in accordance with contemporary national guidance[74,172]) were excluded from analysis.

3.2.2 Outcomes

Participants were linked to national TB surveillance records held by Public Health England, including all statutory TB notifications, to identify those notified with TB (until 31/12/2017). Notified TB cases were validated by local record review and included culture-confirmed TB or clinically diagnosed with radiological or

histological evidence of TB, where a clinician had prescribed treatment with a full course of anti-TB treatment.

3.2.3 Statistical analysis

Analyses were conducted in Stata (version 15) or R (version 3.5.1). QFT-Plus results were interpreted as per manufacturer guidance, with TB antigen responses calculated as TB antigen interferon-γ minus unstimulated control interferon-γ for all analyses[64]. I calculated incidence rates and rate ratios relative to the negative test category, along with sensitivity, specificity and predictive values, including the full duration of follow-up.

To assess the correlation and agreement between the $CD4^+$- $CD8^+$-stimulating tubes, I plotted scatterplots and Bland-Altman plots of TB1 versus TB2 and calculated Spearman rank correlation (in view of non-normal distributions) and mean differences between tubes[186]. I then compared ROC curves and AUROCs when using TB1 only, TB2 only, and the maximal TB antigen tube (TBmax; higher of TB1 and TB2), in order to further evaluate potential incremental value of adding the $CD8^+$-stimulating tube in predicting incident TB cases. I also plotted a ROC curve for the calculated difference between the TB1 and TB2 tubes (TB2-TB1) as a surrogate for the $CD8^+$-specific response, since it has been hypothesised that this may identify contacts with recently acquired *M. tuberculosis* infection, who are at highest risk of TB disease[93].

3.3 Results

3.3.1 Overview of study cohort

A total of 623 contacts were recruited (Figure 3-1), of whom 532 (85.4%) had QFT-Plus results (89 missing; 2 indeterminate) and were followed-up for a median 1.93 years (IQR 1.65-2.21). QFT-Plus results were positive in 180/532 (33.8%) (Table 3-1), of whom 39 (21.7%) commenced preventative therapy. After excluding participants with missing results, those commencing preventative treatment, and with prevalent TB (n=1; notified 3 days after recruitment), I included 492 participants in the primary analysis.

Baseline characteristics were similar between included and excluded participants, except for those who commenced preventative therapy being younger than those who did not (Table 3-1).

Figure 3-1: Study flow chart showing participants included and excluded in primary analysis.

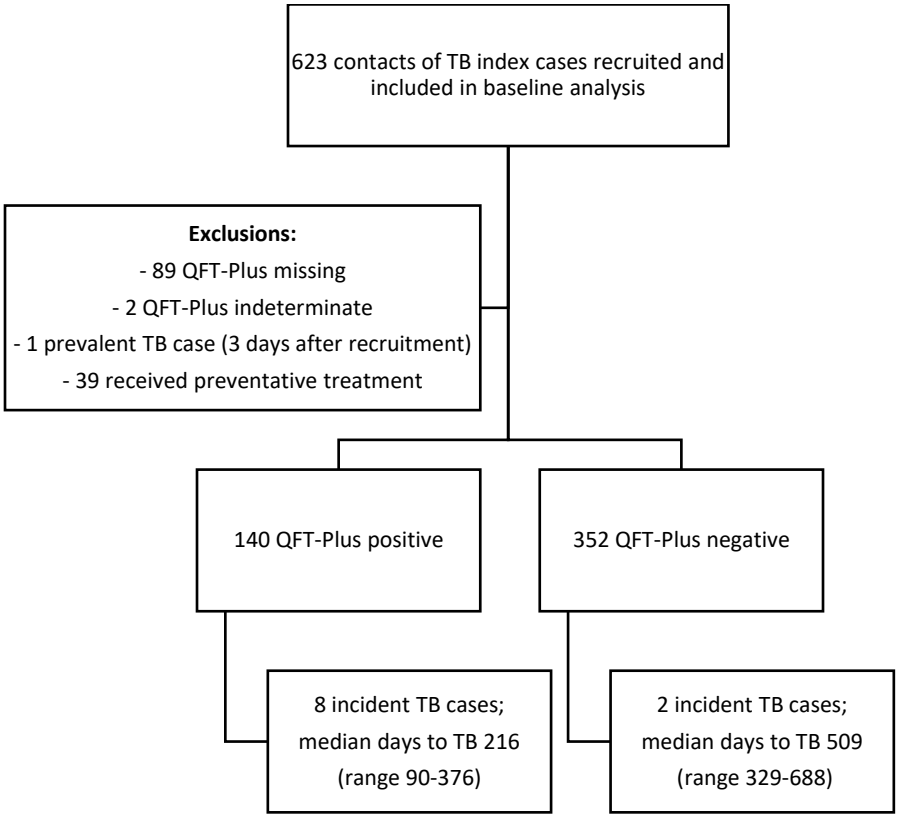


Table 3-1: Baseline characteristics of study cohort, stratified by QuantiFERON-TB Gold Plus (QFT-Plus) results and provision of preventative therapy (PT).

Data presented as n(%), unless stated otherwise. IQR = interquartile range.

	QFT-Plus -[#]	QFT-Plus +		QFT-Plus missing*	All
		No PT[#]	*PT*		
Age					
Median (IQR)	31 (25-43)	43 (32-54)	30 (26-35)	31.5 (23.7-49)	33 (25-46)
Missing	3 (0.9)	2 (1.4)	0 (0)	1 (1.1)	6 (1)
Gender					
Male	165 (46.9)	76 (53.9)	24 (61.5)	37 (40.7)	302 (48.5)
Female	180 (51.1)	62 (44)	15 (38.5)	51 (56)	308 (49.4)
Missing	7 (2)	3 (2.1)	0 (0)	3 (3.3)	13 (2.1)
Ethnicity					
White	95 (27)	27 (19.1)	9 (23.1)	31 (34.1)	162 (26)
South Asian	117 (33.2)	55 (39)	13 (33.3)	33 (36.3)	218 (35)
Black African or Caribbean	67 (19)	30 (21.3)	7 (17.9)	15 (16.5)	119 (19.1)
Other	63 (17.9)	24 (17)	10 (25.6)	9 (9.9)	106 (17)
Missing	10 (2.8)	5 (3.5)	0 (0)	3 (3.3)	18 (2.9)
UK born					
No	235 (66.8)	126 (89.4)	33 (84.6)	66 (72.5)	460 (73.8)
Yes	111 (31.5)	11 (7.8)	6 (15.4)	24 (26.4)	152 (24.4)
Missing	6 (1.7)	4 (2.8)	0 (0)	1 (1.1)	11 (1.8)
Contact type					
Household	210 (59.7)	96 (68.1)	30 (76.9)	49 (53.8)	385 (61.8)
Family non-household	19 (5.4)	7 (5)	2 (5.1)	3 (3.3)	31 (5)
Work or Social	62 (17.6)	19 (13.5)	4 (10.3)	14 (15.4)	99 (15.9)
Other	13 (3.7)	3 (2.1)	2 (5.1)	2 (2.2)	20 (3.2)
Missing	48 (13.6)	16 (11.3)	1 (2.6)	23 (25.3)	88 (14.1)
Diabetes					
No	318 (90.3)	120 (85.1)	38 (97.4)	83 (91.2)	559 (89.7)
Yes	20 (5.7)	18 (12.8)	0 (0)	6 (6.6)	44 (7.1)
Missing	14 (4)	3 (2.1)	1 (2.6)	2 (2.2)	20 (3.2)
HIV					
No	331 (94)	137 (97.2)	37 (94.9)	84 (92.3)	589 (94.5)
Yes	4 (1.1)	0 (0)	1 (2.6)	2 (2.2)	7 (1.1)
Missing	17 (4.8)	4 (2.8)	1 (2.6)	5 (5.5)	27 (4.3)
Total	**352**	**141**	**39**	**91**	**623**

3.3.2 Prognostic value of QFT-Plus

Ten incident TB cases were notified during follow-up (median 222 days after recruitment; range 90-688). TB incidence rates (per 1,000 person-years) were 30.6 (95% CI 15.3-61.1) and 3.0 (0.8-12.1) in the QFT-Plus positive and negative groups, respectively, giving an IRR of 10.1 (2.2-47.7; Table 3-2). QFT-Plus sensitivity for incident TB was 80.0% (44.4-97.5). The positive and negative predictive values were 5.7% (2.5-10.9) and 99.4% (98.0-99.9), respectively. Characteristics of the participants notified with prevalent or incident TB during follow-up are shown in Table 3-3.

Table 3-2: Incidence rates, rate ratios, and predictive values for incident TB during follow-up, stratified by QuantiFERON-TB Gold Plus (QFT-Plus) result.

Data presented as point estimates (95% CI).

	QFT-Plus +	**QFT-Plus -**
TB cases	8	2
Total participants	140	352
Person-years	261.6	663.0
Incidence rate per 1,000 person-years	30.6 (15.3-61.1)	3.0 (0.8-12.1)
Incidence rate ratio	10.1 (2.2-47.7)	
Positive predictive value	5.7 (2.5-10.9)	
Negative predictive value	99.4 (98.0-99.9)	
Sensitivity	80.0 (44.4-97.5)	
Specificity	72.6 (68.4-76.5)	

Table 3-3: Characteristics of prevalent and incident TB cases notified during follow-up.

Gender	Age (years)	Ethnicity	QFT-Plus	Interval to TB (days)	TB Culture	Site of disease	Immunocompromise
Female	32	Indian	Positive	3 (prevalent)	Positive	Uterus	None
Male	16	Black African	Positive	90	Negative	Pleural	None
Male	21	Mixed	Positive	137	Negative	Lymph node	None
Female	44	Indian	Positive	182	Positive	Pulmonary	None
Female	21	Mixed	Positive	210	Negative	Lymph node	None
Male	33	Indian	Positive	222	Positive	Lymph node	None
Male	52	Indian	Positive	296	Positive	Pulmonary	Diabetic
Male	24	Black African	Negative	329	Positive	Pleural	None
Female	27	Black African	Positive	342	Negative	Pleural	None
Male	44	Mixed	Positive	376	Negative	Pleural	None
Male	27	Black African	Negative	688	Negative	Lymph node	None

3.3.3 Incremental value of TB2 tube

There was strong correlation and agreement between the TB1 and TB2 interferon-γ responses (Spearman rank *R* = 0.9; p <0.001; mean difference 0.03 IU/mL; Figure 3-2). ROC curves for prediction of incident TB during all follow-up were similar for the TB1, TB2 and maximal TB antigen responses (AUROCs 0.80-0.82; Figure 3-3). TB2 minus TB1, however, did not discriminate TB progressors from non-progressors (AUROC 0.44; 95% CI 0.20-0.68).

Figure 3-2: Association of interferon-γ responses in the TB1 and TB2 tubes shown as (a) scatterplot; and (b) Bland-Altman plot.

R indicates Spearman rank correlation with accompanying p value. Adapted from [187].

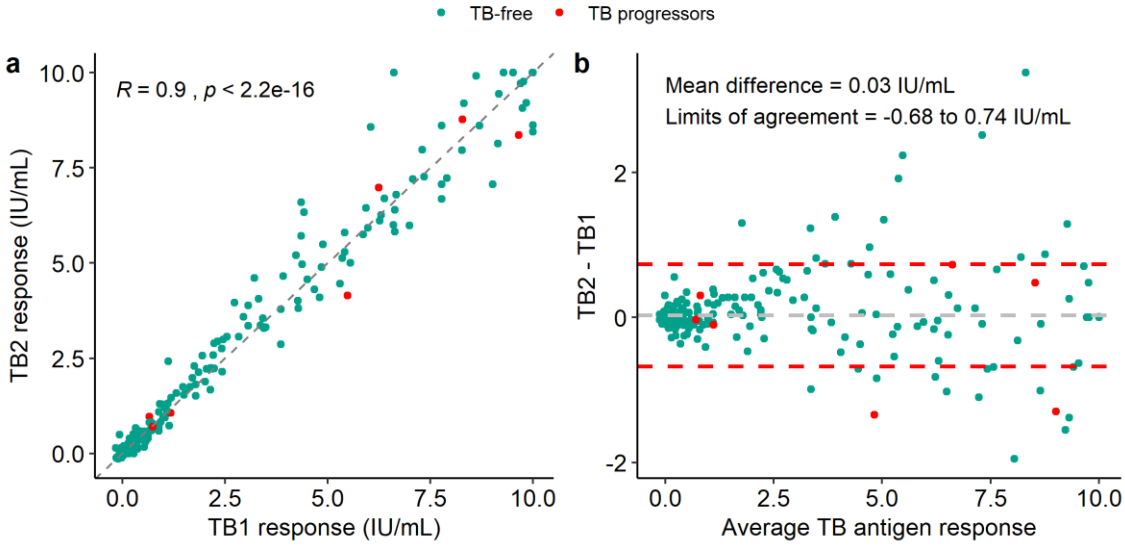
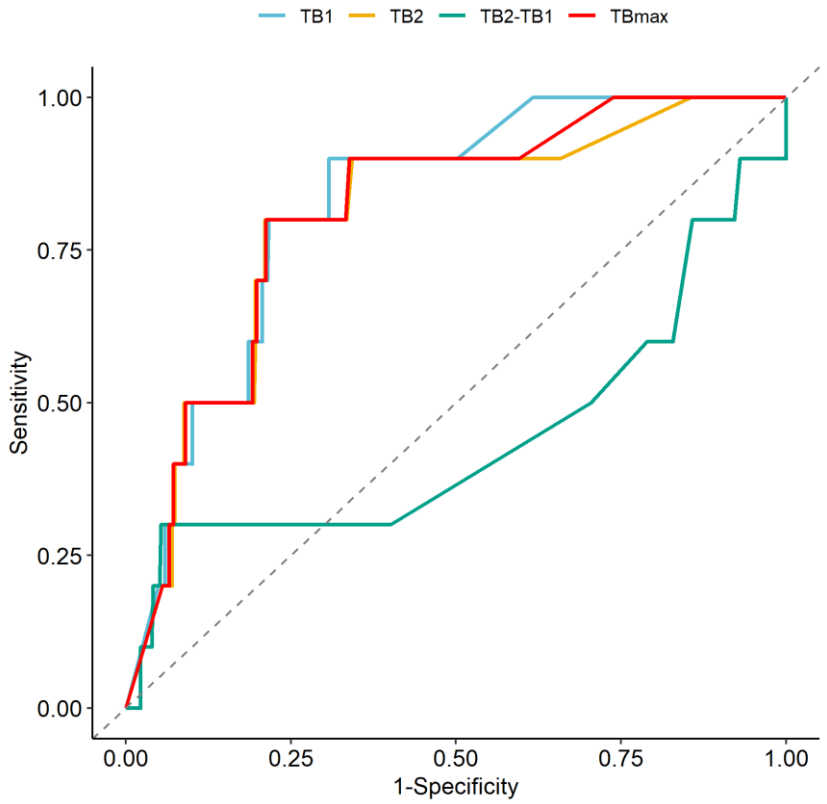


Figure 3-3: ROC curves showing performance of QFT-Plus for predicting incident TB.

TB antigen tubes (TB1 and TB2) are shown separately, along with difference between them (TB2-TB1) and maximum (TBmax). Adapted from [187].



3.4 Discussion

3.4.1 Summary of key findings

In this study, QFT-Plus performance appeared comparable to published evaluations of QFT-GIT and T-SPOT.TB, with an IRR of 10.1, 80% sensitivity for detection of incident TB, and an overall positive predictive value for incident TB of 5.7%[80]. Interferon-γ responses in the TB1 and TB2 tubes had strong correlation and agreement, and ROC curves therefore showed minimal difference between them for predicting incident TB. As a result, the calculated difference between TB1 and TB2 responses, as a proxy for the CD8-specific response, did not predict incident TB. My cohort study findings support and extend previous data showing similar performance of QFT-Plus to QFT-GIT in cross-sectional studies[188].

3.4.2 Strengths of the study

This is the first published evaluation, to our knowledge, of the prognostic value of QFT-Plus test. The prospective design allowed the collection of detailed clinical, demographic and laboratory data. Participants were recruited while attending routine contact-tracing services, ensuring the study population was representative of TB contacts. Moreover, follow-up was robust through linkage to national surveillance records using a validated matching algorithm[184], minimising risk of missing incident TB cases. In addition, QFT-Plus results measured in this study were not used to inform clinical management decisions, since this test was not routinely implemented in the NHS during the study period, reducing the risk of differential work-up bias.

3.4.3 Study limitations

Despite recruitment of 492 recent TB contacts with valid test results, the number of TB progressors was small, reflecting low progression risk even among contacts, meaning that the study is not powered to detect differences in prognostic value for incident TB between the TB1 and TB2 antigen tubes. Nonetheless, our findings that TB1 and TB2 responses had strong correlation and agreement, and that the difference between the tubes (representing the CD8-specific response) had no discrimination, together suggest that any incremental prognostic value of the second tube is likely to be limited. Future studies could

seek to evaluate the prognostic ability of QFT-Plus with greater precision using larger sample sizes and with direct head-to-head comparisons to QFT-GIT, T-SPOT.TB and TST among the same participants.

A second limitation was that the provision of preventative therapy to a subset of the QFT-Plus positive patients could have led to selection bias. In particular, preventative treatment may have been more likely to be offered to people at higher risk of developing TB, which may have led to underestimation of the prognostic ability of QFT-Plus. However, while the subgroup who received preventative therapy were younger than those who did not (reflecting national policy at the time of the study[74,172]), other measured characteristics were similar, suggesting that the magnitude of this bias is likely to be relatively small.

Third, the TB contacts included both pulmonary and extra-pulmonary index cases, reflecting national contact screening policy during the study period[172]. positive predictive value of QFT-Plus may be higher among populations including only pulmonary TB contacts, due to higher pre-test probability of incident TB disease. However, previous evaluations of the QFT-GIT and T-SPOT.TB have also included extra-pulmonary contacts, which allows the current study findings to be put into this context[80].

Serial testing (before and after exposure) was not performed, meaning that I was unable to assess QFT-Plus conversions over time, which may provide a more reliable measure of recent *M. tuberculosis* infection. QFT-Plus results were also missing or indeterminate for 91/623 patients (14.6%), in keeping with the proportion of missing results for other IGRAs in a recent evaluation[80]. However, these patients' characteristics were similar to the overall study population, suggesting that risk of subsequent selection bias was likely small.

I included both microbiologically confirmed and clinically diagnosed TB cases in my outcome definition, in keeping with previous IGRA evaluations[80,88,189–192]. The rationale for this is that extra-pulmonary TB, which occurs frequently among TB cases occurring in foreign-born people living in the UK[193], is often challenging to prove microbiologically. However, all TB cases diagnosed during the study received a full course of TB therapy, and none were de-notified. It is therefore

likely that these represented true TB cases, with low risk of outcome misclassification.

Finally, in this study, I did not seek to assess whether TB1 and TB2 could be considered as independent predictors in a prognostic prediction model. However, this is not how QFT Plus is currently implemented and any such approach is likely to be limited by the very strong correlation and agreement between the two tubes.

3.4.4 Conclusion

In summary, in this evaluation of the prognostic ability of QFT-Plus for incident TB, performance was comparable to other commercial IGRAs. Addition of the TB2 tube is likely to have limited, if any, incremental prognostic value.

3.5 Contribution statement

This primary data for this chapter were collected in the 'Next Generation Tests for Latent Tuberculosis Infection' study, led by Professor Ibrahim Abubakar. I led the analyses presented in this chapter from protocol development to completion and dissemination. I also coordinated data linkage to the UK Enhanced TB Surveillance system, in order to identify incident TB cases during follow-up.

3.6 Outputs relating to this chapter

This study is published in *Annals of the American Thoracic Society*:

Gupta RK, Kunst H, Lipman M, Noursadeghi M, Jackson C, Southern J, Imran A, Lozewicz S, Abubakar I (2020). Evaluation of QuantiFERON-TB Gold Plus for predicting incident tuberculosis among recent contacts: a prospective cohort study. **Annals ATS.** https://doi.org/10.1513/AnnalsATS.201905-407RL

4 Objective 3: Development and validation of a personalised risk predictor for incident TB in settings aiming towards pre-elimination: an individual participant data meta-analysis

4.1 Introduction

The risk of TB among individuals with a clinical diagnosis of LTBI is highly variable between study populations, with incidence rates ranging from 0.3-84.5 per 1,000 person-years of follow-up[66,76]. Quoting the 5-15% lifetime estimate is therefore likely to be inaccurate for many people. Thus, improved risk-stratification is required to enable more precise delivery of preventative treatment to those most likely to benefit[16,194]. While the magnitude of the T cell response to *M. tuberculosis* is associated with incident TB risk, implementing higher diagnostic thresholds alone does not improve prediction on a population level due to a marked loss of sensitivity with this approach (Chapter 2)[176].

In this study, I first sought to characterise the population risk of TB among people tested for LTBI using an IPD-MA. In order to study progression from LTBI to TB disease with low risk of re-infection with *M. tuberculosis* during follow-up, I focused on settings with low transmission (defined as annual incidence ≤20/100,000 persons). I then aimed to develop and validate a directly data-driven personalised risk predictor for incident TB (PERISKOPE-TB) that combines a quantitative T cell response measure with key clinical covariates.

4.2 Methods

4.2.1 Systematic review and pooling of individual participant data

I conducted a systematic review and IPD-MA, in accordance with Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data standards[195], to investigate the risk of progression to TB disease among people tested for LTBI in low transmission settings (PROSPERO CRD42018115357). I searched Medline and Embase on 09/01/2019 for studies published 01/01/2002 to 31/12/2018 using comprehensive terms for ‘TB’, ‘IGRA’, ‘TST’, ‘latent TB’, and ‘predictive value’, without language restrictions (Table 4-1).

Table 4-1: Medline search strategy for systematic review and IPD-MA to investigate risk of progression to TB among people tested for LTBI.

1. exp TUBERCULOSIS/ or tuberculosis.mp. or exp MYCOBACTERIUM TUBERCULOSIS/ or tb.mp.
2. exp Interferon-gamma Release Tests/
3. interferon gamma release.mp.
4. igra.mp.
5. t-spot*.mp.
6. tspot*.mp.
7. quantiferon*.mp.
8. qft*.mp.
9. 2 or 3 or 4 or 5 or 6 or 7 or 8
10. exp TUBERCULIN TEST/
11. tuberculin skin test.mp.
12. tst.mp.
13. purified protein derivative.mp.
14. ppd.mp.
15. mantoux.mp.
16. 10 or 11 or 12 or 13 or 14 or 15
17. exp latent tuberculosis/
18. latent.mp.
19. 17 or 18
20. 9 or 16 or 19
21. exp "Predictive Value of Tests"/
22. predict*.mp.
23. prognos*.mp.
24. progress*.mp.
25. ((tb adj2 developed) or (tuberculosis adj2 developed)).mp.
26. ppv.mp.
27. npv.mp.
28. (incidence rate ratio or irr).mp.
29. 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28
30. 1 and 20 and 29
31. limit 30 to yr="2002 -2018"
32. limit 31 to "humans only (removes records about animals)"
33. remove duplicates from 32

Longitudinal studies that primarily aimed to assess the risk of progression to TB disease among individuals tested for LTBI and that were conducted in a low TB transmission setting were eligible for inclusion (Table 4-2).

Table 4-2: Eligibility criteria for systematic review and IPD-MA to investigate risk of progression to TB among people tested for LTBI.

Inclusion criteria
1. Primary objective is to assess of risk of progression to TB disease among individuals tested for LTBI.
2. Longitudinal study design (prospective or retrospective)
3. Minimum median duration of follow-up one year.
4. Conducted in a low-TB incidence setting (defined as annual incidence ≤20/100,000 persons) at midpoint of study.
5. Commercially available interferon-gamma release assay (IGRA) or the tuberculin skin test (TST) included as one of the diagnostic tests for LTBI.
6. Minimum individual level exposure variables recorded should include age, gender, indication for LTBI screening, result of LTBI screening test (positive or negative), and whether preventative therapy was provided.
7. Minimum individual level outcome variables recorded should include presence or absence of active TB during follow-up, date of TB diagnosis, date of follow-up censor.
8. Study participants should be recruited after 1st January 2001.
Exclusion criteria
1. Studies in which all patients with evidence of LTBI received preventative therapy.
2. Studies testing for LTBI with TST only that do not have data on prior BCG vaccination.
3. Studies limited to single contact investigations.
4. Studies without associated full-text publications.

Two independent reviewers conducted first and second screens for the systematic review, with arbitration by a third reviewer where required. All titles and abstracts identified by the search were included in the first screen; relevant articles were selected for the second screen, which included full text review. Following confirmation of study eligibility, corresponding authors of original studies were invited to contribute individual participant data via a data safe haven. Received data were mapped to a master variables list, and the integrity of the data were examined by comparing original reported results with re-analysed results using contributed data. Quality assessment was performed using a modified version of the Newcastle-Ottawa scale for cohort studies[196].

4.2.2 Definitions

Participants entered the cohort on the day of LTBI screening or diagnosis, and exited on the earliest of censor date (last date of follow-up), active TB diagnosis date, date of death, or date of loss to follow-up (where available). LTBI was defined as any positive LTBI test (TST or commercial IGRA), using TST thresholds as defined by the contributing study (a 10mm cut-off was used for studies that assessed multiple thresholds). Quantitative IGRA thresholds were calculated according to standard manufacturer guidelines. IGRAs included three generations of QuantiFERON TB assays (QFT-GIT, QuantiFERON Gold, QFT-Plus), which were assumed to be equivalent[187], and T-SPOT.TB. Microbiologically confirmed and/or clinically diagnosed TB cases were included, as per contributing study definitions. Death was not included as a competing risk due to the low risk of mortality among the study cohorts.

In the absence of a widely accepted temporal distinction between prevalent and incident disease, prevalent TB at the time of screening was arbitrarily defined as a TB diagnosis within 42 days of enrolment; these cases were omitted from the primary analysis. Alternative shorter and longer temporal definitions were tested as sensitivity analyses. Participants with missing outcomes or durations of follow-up were considered lost to follow-up and excluded. 'Preventative treatment' was defined as any LTBI treatment regimen recommended by the WHO[62]. All contributing studies included regimens consistent with this guidance; the effectiveness of each regimen was assumed to be equivalent[152].

4.2.3 Population-level analysis

4.2.3.1 One-stage IPD-MA: survival analysis

In a one-stage IPD-MA approach, I used flexible parametric survival models, with a random effect intercept by source study to account for between study heterogeneity, to examine population level cumulative TB incidence over two and five years. The rationale for using flexible parametric survival models was that, in contrast to Cox proportional hazards models, they enable estimation of baseline risk throughout follow-up[197]. Flexible parametric survival models also estimate baseline hazard functions more flexibly than Poisson models (which assume constant hazards) or Weibull models (which assume monotonic changes in

hazards over time) by using restricted cubic splines[171]. I examined cumulative incident TB risk stratified by LTBI screening result (positive *vs* negative) and provision of LTBI treatment (commenced *vs.* not commenced). I then further examined cumulative incident TB risk among untreated participants with LTBI, stratified by indication for screening (recent child contacts (<15 years) *vs* adult contacts *vs* migrants *vs* immunocompromised), by separately fitting random-effect flexible parametric survival models to each risk group. Child contacts were further stratified by age (<5 *vs.* 5 to 14 years).

4.2.3.2 Two-stage IPD-MA: incidence rates

I also calculated TB incidence rates (per 1,000 person-years) in a two-stage IPD-MA approach, stratified by LTBI screening result, provision of LTBI treatment, and indication for screening. The rationale for this analysis was to complement the one-stage IPD-MA survival approach through forest plot visualisation and meta-analysis of study-level incidence rates. In contrast to the one-stage survival analyses, rates were calculated separately for the 0-2 year and 2-5 year follow-up intervals, since risk of disease has been consistently reported to be higher in the initial two-year interval[16]. Pooled incidence rate estimates for each risk group and follow-up interval were derived using random intercept Poisson regression models, without continuity correction for studies with zero events, in the meta package in R[198]. I adopted this approach in preference to an inverse variance meta-analysis method (with continuity correction) since the latter is prone to biased estimates when outcomes are rare[199], as may be the case for incident TB.

4.2.4 Prediction model analysis

4.2.4.1 Variables of interest

I then developed and validated a personalised prediction model for incident TB, reported in accordance with transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidance[200]. For this analysis, I included studies that reported quantitative LTBI test results, proximity and infectiousness (based on sputum smear status) of index cases for contacts, and country of birth and time since entry for migrants, since I considered these variables fundamental *a priori*. Using this subset of the data, I examined the availability of a range of variables of interest, specified *a priori*, in the contributing

datasets to determine eligibility for inclusion as candidate predictors in the model (Table 4-3). This list of variables was based on the UK PREDICT data dictionary[80] and was considered as a starting point for exploration of data availability and definitions across studies.

Table 4-3: Pre-specified variables of interest as potential candidate predictors for prognostic model.

Category	Variables
Demographics	Age, gender
Exposure to M. tuberculosis	Country of birth, year of migration, presence/absence of recent contact with index case, contact proximity, time since exposure, index case site of disease, index case sputum smear status, index case drug susceptibility.
Clinical risk factors for progression	Body mass index, diabetes, transplant receipt, chronic renal failure, malignancy, anti-TNF therapy, corticosteroid exposure, silicosis, HIV status, blood CD4 count, smoking, problem drug use, alcohol excess, history of homelessness, imprisonment, BCG vaccination history, previous TB history
LTBI screening	IGRA / TST result (binary and quantitative), TST date and result, LTBI treatment acceptance and completion

I determined that the following predictors were available from a sufficient number of datasets for further evaluation: age, gender, quantitative LTBI test result, previous BCG vaccination, recent contact (including proximity and infectiousness of index case), migration from a high TB incidence setting, time since migration, solid organ or haematological transplant receipt, HIV status and TB preventative treatment commencement.

4.2.4.2 Variable transformations

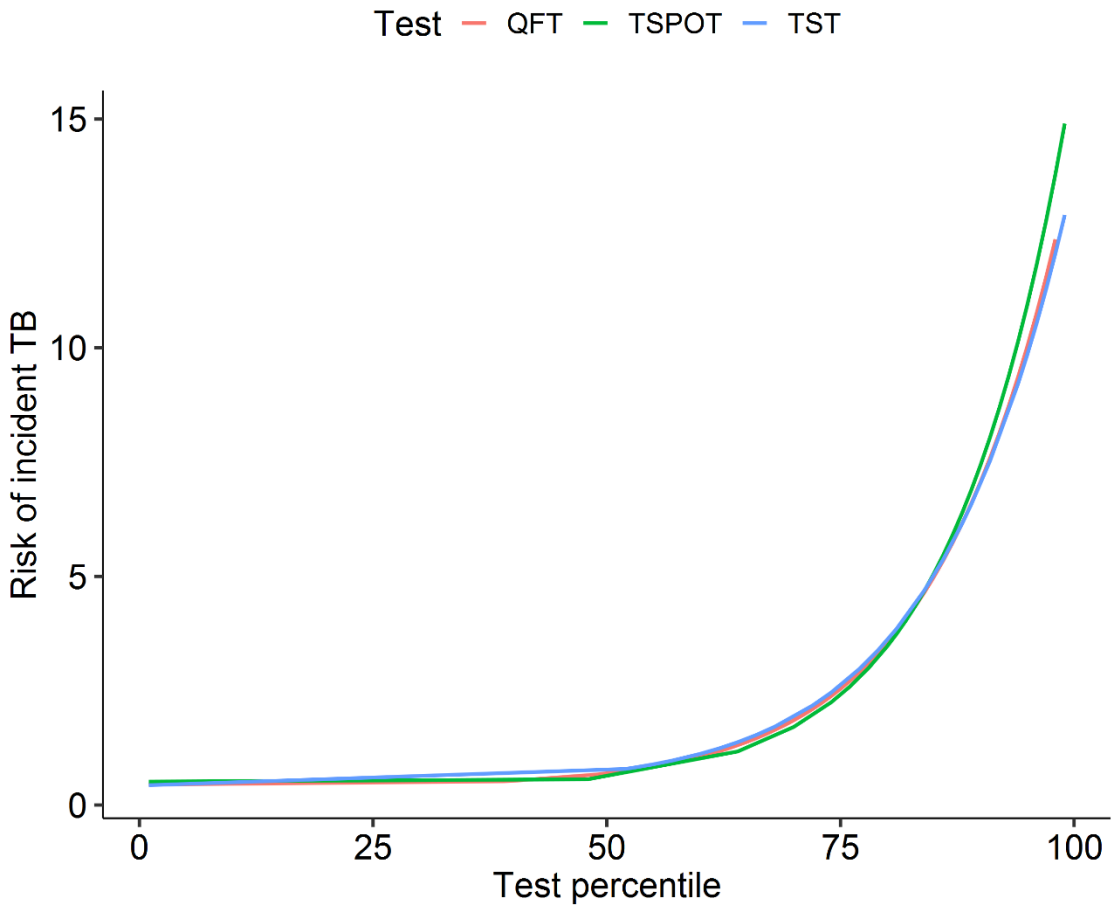
In Chapter 2, I demonstrated that higher quantitative TST, QFT-GIT and T-SPOT.TB results are associated with risk of incident TB[176]. However, each LTBI test is reported using different scales, and it has hitherto been unclear whether quantitative values of each test are equivalent with respect to incident TB risk.

To assess this further, I examined a sub-population of the entire cohort where all three tests were performed among the same participants in head-to-head studies. In view of their skewed distributions, I normalised quantitative results for the TST, QFT-GIT and T-SPOT.TB to a percentile scale using this head-to-head population, and examined the association between normalised result and incident

TB risk using Cox proportional hazards models with restricted cubic splines. Since TST cut-offs are frequently stratified by BCG vaccination and HIV status[201,202], I also examined whether these variables modified the association between quantitative TST measurement and incident TB risk in the head-to-head subpopulation. Since there was no evidence that including interaction terms for either BCG vaccination or HIV improved model fit (based on Akaike Information Criteria (AIC)), I used unadjusted TST measurements from this point. This analysis revealed that the normalised percentile results for each test (unadjusted TST, QFT-GIT and T-SPOT.TB) appeared to be associated with similar risk of incident TB (Figure 4-1).

Figure 4-1: Risk of incident TB with percentile normalised quantitative test results for the TST, QFT-GIT and T-SPOT.TB.

Modelled using Cox regression among head-to-head population among whom all three tests were performed (n=8,335; 158 TB cases; 3 studies). Material from [203].



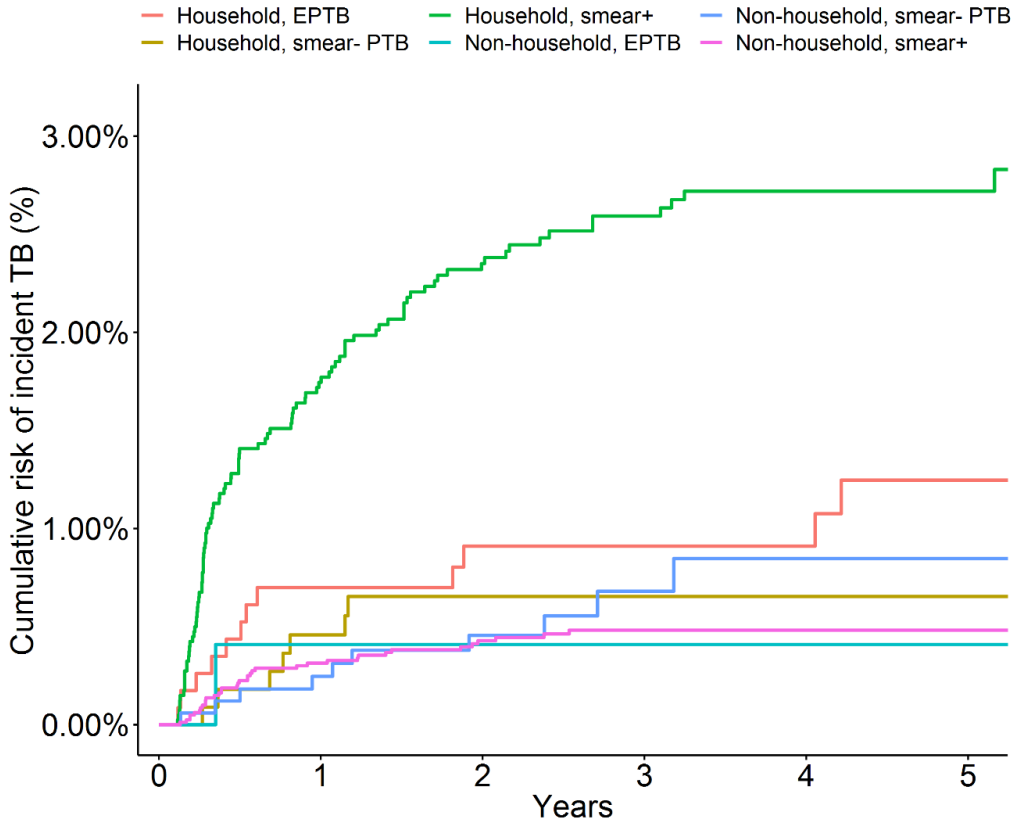
The LTBI tests implemented differed between contributing studies. From this point, all LTBI test results were therefore normalised to this percentile scale to

enable data harmonisation across studies, by transforming raw quantitative results to the relevant percentile using look-up tables derived from the head-to-head population (Supplementary Material; Chapter 8.1). Since most people evaluated for LTBI under routine programmatic conditions have a single test performed, I only included one test result per participant in the prediction model. I preferentially included tests where quantitative results were available. Where quantitative results were available for more than one test, I preferentially included the QuantiFERON result (since this was the most commonly used test in the dataset), followed by T-SPOT.TB, and then TST.

Recent TB contacts were categorised as either 'smear positive and household' or 'other' contacts, since there was no evidence of separation of risk among additional subgroups of the 'other' contacts stratum during exploratory univariable analyses (Figure 4-2).

Figure 4-2: Cumulative TB incidence, stratified by proximity and infectiousness of index cases among contacts.

PTB = pulmonary TB index case; EPTB = extra-pulmonary TB index case. Material from [203]. Shows by years since LTBI testing. Confidence intervals not shown to aid visual clarity.



Since I considered migration from a high TB burden country (defined as annual TB incidence ≥100/100,000 persons at the year of migration) to be a proxy for prior TB exposure, I included this in a composite 'TB exposure' variable, which included four mutually-exclusive levels: household contact of smear-positive index case; 'other' contact; migrant from country with high TB incidence, without recent contact; and no exposure. There was no evidence of separation of incident TB risk when stratified by TB incidence in country of birth above the binary country of birth threshold (TB incidence ≥100/100,000 persons) among migrants, or when stratified by country of birth among recent contacts (Figure 4-3, Figure 4-4).

Figure 4-3: Cumulative TB incidence, stratified by incidence in country of birth among migrants from high TB burden countries.

P value represents Log-rank test. Material from [203].

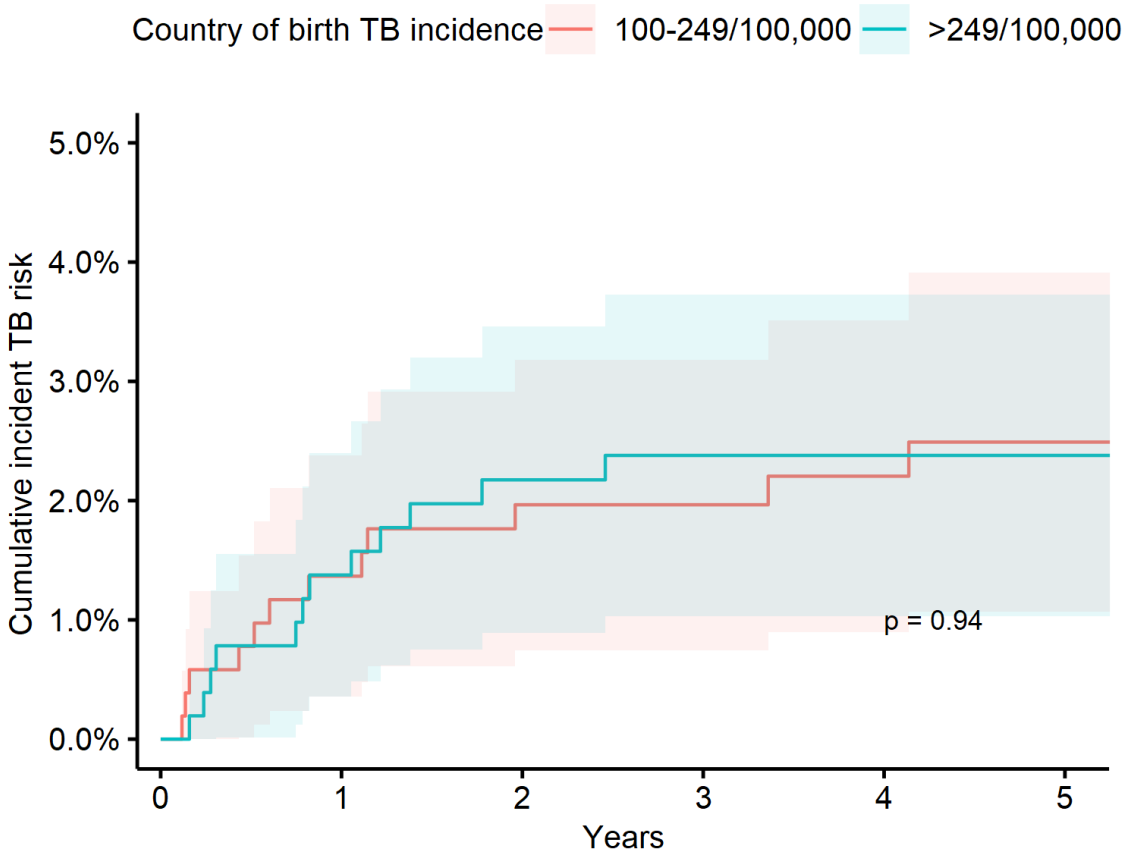
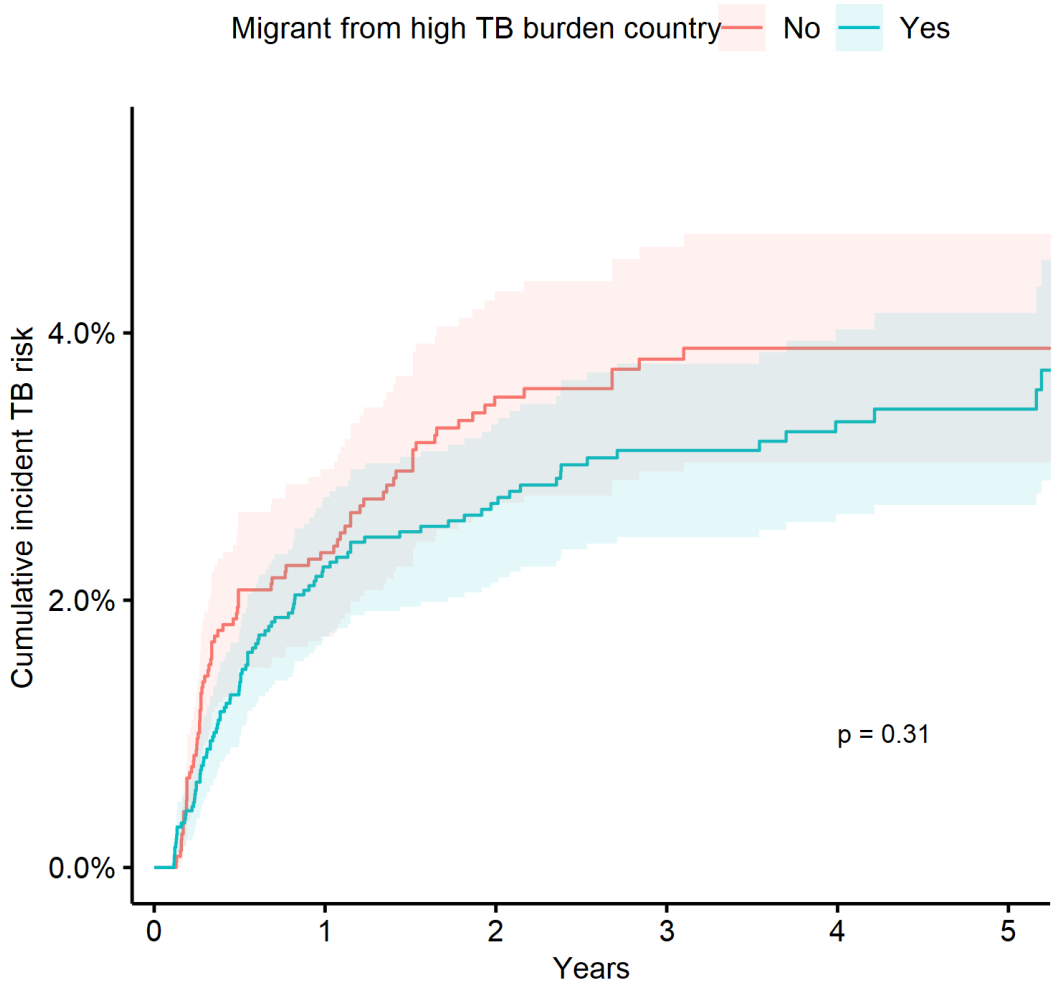


Figure 4-4: Cumulative TB incidence, stratified by country of birth among recent contacts.

P value represents Log-rank test. Material from [203].



Age and normalised test result variables were modelled using restricted cubic splines (using a default of five knots placed at recommended intervals[175]) to account for their non-linear associations with incident TB, as described in Chapter 2.2.3. These transformations were pre-specified and were done prior to multiple imputation. A data dictionary and summary of all candidate predictor variables considered for inclusion in the prognostic model is provided in Table 4-4.

Table 4-4: Data dictionary of candidate predictors considered for inclusion in prognostic model.

Variable	Description	Levels
Age	Age in years	Numeric
Sex	Gender	Male, Female

BCG	Previously received BCG vaccination (self-report, presence of scar or documented evidence)	Yes, No
Percentile LTBI test result	Normalised LTBI test result	Numeric
Exposure	Composite TB exposure variable	Household contact of smear positive index case 'Other' contacts Migrant from high TB burden country with no contact No exposure
Months since migration	Months since migration for migrants with no TB contact	Numeric
HIV	HIV status	Positive, Negative
Transplant	Previously received haematological or solid organ transplant (assumed negative when missing)	Yes, No
LTBI treatment	Commenced preventative treatment	Yes, No

4.2.4.3 Multiple imputation

Age, sex, time since migration for migrants, quantitative LTBI test results and LTBI treatment commencement were available for >95% of participants. BCG vaccination history was systematically missing for 5/15 studies and HIV status was systematically missing from 6/15 studies, all conducted among recent TB contacts. A full summary of missingness of candidate predictor variables is provided in the Supplementary Material (Chapter 8.1).

I performed multi-level multiple imputation to account for sporadically and systematically missing data, while respecting clustering by source study, in accordance with recent guidance[204] using the micemd package in R[205]. This approach assumes missingness at random, whereby the pattern of missingness can be explained by the observed data[206]. I used predictive mean matching for continuous variables, due to their skewed distributions. I included all variables (including transformations) assessed in the downstream prediction model in the imputation model, along with auxiliary variables, to ensure compatibility of the imputation and analysis models. The incident TB outcome and Nelson-Aalen estimator variables were also included in the imputation model.

Multi-level imputation was done separately for contacts and non-contacts due to expected heterogeneity between these groups. I generated ten multiply imputed

datasets, with 25 between-imputation iterations. Model convergence was assessed by visually examining plots of imputed parameters against iteration number. All downstream analyses were done in each of the ten imputed datasets; model coefficients and standard errors were combined using Rubin's rules[207]. No imputation was done for participants missing binary LTBI test results, or for those with missing outcomes; these individuals were excluded. For recent TB contacts or people screened due to HIV infection with missing data on transplant status, this was assumed negative due to the very low prevalence of transplant receipt when observed among these risk groups (<0.5%).

4.2.4.4 Variable selection and final model development

I then performed backward selection of the nine candidate predictors in each of the imputed datasets, using AIC. This process starts with a full model including all candidate variables. Predictors are then iteratively removed, and model fit is re-assessed based on AIC; predictors which do not lead to an increase in AIC when removed are not included in the selected model. I included predictors that were retained in more than 50% of the imputed datasets in the final model.

T cell responses to *M. tuberculosis* may be impaired in the context of immunosuppression (including among people with HIV or transplant recipients)[66]. I therefore also tested whether there was a significant interaction between HIV or transplant and the normalised percentile test result variable, in order to assess whether the association between the quantitative test result and incident TB risk varied according to HIV or transplant status. This analysis showed no evidence of effect modification, based on AIC, thus these interaction terms were not included in the final model.

Similar to the population-level analysis, I used flexible parametric survival models for the prediction modelling in order to facilitate estimation of baseline risk throughout the duration of follow-up[208], using the rstpm2 package in R[209]. I examined a range of degrees of freedom for the baseline hazard, using proportional hazards and odds scales, and selected the final model parameters based on the lowest AIC across the imputed datasets. Visual inspection of survival curves suggested non-proportional hazards for the composite exposure category; I therefore assessed whether including this variable as a time-varying

covariate (by including an interaction between the composite exposure covariate of interest and time) improved model fit[210]. Since the AIC for the time-varying covariate model was lower across all imputed datasets, this time-varying covariate approach was used for the final model.

4.2.4.5 Internal-external cross-validation

Following selection of the final model, I used the internal-external cross-validation (IECV) framework for model validation[211]. In this process, one entire contributing study dataset is iteratively discarded from the model training set and used for external validation. This process is repeated until each dataset has been used once for validation. Compared to an approach of splitting the full dataset into development and validation cohorts, this framework has the advantage of utilising all contributing data for final model development, while facilitating concurrent assessment of between-study heterogeneity and generalisability across contributing studies[211,212].

The primary outcome for IECV was two-year risk of incident TB. I included datasets with a minimum of five incident TB cases, and where participants had been included regardless of LTBI test result, as the primary validation sets. During validation, I assessed discrimination (how well the predictions differentiate incident TB cases from non-progressors), and calibration (how well predicted risk matched observed risk)[197].

I assessed model discrimination using the C-statistic, equivalent to the AUROC (as discussed in Chapter 2.2.3), for two-year TB risk. Model calibration was assessed by visually examining calibration plots of predicted risk *vs.* Kaplan Meier estimated observed two-year risk in quintiles, and using the calibration slope and calibration-in-the-large statistics[177]. Calibration slopes >1 suggest under-fitting (predictions are not varied enough), while slopes <1 indicate over-fitting (predictions are too extreme). Slopes were calculated by fitting survival models with the model linear predictor as the sole predictor; the calculated coefficient for the linear predictor provides the calibration slope. Calibration-in-the-large indicates whether predictions are systematically too low (calibration-in-the-large >0) or too high (calibration-in-the-large <0). I calculated calibration-in-the-large for each validation set by fixing all model coefficients from model

development (including the baseline hazard terms), and re-estimating the intercept. The difference between the development model and recalculated validation model intercepts provided the calibration-in-the-large statistic[213].

4.2.4.6 Pooling of IECV parameters and random-effects meta-analysis

IECV was performed on each imputed dataset. Validation set C-statistics, calibration slopes and calibration-in-the-large metrics were pooled for each study across imputations using Rubin's rules[207]. I then meta-analysed these metrics across validation studies with random-effects, using logit-transformed C-statistics as previously recommended[214], to derive pooled discrimination and calibration estimates. The IECV validation sets were also pooled, with averaging of the predicted two-year risk of TB for each individual in the validation sets across imputations, for downstream decision curve analyses as described below.

4.2.4.7 Decision curve analysis

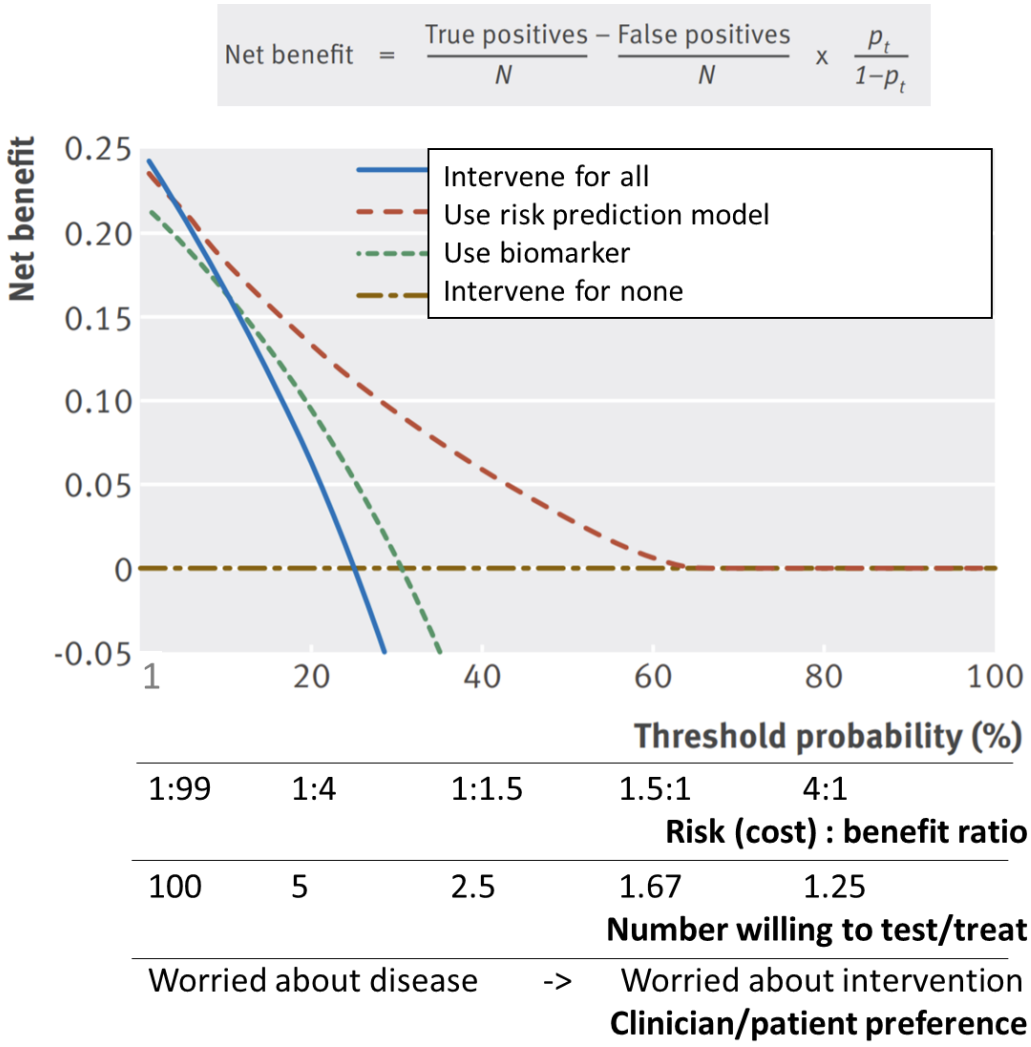
Decision curve analysis extends the evaluation of prediction model validation parameters by assessing the potential clinical utility of a model to inform medical decision-making in practice[215–217]. The key metric in decision curve analysis is 'net benefit'. This measure quantifies the proportion of true positive cases detected minus the proportion of false positives, with weighting by the 'threshold probability'[215].

Net benefit, calculated as per the equation shown in Figure 4-5, is plotted across a range of 'threshold probabilities'. The threshold probability reflects the percentage cut point for the prediction model, above which an intervention (such as treatment initiation) is recommended. The threshold probability also reflects the risk (or cost) / benefit ratio of initiating treatment. For example, a threshold probability of 1% corresponds to a risk: benefit ratio of 1:99 (whereby the benefit of the intervention for true positive patients is considered to be 99 times greater than the risk for false positive patients). Another way to interpret this is as a number willing to treat. For example, a threshold probability of 1% reflects a number willing to treat of 100. Overall, clinicians and patients who are more concerned about a disease endpoint are more likely to have lower threshold probabilities, while those more concerned about adverse effects of medical interventions (such as drug side-effects) will have higher threshold probabilities.

These complementary interpretations can be illustrated by relabelling the threshold probability axis, as in the Figure 4-5 schematic.

Figure 4-5: Decision curve analysis schematic showing net benefit equation and comparison of multiple decision curves.

Adapted from Vickers et al (2016) [217]. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests BMJ; 352 :i6 https://doi.org/10.1136/bmj.i6 under CC BY-NC 3.0 license. p_t = threshold probability; N = total sample size.



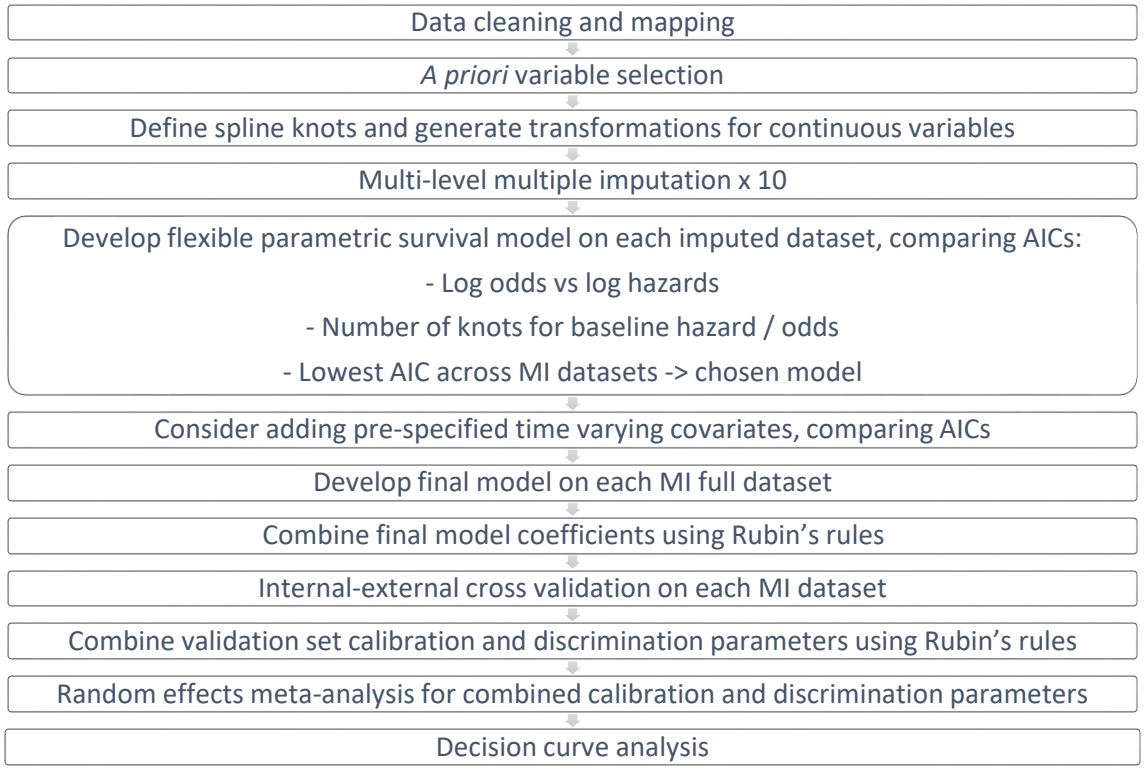
Since threshold probabilities can vary between clinicians and patients according to their preference, net benefit can be calculated across a range of clinically relevant threshold probabilities and plotted as decision curves. Figure 4-5 demonstrates example decision curves for a variety of strategies including: intervening for all people; using a risk prediction model to guide decisions; using a single biomarker; or intervening for none. In general, the best approach is the

one that leads to highest net benefit across a clinically plausible range of threshold probabilities.

In my analysis, I compared the approaches of using the final prognostic model to guide preventative treatment decisions to the default strategies of treating either all or no patients with a positive LTBI test across a range of threshold probabilities of 0-25% (reflecting a number willing to treat range of ≤4 to prevent one incident TB case). I analysed net benefit using the stdca function from the ddsjoberg/dca package in R[218], using the stacked validation sets of untreated participants with positive LTBI tests from IECV (to ensure that each individual for whom a prediction was generated had not been included in the model development set used to derive that prediction).

The full prediction modelling workflow in summarised in Figure 4-6.

Figure 4-6: Flowchart summarising prediction modelling analysis pipeline.



4.2.5 Sensitivity analyses

Sensitivity analyses included:

- Recalculating prediction model parameters using:

(a) alternative definitions of prevalent TB (ranging from diagnosis within 0 to 90 days of recruitment);
(b) a complete case approach (for all variables except for HIV status, which was assuming negative where this was missing);
(c) exclusion of participants who received preventative treatment.

Parameters for each of these models were compared with the primary model (without time-varying covariates to facilitate interpretation).

- Examining IECV discrimination parameters for validation datasets when:
  (a) restricted to participants with positive binary LTBI tests;
  (b) excluding those who received preventative treatment;
  (c) imputing an average quantitative positive or negative LTBI test result (based on the medians among the study population), according on the binary result. This was done to assess model performance in situations where the quantitative test result is not available.

4.2.6 Ethics

This study involved analyses of fully depersonalised data from previously published cohort studies, with data pooling via a safe haven. Ethical approvals for sharing of data were sought and obtained by contributors of individual participant data, where required.

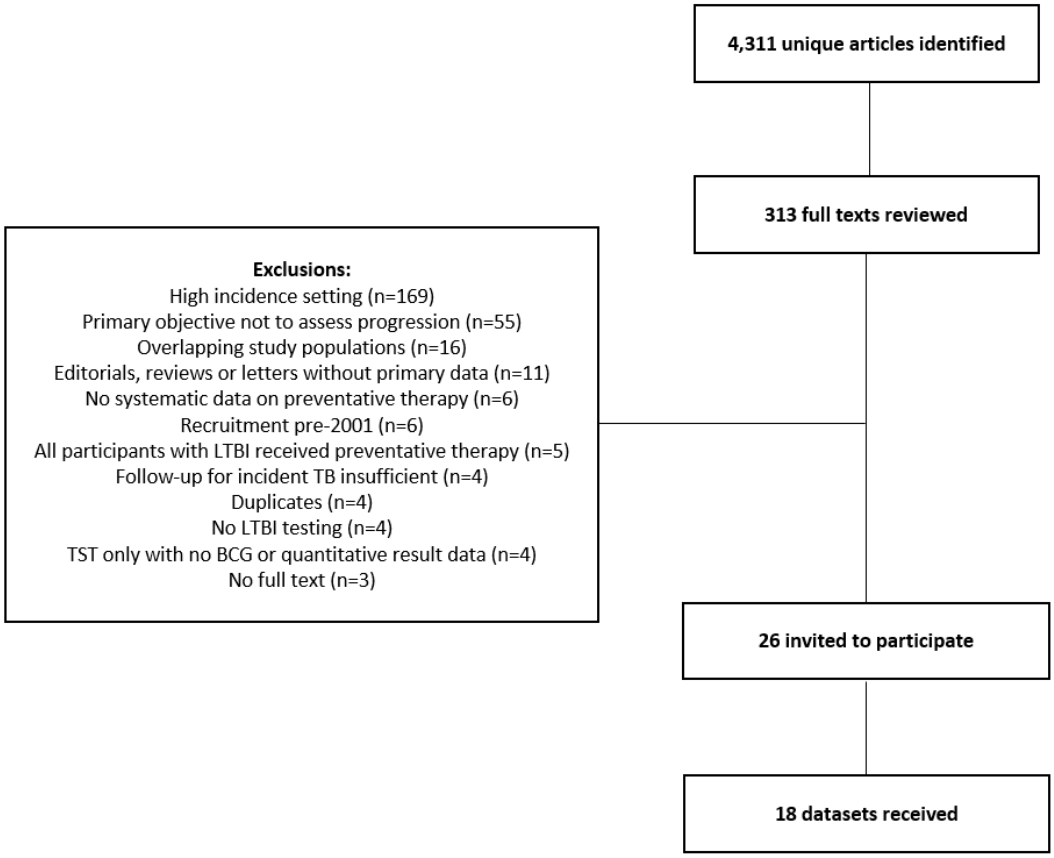
4.3 Results

4.3.1 Systematic review

I identified 26 studies that aimed to assess the risk of progression to TB disease among individuals tested for LTBI in low TB transmission settings; corresponding authors of these studies were invited to contribute individual level data (Figure 4-7). Of these, I received 18 individual level datasets, including participants recruited in 20 countries.

Figure 4-7: Flow chart for systematic review and IPD-MA of studies examining risk of TB among people tested for LTBI in low incidence settings.

Material from [203].



The pooled dataset included a total of 82,360 individual records, of whom 51,697 had evidence of LTBI and 826 were diagnosed with TB. Of the received data, 80,468 participants (including 803 TB cases) had sufficient data for inclusion in the primary analysis (Figure 4-8). The characteristics of the included study datasets are summarised in Table 4-5. Additional characteristics are shown in the Supplementary Material (Chapter 8.1). Characteristics of the eight eligible studies for which individual participant data were not obtained were similar to those included in the analysis (Table 4-6).

Eight studies recruited adults only; the remainder recruited both adults and children. The target population was recent TB contacts in nine studies[87,88,187,189,190,192,219–221], people living with HIV in two studies[222,223], mixed immunocompromised groups in two studies[224,225], transplant recipients in one study[226], mixed population screening in two studies[191,227], recent migrants in one study[228], and a combination of recent contacts and migrants in one study[80]. Median follow-up of all participants was 3.7 years (IQR 2.1-5.3). All contributing studies reported baseline assessments for prevalent TB through routine clinical evaluations, and all included culture-confirmed and clinically diagnosed TB cases in their case definitions. Four studies had a proportion of participants lost to follow up >5%[88,192,223,224]; baseline characteristics of those lost to follow-up were similar to those followed-up in each of these studies (Supplementary Material; Chapter 8.1). All contributing studies achieved quality assessment scores of 6 or 7/7 (full quality assessments shown in Supplementary Material; Chapter 8.1). Four studies did not have data on microbiological confirmation. Where available, culture confirmation was reported for 69.3% of TB cases.

Figure 4-8: Flow chart showing inclusion of participants in the population-level and prediction modelling analyses.

Material from [203].

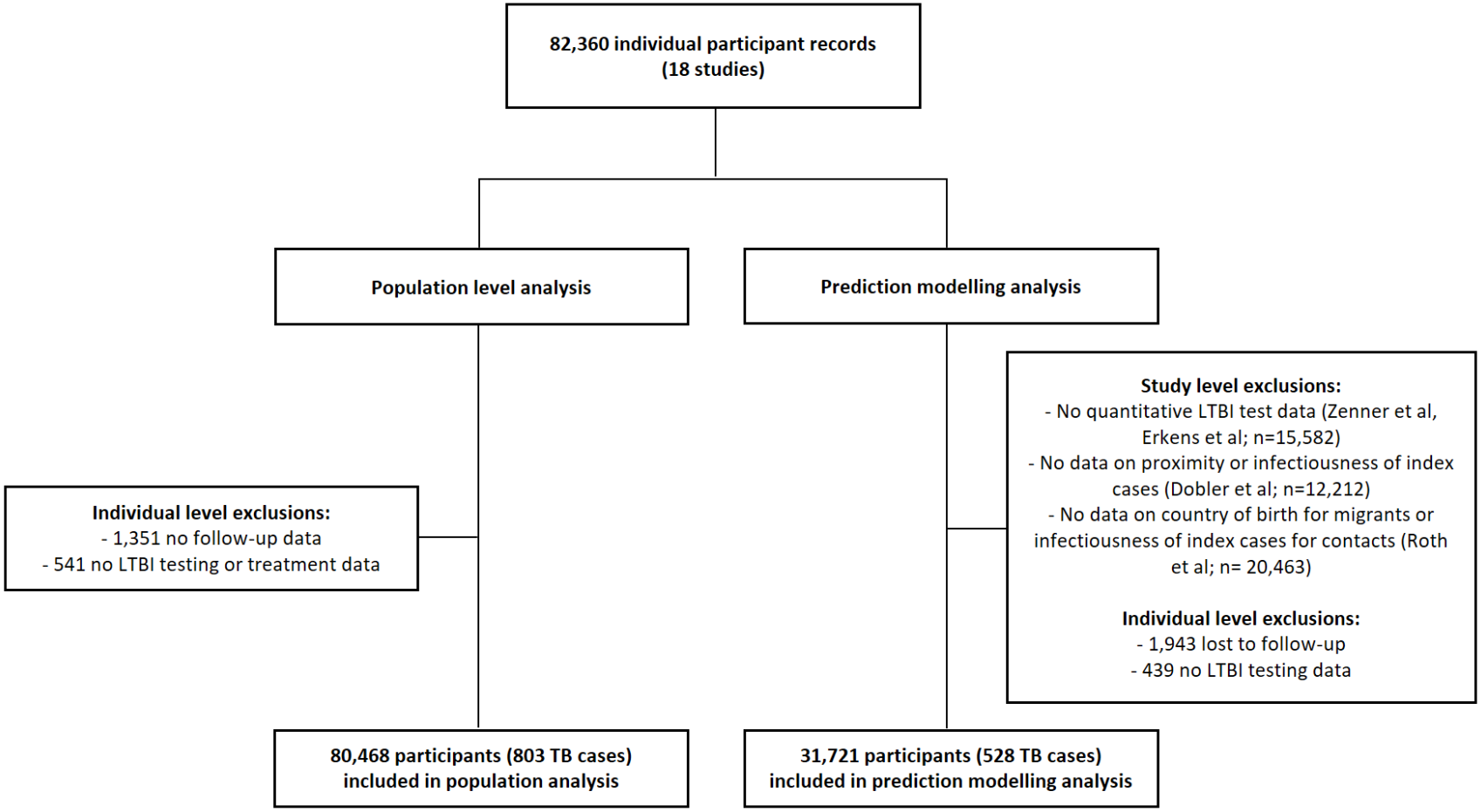


Table 4-5: Characteristics of contributing studies included in Chapter 4 IPD-MA.

Authors	Publication Year	Country	N (total)	Adults / children	Population	Follow-up years (median (IQR))	TB cases	Loss to follow-up	Included in prediction modelling	NOS^
Abubakar et al.[80]	2018	UK	10,045	Adults	Contacts & migrants	4.7 (3.7-5.5)	147	10 (0.1%)	Yes	7/7
Aichelburg et al.[222]	2009	Austria	830	Adults	PLHIV	1.2 (0.7-1.4)	11	25 (3%)	Yes	7/7
Altet et al.[219]	2015	Spain	1,339	Adults & children	Contacts	4 (4-4)	95	0 (0%)	Yes	7/7
Diel et al.[192]	2011	Germany	1,414	Adults & children	Contacts	3.5 (2.5-4.2)	19	381 (26.9%)	Yes	7/7
Dobler & Marks[220]	2013	Australia	12,212	Adults & children	Contacts	4.2 (2-6.9)	94	351 (2.9%)	No*	7/7
Doyle et al.[223]	2014	Australia	919	Adults	PLHIV	2.9 (1.7-3.6)	2	47 (5.1%)	Yes	7/7
Erkens et al.[191]	2016	Netherlands	14,241	Adults & children	Mixed population screening	5.5 (3-7.4)	134	NA	No*	6/6
Geis et al.[87]	2013	Germany	1,283	Adults & children	Contacts	0.8 (0.4-1.1)	33	62 (4.8%)	Yes	6/6
Gupta et al.[187]	2020	UK	623	Adults	Contacts	1.9 (1.6-2.2)	13	0 (0%)	Yes	7/7
Haldar et al.[190]	2013	UK	1,411	Adults & children	Contacts	1.9 (1.3-2.4)	37	30 (2.1%)	Yes	7/7
Lange et al.[224]	2012	Germany	456	Adults	Immunocompromised	2.8 (2-3.1)	1	42 (9.2%)	Yes	7/7
Munoz et al.[226]	2015	Spain	76	Adults	Transplant recipients	4.3 (3.6-4.8)	2	0 (0%)	Yes	7/7
Roth et al.[227]	2017	Canada	22,949	Adults & children	Mixed population screening	3 (1.8-4.3)	58	NA	Subset*	6/6
Sester et al.[225]	2014	Multiple European countries	1,464	Adults	Immunocompromised	2.7 (1.5-3.5)	11	7 (0.5%)	Yes	7/7
Sloot et al.[189]	2014	Netherlands	5,895	Adults & children	Contacts	5.9 (3.6-7.7)	81	NA	Yes	7/7
Yoshiyama et al.[221]	2015	Japan	625	Adults & children	Contacts	1.8 (1.4-2)	12	0 (0%)	Yes	6/7
Zellweger et al.[88]	2015	Multiple European countries	5,237	Adults & children	Contacts	2.6 (1.9-3.5)	55	1339 (25.6%)	Yes	7/7
Zenner et al.[228]	2017	UK	1,341	Adults	Migrants	3.7 (3-4.8)	21	NA	No*	7/7
Total			**82,360**			**3.7 (2.1-5.3)**	**826**	**2294 (2.8%)**		

**Not included in prediction modelling due to lack of data on proximity or infectiousness of index cases[220], or absent quantitative LTBI test data[191,228]. A subset of the dataset was included in the prediction model for the Roth et al study[227]; contacts and migrants were excluded due to no data being available on country of birth or infectiousness of index cases, respectively.*

^Modified version of the Newcastle Ottawa Scale for cohort studies

Table 4-6: Characteristics of eligible studies that did not contribute individual participant data.

Authors	Publication Year	Country	N (total)	Adults / children	Population	Follow-up (years)	Incident TB cases	Reason for exclusion
Bergot et al.[229]	2012	France	687	Adults & children	Contacts	Mean 2.8 (SD 0.9)	2	Unable to contact authors
Blount et al.[230]	2016	USA	1,152	Adults & children	Migrants with suspected TB, previous TB, TB contact, or LTBI	Median 6.7 (IQR 5.1 - 8.2)	7	Pending dataset receipt at database closure
Hand et al.[231]	2018	USA	148	Adults	Transplant recipients	Median 2.5	4	Unable to contact authors
Harstad et al.[232]	2010	Norway	823	Adults	Migrants	Range 1.9 - 2.7	9	Data no longer available
Kik et al.[233]	2010	Netherlands	339	Adults	Contacts who are migrants (TST ≥5mm)	Median 1.8 (IQR 1.3 - 2.0)	6	Pending dataset receipt at database closure
Pullar et al.[234]	2014	Norway	298	Adults	PLHIV	2	0	Unable to contact authors
Reichler et al.[235]	2018	USA and Canada	4,490	Adults	Contacts	2	77	Governance approvals not possible
Winje et al.*[85]	2018	Norway	44,875	Adults & children	Mixed population screening	Median 3.6 (IQR 2.4 - 5.1)	257	Pending governance approvals at database closure

SD = standard deviation; IQR = interquartile range

**Only a subset of the cohort were eligible for inclusion, since indication for LTBI screening was not available for most participants.*

4.3.2 Population-level analysis

In the pooled dataset, the two-year cumulative risk of incident TB was estimated as 4.0% (95% CI 2.6-6.3) among people with LTBI who did not receive preventative therapy, 0.7% (0.4-1.3) in people with LTBI who commenced preventative therapy and 0.2% (0.1-0.4) in people without LTBI (Figure 4-9; Table 4-7). The corresponding five-year risk of incident TB among these groups was 5.4% (3.5-8.5), 1.1% (0.6-2.0) and 0.3% (0.2-0.5), respectively.

Among untreated people with LTBI, two-year risk of incident TB was 14.6% (7.5-27.4) among recent child (<15 years) contacts, 3.7% (2.3-6.0) among adult contacts, 4.1% (1.3-12.0) among migrants, and 2.4% (0.8-6.8) among people screened due to immunocompromise (without an index exposure). Corresponding five-year risk was 15.6% (8.0-29.2) among recent child contacts, 4.8% (3.0-7.7) among adult contacts, 5.0% (1.6-14.5) among migrants, and 4.8% (1.5-14.3) among people screened due to immunocompromise.

Among recent child contacts, risk was markedly higher among those aged <5 years, compared to those aged 5-14 years (two-year risk 26.0% (9.4-60.1) *vs.* 12.4% (5.7-25.6); Figure 4-9), with 85.4% and 93.7% of cumulative risk being accrued in the first one and two years of follow-up, respectively. Among adult contacts and migrants, the annual risk also declined markedly with time. Of the cumulative five-year risk, 58.0% and 77.5% was accrued in the first one and two years of follow-up for adult contacts, with corresponding values among migrants of 66.4% and 81.6%. There was a more even distribution of risk during follow-up in the immunocompromised group.

Figure 4-9: Population-level cumulative TB incidence during follow-up.

Cumulative incidence curves are derived from flexible parametric survival models fitted to each risk group with random-effect intercepts by source study. PT = preventative treatment. Material from [203].

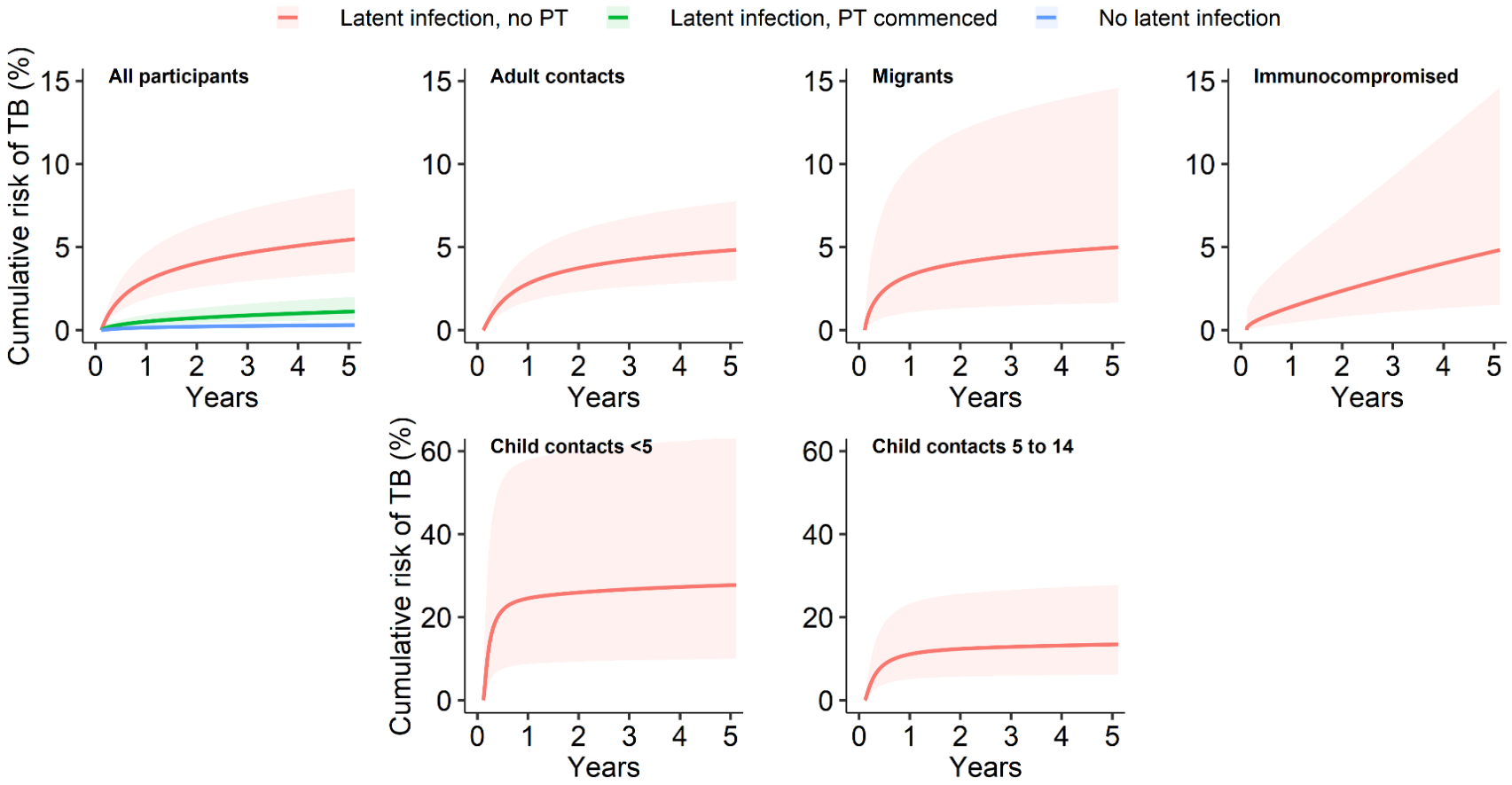


Table 4-7: Cumulative risk of incident tuberculosis during follow-up.

Cumulative incidence estimates are derived from flexible parametric survival models fitted to each risk group with random-effect intercepts by source study. PT = preventative treatment.

Group	**n**	**TB**	**Studies**	**Cumulative risk of incident TB by follow-up year (shown as risk (95% CI); cumulative proportion)**				
				1	*2*	*3*	*4*	*5*
No latent infection	29215	71	15	0.2 (0.1-0.3); 52.7%	0.2 (0.1-0.4); 71.9%	0.3 (0.2-0.4); 83.7%	0.3 (0.2-0.5); 92.6%	0.3 (0.2-0.5)
Latent infection, PT commenced	19032	125	15	0.5 (0.3-0.9); 46.3%	0.7 (0.4-1.3); 66.1%	0.9 (0.5-1.6); 79.9%	1 (0.6-1.8); 90.7%	1.1 (0.6-2)
Latent infection, no PT	32221	607	18	3 (1.9-4.7); 54.7%	4 (2.6-6.3); 74%	4.6 (3-7.2); 85.3%	5.1 (3.2-7.9); 93.5%	5.4 (3.5-8.5)
Adult contacts	10529	366	11	2.8 (1.7-4.5); 58.2%	3.7 (2.3-6); 77.6%	4.2 (2.6-6.8); 87.9%	4.6 (2.8-7.3); 94.7%	4.8 (3-7.7)
All child contacts	551	119	9	13.3 (6.8-25.2); 85.4%	14.6 (7.5-27.4); 93.7%	15.1 (7.8-28.2); 96.6%	15.4 (7.9-28.7); 98.5%	15.6 (8-29.1)
Child contacts <5	110	38	7	24.6 (8.8-57.9); 88.6%	26 (9.4-60.1); 93.7%	26.7 (9.7-61.4); 96.5%	27.3 (9.9-62.4); 98.4%	27.7 (10-63.2)
Child contacts 5 to 14	441	81	8	11.1 (5.1-23.2); 82.5%	12.4 (5.7-25.6); 92.2%	12.9 (6-26.6); 95.8%	13.2 (6.1-27.2); 98.2%	13.4 (6.2-27.7)
Migrants	1719	48	2	3.3 (1.1-9.9); 66.4%	4.1 (1.3-12); 81.6%	4.5 (1.5-13.1); 89.8%	4.7 (1.6-13.9); 95.5%	5 (1.6-14.5)
Immunocompromised	459	18	6	1.4 (0.4-4.4); 29.6%	2.4 (0.8-6.8); 49.9%	3.2 (1.1-9.3); 68%	4 (1.3-11.8); 84.5%	4.7 (1.5-14.3)

TB incidence rates in years 0-2 and 2-5 of follow-up among people with untreated LTBI, stratified by risk group for LTBI screening, are shown in Figure 4-10 and Figure 4-11. Within each of the risk groups assessed, incidence rates among untreated people with LTBI were markedly higher in the 0-2 year interval, compared to the 2-5 year interval, but were highly heterogeneous across studies (I^2 statistics, representing the proportion of variance that is considered due to between-study heterogeneity, ranged from 54-91% for incidence rates during the 0-2 year interval among untreated people with LTBI, when stratified by indication for screening). These findings suggest highly variable TB risk among people with LTBI, even within risk groups.

Figure 4-10: TB incidence rates by source study among people with untreated LTBI, during 0-2 year interval following LTBI testing.

Pooled estimates calculated using random effect Poisson models, without continuity correction for studies with zero events. Material from [203].

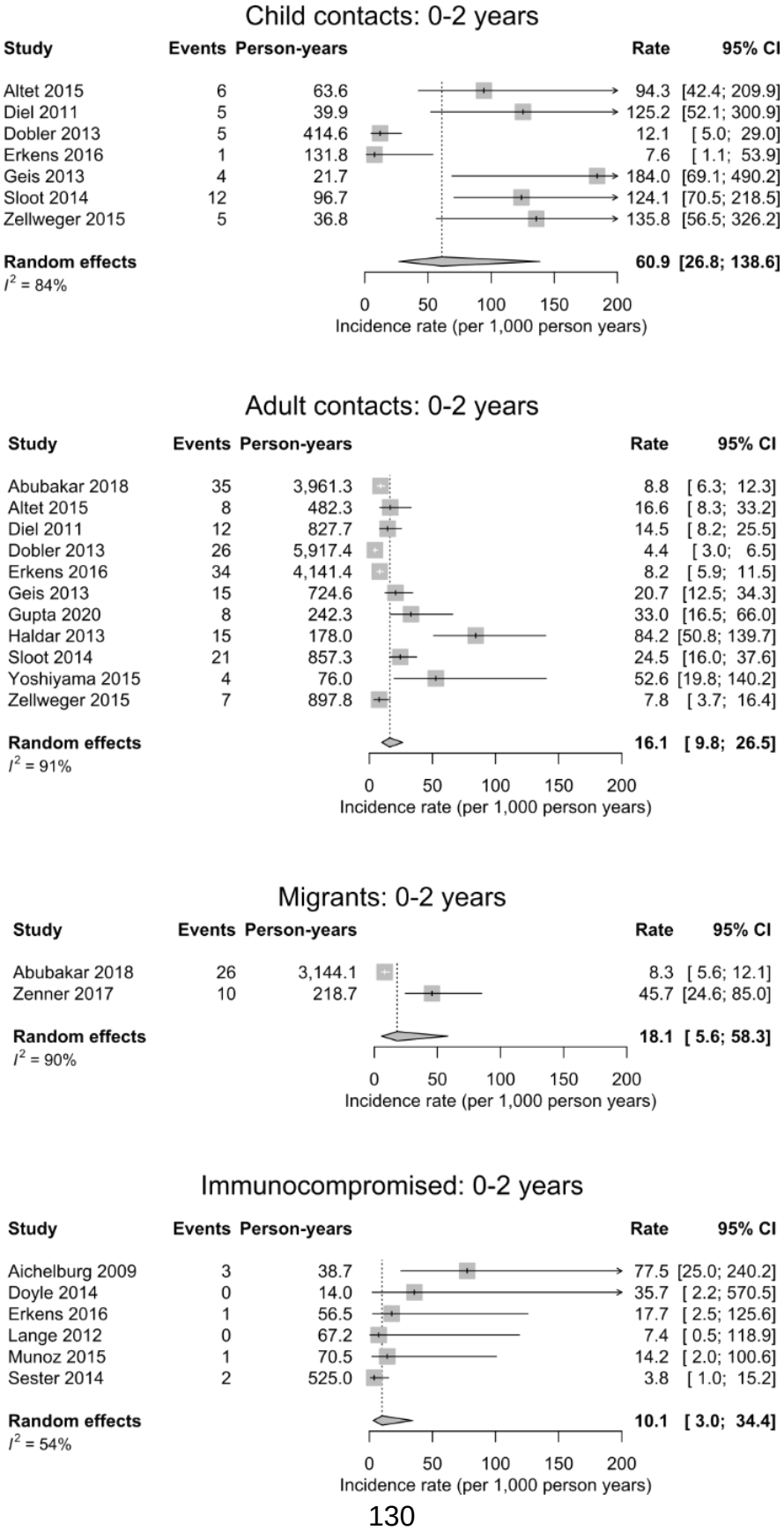
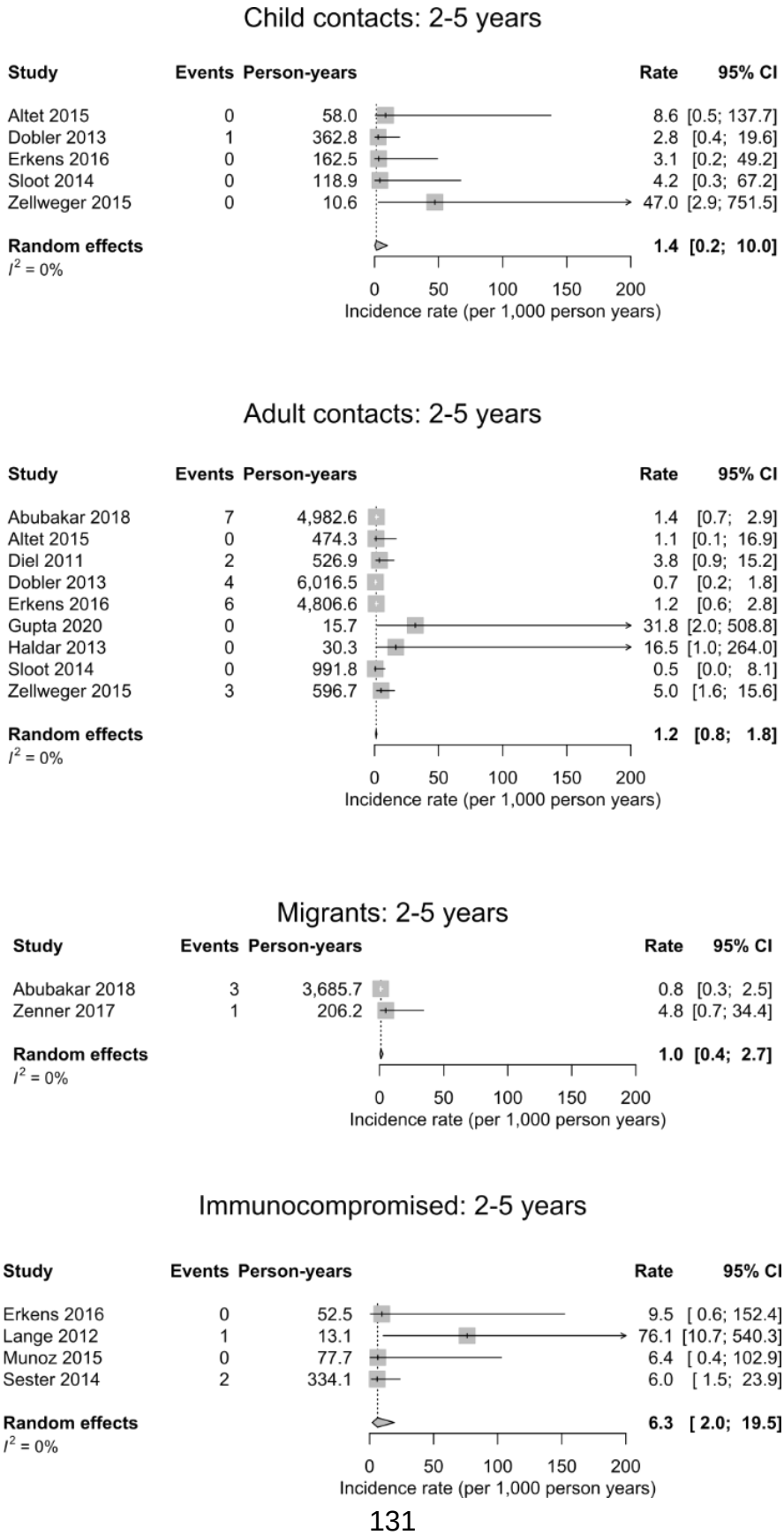


Figure 4-11: TB incidence rates by source study among people with untreated LTBI, during 2-5 year interval following LTBI testing.

Pooled estimates calculated using random effect Poisson models, without continuity correction for studies with zero events. Material from [203].



4.3.3 Prediction model development

The observed heterogeneity in TB incidence rates across studies, even after stratification by binary LTBI result, commencement of preventative treatment and indication for screening, suggests that an individual level approach to risk-stratification may be of benefit. I therefore developed a personalised risk prediction model using a subset of the received data (where sufficient individual level variables were available) including 528 TB patients among 31,721 participants from 15 studies (Figure 4-8). All 15 of these datasets were used for model development and validation, using the IECV framework[211].

Characteristics of the studies included in prediction model development and validation were similar to those that were not (Table 4-5). I selected a flexible parametric survival model with two degrees of freedom on a proportional hazards scale for the final model, since this showed the best fit (as measured by AIC) in each imputed dataset. From the nine candidate predictors (Table 4-4), only previous BCG vaccination and gender were not retained following backward elimination. The final prognostic model therefore included: age, a composite 'TB exposure' variable (modelled with time-varying covariates to account for non-proportional hazards), time since migration for migrants from countries with high TB incidence, HIV status, solid organ or haematological transplant receipt, normalised LTBI test result and preventative treatment commencement. Associations between each of these predictors and incident TB risk are shown visually in Figure 4-12 to aid interpretation; the final model coefficients and standard errors, pooled across multiply imputed datasets, are provided in Table 4-8.

Figure 4-12: Adjusted predictor effects for each predictor in final prognostic model.

Plots demonstrate associations between each predictor and incident TB risk. Illustrative estimates are shown for a 33-year old migrant from a high TB burden setting. The example 'base case' patient does not commence preventative treatment, is not living with HIV, has not received a previous transplant, and has an 'average' positive latent TB test. One of these predictors is varied in each plot. Panel (b) indicates latent TB test result, normalised to a percentile scale. Material from [203].

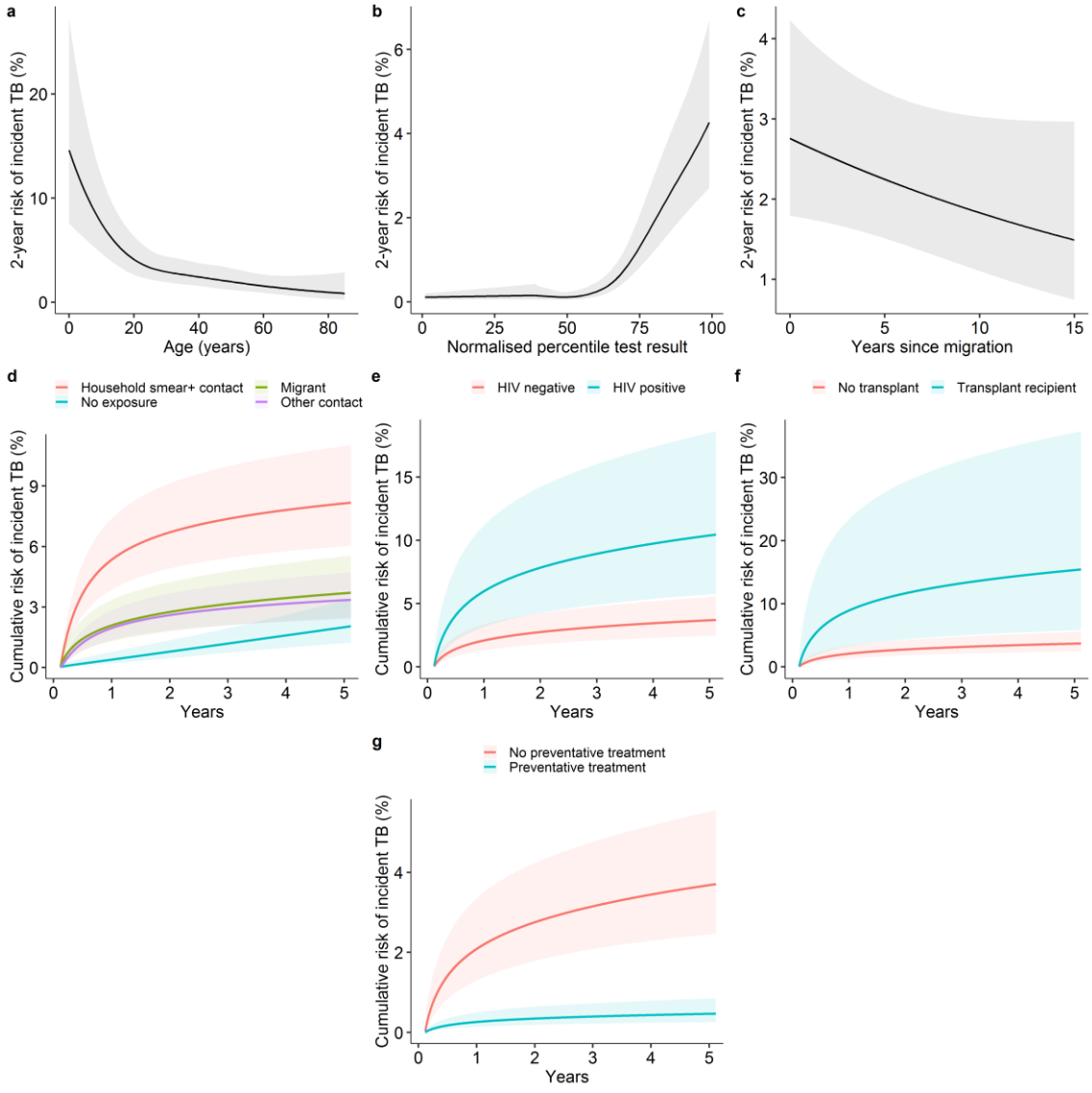


Table 4-8: Final coefficients for PERISKOPE-TB prognostic model.

Coefficients were pooled across multiple imputed datasets using Rubin's rules.

	Coefficient	**Standard error**	**Relative hazard (95% CI)**
Age			
Age spline 1	-0.070	0.017	0.93 (0.9 - 0.97)
Age spline 2	0.129	0.100	1.14 (0.94 - 1.38)
Age spline 3	-0.407	0.619	0.67 (0.2 - 2.24)
Age spline 4	0.266	0.854	1.3 (0.24 - 6.96)
Exposure			
Household, smear positive contact	0.867	1.472	2.38 (0.13 - 42.61)
Other contacts	-1.165	1.659	0.31 (0.01 - 8.06)
Migrant from high TB burden country	0.803	1.717	2.23 (0.08 - 64.61)
*Months since migration**	-0.003	0.002	0.997 (0.99 - 1)
Immune function			
HIV positive	1.072	0.288	2.92 (1.66 - 5.14)
Transplant received	1.490	0.486	4.44 (1.71 - 11.5)
LTBI test result			
Normalised test result spline 1	0.065	0.050	1.07 (0.97 - 1.18)
Normalised test result spline 2	-0.505	0.344	0.6 (0.31 - 1.18)
Normalised test result spline 3	1.796	0.979	6.03 (0.88 - 41.03)
Normalised test result spline 4	-2.438	1.036	0.09 (0.01 - 0.67)
Preventative therapy			
Received preventative therapy	-2.081	0.222	0.12 (0.08 - 0.19)
Baseline hazard			
Intercept	-10.520	1.384	
γ1	5.784	2.272	
γ2	4.323	0.810	
Time-varying covariates (non-proportional hazards)			
*γ1*Household, smear positive contact*	3.327	2.503	
*γ1*Migrant from high TB burden country*	1.721	2.948	
*γ1*Other contacts*	5.103	2.834	
*γ2*Household, smear positive contact*	-1.371	0.850	
*γ2*Migrant from high TB burden country*	-1.502	0.939	
*γ2*Other contacts*	-0.757	0.912	

**for migrants from high TB burden countries who are not recent contacts only.*

4.3.4 Internal-external cross-validation

Model discrimination and calibration parameters for two-year risk of incident TB from the primary validation studies from IECV are shown in Figure 4-13. C-statistics ranged from 0.78 (95% CI 0.47-1.0) in a study of immunocompromised participants with a small number of incident TB cases[225] to 0.97 (0.94-0.99) in a study of TB contacts[192]. The random-effects meta-analysis estimate of the C-statistic was 0.88 (0.82-0.93). Visual calibration plots suggested reasonable

calibration in most studies (Figure 4-14). Since incident TB is an infrequent outcome, predictions were appropriately low, with average predicted risk <10% in all quintiles of risk. The pooled random-effects meta-analysis calibration-in-the-large estimate was 0.14 (-0.24-0.53), with evidence of systematic under-estimation of risk in one study (calibration-in-the-large 1.02 ; 0.61-1.43)), and over-estimation in one study (calibration-in-the-large -0.64 (-1.09-0.19)). The pooled random-effects meta-analysis calibration slope estimate was 1.11 (0.83-1.38). Slopes appeared heterogeneous, though visual assessment of calibration plots suggested that these were prone to being extreme due to the skewed distribution of predicted and observed risk, likely reflecting the relatively rare occurrence of incident TB events.

Figure 4-13: Forest plots showing model discrimination and calibration metrics for predicting two-year risk of incident TB from internal-external cross-validation.

Shown for each study as point estimate [95% CI]. Pooled estimates derived from random effects meta-analysis. Material from [203].

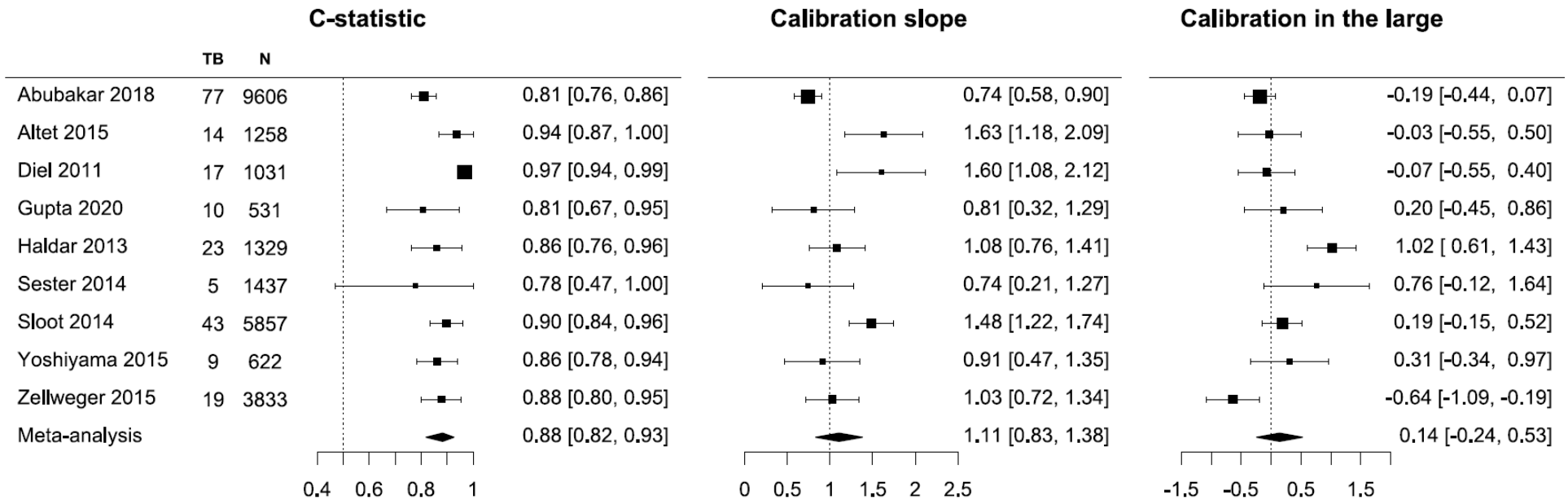
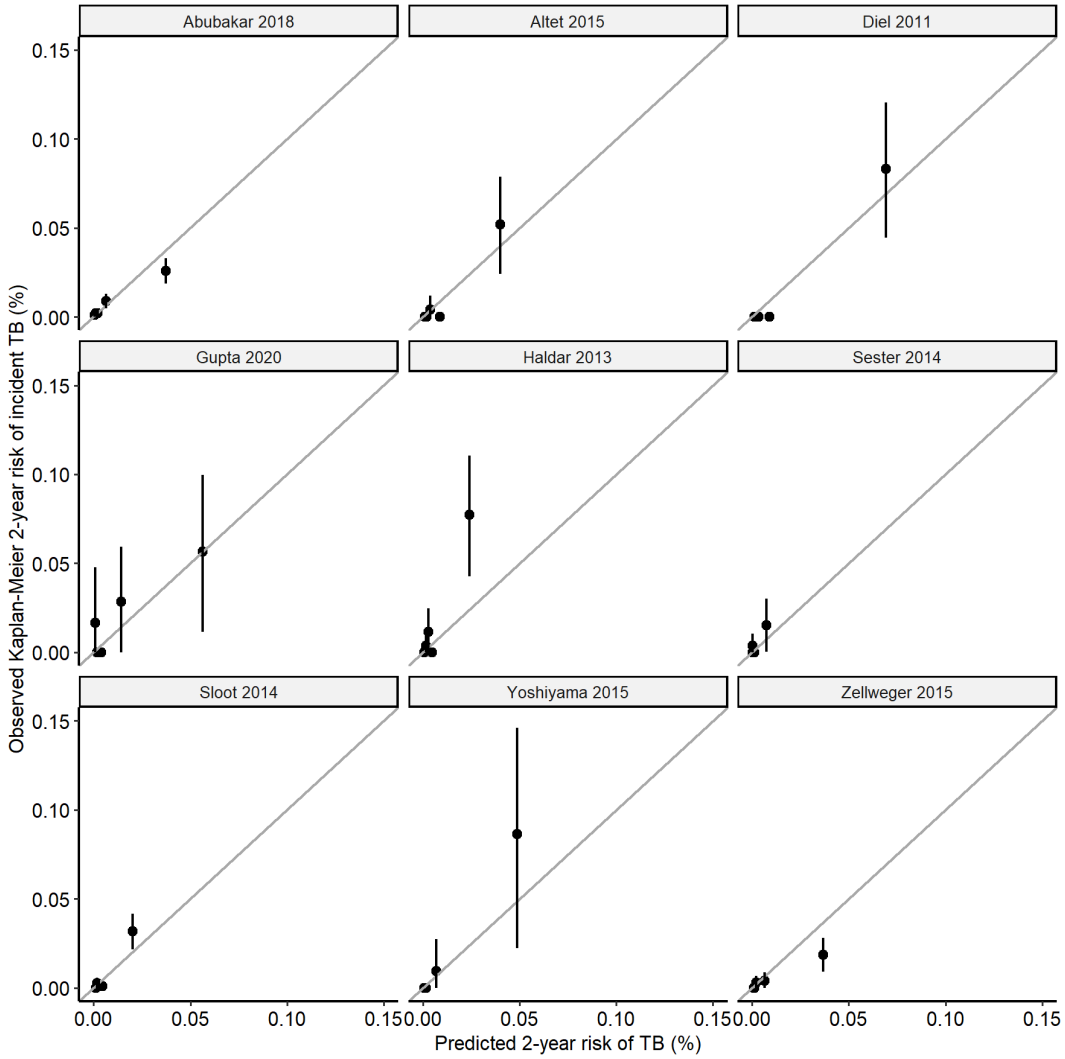


Figure 4-14: Calibration plots from internal-external validation of prediction model, stratified by validation study.

X-axis shows predicted risk, in quintiles, with corresponding Kaplan Meier two-year risk of incident TB on the Y-axis. Material from [203].



4.3.5 Distribution of predicted risk and individual predictions

Figure 4-15 shows the distributions of predicted TB risk among participants who did not commence preventative treatment from the pooled IECV validation sets, stratified by: (a) binary LTBI test result; and (b) indication for screening (among those with a positive test). The median predicted two-year TB risk was 2.0% (IQR 0.8-3.7) and 0.2% (0.1-0.3) among participants with positive and negative binary LTBI test results, respectively. I then examined incident TB risk in four quartiles of predicted risk among untreated participants with positive LTBI tests from the pooled validation sets. Kaplan-Meier plots of the four quartiles showed clear separation of observed risk among these four groups (Figure 4-15).

Figure 4-15: Distribution of predictions and risk of incident tuberculosis in four quartiles of risk for people with positive latent TB tests.

Stratified by (a) binary latent TB test result and (b) indication for screening among untreated people with positive LTBI tests. (c) Randomly sampled individual patients from each risk quartile (characteristics shown in Table 4-9). (d) Shows Kaplan-Meier plots for quartile risk groups (1=lowest risk) of untreated individuals with positive LTBI tests. P value represents Log-rank test. Material from [203].

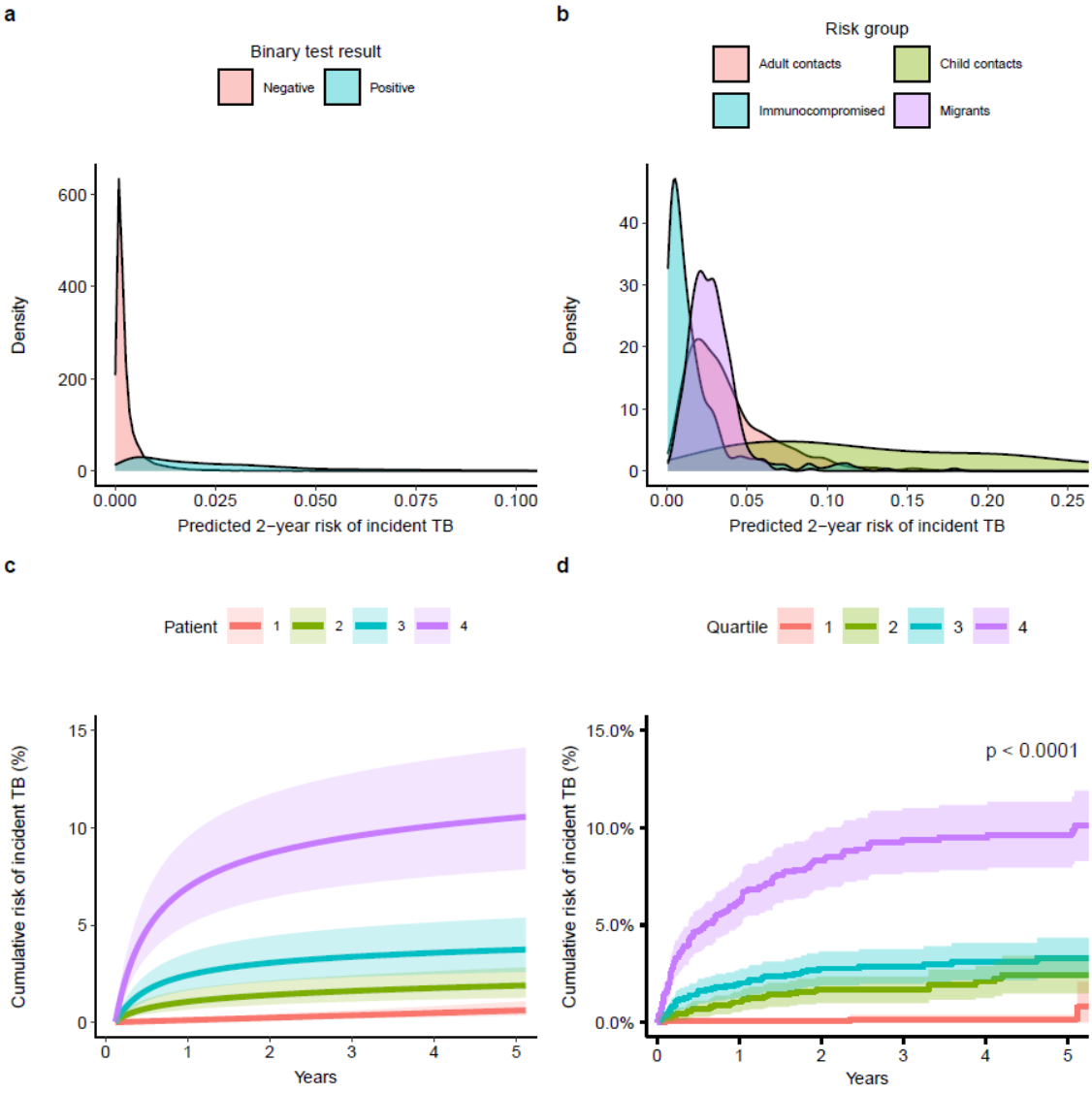


Table 4-9: Characteristics of sampled participants shown in Figure 4-15c.

Patient	Age	Exposure	Test result (percentile)	QFT equivalent (IU / mL)	T-SPOT.TB equivalent (spots)	TST equivalent (mm)	Predicted 2 year risk (%)
1	22	No contact, non-migrant	68	0.10 to 0.12	2	10	0.2 (0.1-0.4)
2	41	Migrant (3.8 years since entry)	80	0.52 to 0.60	7	14	1.4 (0.9-2.1)

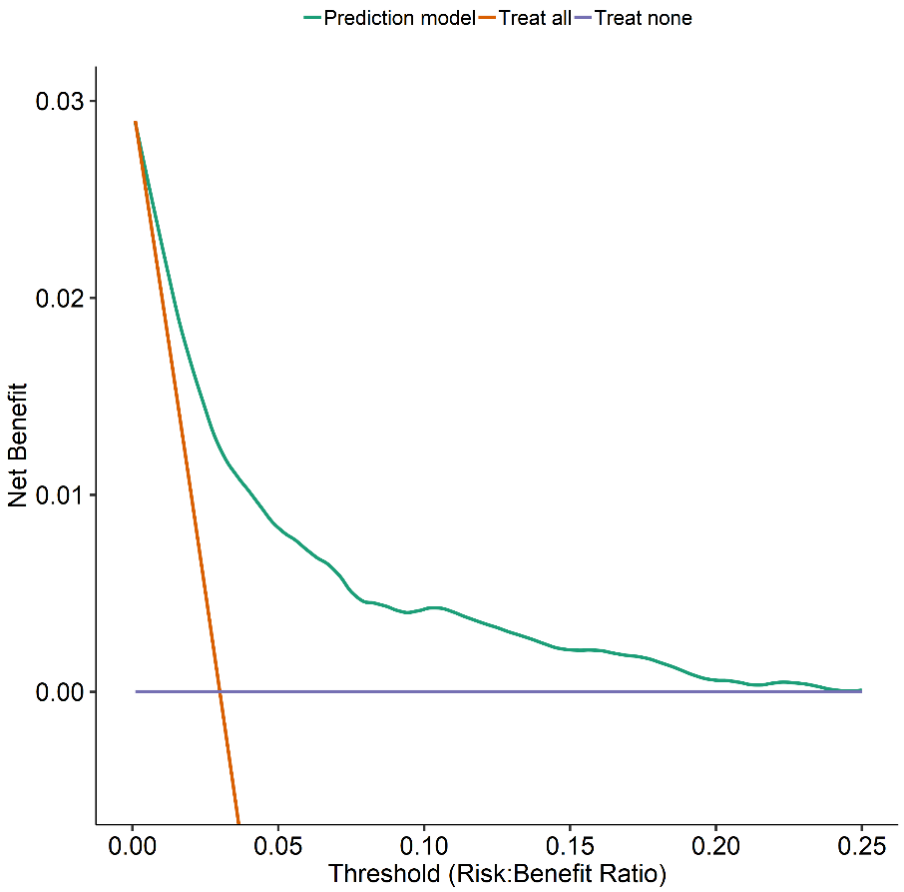
3	51	Household, smear+ contact	79	0.45 to 0.52	6	14	3.1 (2.1-4.4)
4	33	Household, smear+ contact	94	5.10 to 6.29	50 to 58	21	8.7 (6.4-11.7)

4.3.6 Decision curve analysis

Among untreated participants with LTBI from the pooled validation sets in IECV, net benefit for the prediction model was greater than either treating all LTBI patients, or treating none, throughout a range of threshold probabilities from 0-25% (reflecting a range of clinician and patient preferences) (Figure 4-16).

Figure 4-16: Decision curve analysis for PERISKOPE-TB prognostic model.

Shown as net benefit of the prediction model among untreated participants from the stacked validation sets with positive binary latent TB tests (n=6,418 participants), compared to 'treat all' and 'treat none' strategies across a range of threshold probabilities (x-axis). Material from [203].



4.3.7 Sensitivity analyses

Model parameters for a range of sensitivity analyses are presented in Table 4-10. Recalculation of model predictor parameters revealed similar directions and magnitudes of effect to the primary model when using shorter and longer temporal definitions of prevalent TB, though baseline risk was expectedly higher with shorter definitions. Model parameters also appeared similar to the primary analysis when excluding participants who received preventative treatment. Model parameters were more extreme when using a complete case approach.

The pooled random-effects meta-analysis C-statistic from IECV when limiting to participants who did not receive preventative treatment was 0.89 (95% CI 0.82-0.93), similar to the primary analysis and suggesting consistent discrimination when restricted to untreated participants (Figure 4-17a). The pooled random-effects meta-analysis C-statistic including only participants with a positive binary LTBI test was 0.77 (0.70-0.83; Figure 4-17b). This finding indicates good discrimination even among participants with a conventional diagnosis of LTBI, albeit lower than discrimination when also including participants with a negative binary LTBI test, likely reflecting the high negative predictive value of a negative result

Finally, in order to assess model performance in situations where the quantitative test results are not available, I assumed an average quantitative positive or negative LTBI test result (based on the medians among the study population), according to the binary result in the validation sets. This analysis provided a pooled random-effects meta-analysis C-statistic of 0.86 (0.76-0.93; Figure 4-17c), and net benefit appeared higher when using this model than either the strategies of treating all patients with evidence of LTBI, or no patients, across the range of threshold probabilities (Figure 4-17d). However, the model using a binary test result had a lower C-statistic, and slightly lower net benefit across most threshold probabilities, compared to the full model using quantitative test results.

Table 4-10: Recalculation of PERISKOPE-TB model parameters in sensitivity analyses.

Using: alternative definitions of prevalent TB; a complete case approach (for variables other than HIV, which was assumed negative where missing); and excluding participants who received preventative therapy (PT), with comparison to primary model. All coefficients are shown on relative hazards scale, without time-varying covariates to aid interpretation.

	Primary	**Alternative prevalent TB definitions**				**Complete case**	**Excluding patients receiving PT**
		0 days	*<28 days*	*<56 days*	*<90 days*		
Age							
Age spline 1	0.932	0.907	0.932	0.936	0.967	0.942	0.923
Age spline 2	1.137	1.289	1.168	1.097	0.967	1.072	1.182
Age spline 3	0.674	0.338	0.547	0.879	1.550	1.064	0.519
Age spline 4	1.271	2.914	1.712	0.870	0.486	0.648	1.868
Exposure							
Household, smear positive contact	5.961	7.440	6.124	5.919	5.863	8.351	6.591
Other contacts	2.575	2.773	2.529	2.522	2.426	3.333	2.761
Migrant from high TB burden country	2.335	3.036	2.441	2.333	2.466	2.672	2.454
*Months since migration**	0.997	0.997	0.996	0.996	0.996	0.996	0.997
Immune function							
HIV positive	2.803	2.710	2.777	2.799	2.893	5.475	3.114
Transplant received	4.520	4.385	4.331	4.562	5.176	5.637	5.061
LTBI test result							
Normalised test result spline 1	1.068	1.083	1.052	1.063	1.057	1.060	1.080
Normalised test result spline 2	0.602	0.561	0.667	0.622	0.663	0.585	0.557
Normalised test result spline 3	6.087	7.324	4.502	5.430	4.362	7.330	7.593
Normalised test result spline 4	0.086	0.073	0.120	0.102	0.138	0.060	0.068
Preventative therapy							
Received preventative therapy	0.124	0.096	0.122	0.117	0.120	0.097	NA

**for migrants from high TB burden countries who are not recent contact.*

Figure 4-17: Model validation sensitivity analyses.

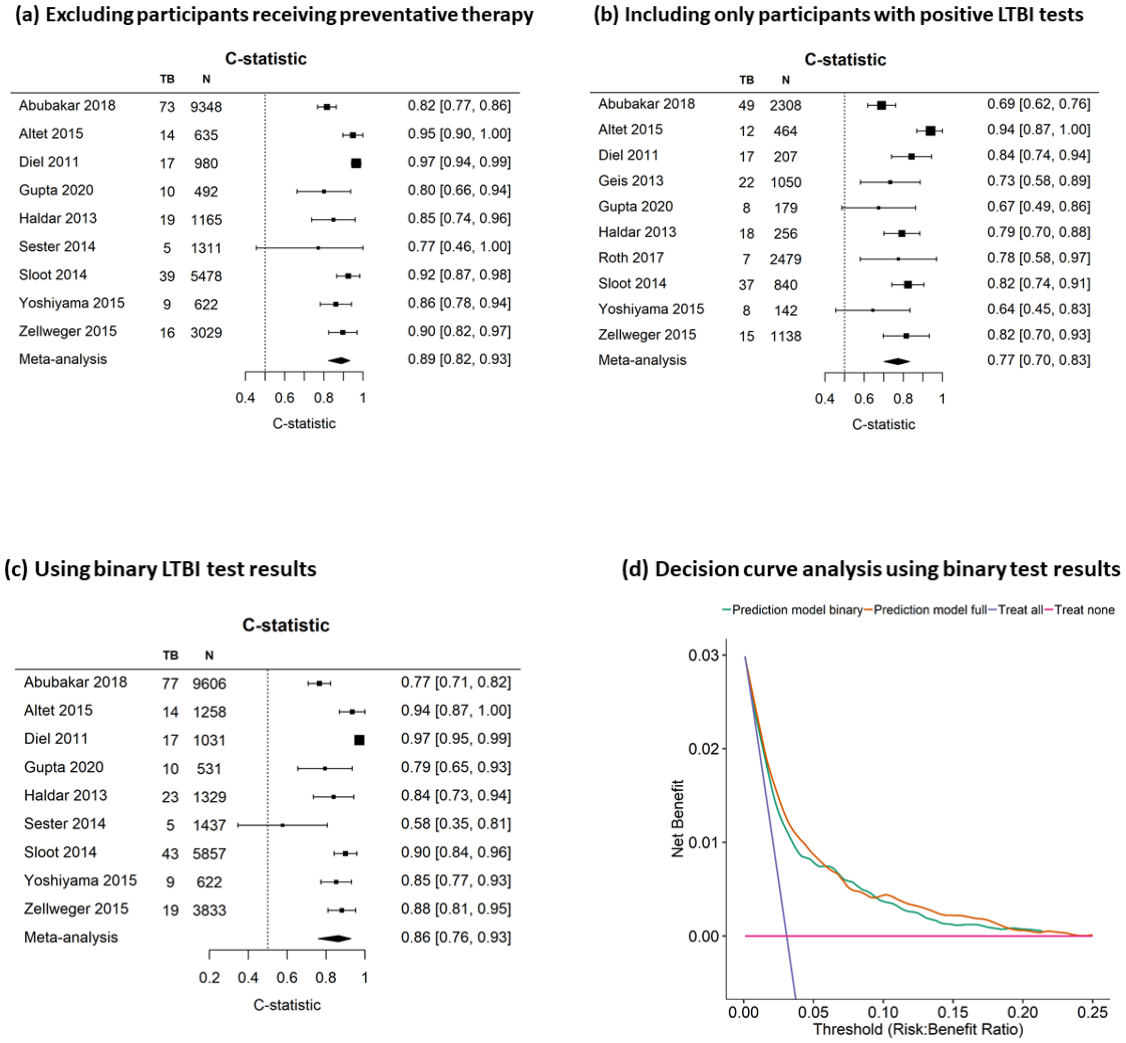
Material from [203]. Recalculation of the C-statistics from internal-external cross validation, limiting validation sets to:

(a) participants who did not receive preventative therapy;

(b) participants with a positive LTBI test;

(c) binary LTBI test results (using an average quantitative positive or negative LTBI test result as appropriate, based on the medians among the study population).

Panel (d) shows decision curve analyses when using the prediction model using a binary LTBI test result, compared to the full prediction model, 'treat all' and 'treat none' strategies across a range of threshold probabilities (x-axis).



4.4 Discussion

4.4.1 Summary of key findings

In this study, I examined population-level incident TB risk in a pooled dataset of >80,000 individuals tested for LTBI in 20 countries with low *M. tuberculosis* transmission. I found cumulative five-year risk of incident TB among people with untreated evidence of LTBI approaching 16% among child contacts, and approximately 5% among recent adult contacts, migrants from high TB burden settings, and immunocompromised individuals. A majority of cumulative five-year risk was accrued during the first year among risk groups with an index exposure, supporting previous data suggesting that risk of progressive TB declines markedly with increasing time since infection[16]. However, there was substantial variation in incidence rates even within these risk groups.

I therefore developed a directly data-driven model that incorporates the magnitude of the T cell response to *M. tuberculosis* with readily available clinical variables in order to capture heterogeneity within risk groups, and generate personalised risk predictions for incident TB in settings aiming towards pre-elimination. Clinical covariates in the final model included age, recent contact (including proximity and infectiousness of the index case), migration from high TB burden countries (and time since arrival), HIV status, solid organ or haematological transplant receipt, and commencement of preventative treatment. The model was validated by quantifying the meta-analysis C-statistic for predicting incident disease over two years, and by evaluating its calibration, using recommended methods[212]. Most importantly, the model showed clear clinical utility for informing the decision to initiate preventative treatment, compared to treating all or no patients with LTBI.

The results of the current analyses are consistent with, and extend existing evidence. Recent analyses report similar population-level TB incidence rates among adult contacts[76], with markedly higher risk among young children[165]. Moreover, these recent meta-analyses confirm highly heterogeneous population-level estimates, thus justifying an individual-level approach to risk estimation[76,165]. Previous models developed and validated in Peru, a high transmission setting, have generated individual or household-level TB risk estimates for TB

contacts[236–238]. Another model ("TSTin3D"), parameterised using aggregate data estimates from multiple global sources, seeks to estimate TB risk following LTBI testing in all settings[166]. However, it is currently validated only in Canada, with evidence moderate discrimination and overestimation of risk[168], and the model omits key predictor variables identified in the current study (including the magnitude of the T cell response and infectiousness of index cases)[166]. TSTin3D includes relative risk estimates for a wide range of co-morbidities. Future studies could compare the performance of PERISKOPE-TB and TSTin3D for predicting incident TB using simulated or real patient cohorts. In addition, TSTin3D also provides estimates of serious hepatotoxicity risk to further facilitate risk: benefit discussions regarding preventative treatment initiation. Such estimates could be added to PERISKOPE-TB in future iterations, ideally using personalised toxicity estimates based on age, comorbidities and preventative treatment regimen (Chapter 1.7.3).

4.4.2 Policy implications

The personalised predictions from the PERISKOPE-TB model may enable more precise delivery of preventative treatment to those at highest risk of TB disease, while concurrently reducing toxicity and costs related to treatment of people at lower risk. Moreover, the model may allow clinicians and patients to make more informed and individualised choices when considering initiation of preventative treatment, through shared decision-making. The PERISKOPE-TB model also challenges the fundamental notion of an arbitrary binary test threshold for diagnosis of LTBI. By incorporating a quantitative measure of immunosensitisation to *M. tuberculosis*, we facilitate a shift from the conventional paradigm of LTBI as a binary diagnosis, towards personalised risk-stratification for progressive TB. This approach takes advantage of stronger T cell responses being a correlate of risk, while guarding against a loss of sensitivity by arbitrarily introducing higher test thresholds programmatically[176].

The PERISKOPE-TB model has been made available at http://periskope.org. Following user input of the required variables, the tool presents incident TB risk estimates numerically and graphically. Importantly, the tool is currently intended to be clinician-facing. Future qualitative work with both clinicians and patients could be undertaken to optimise the tool to ensure optimal risk communication

with appropriate contextualisation, and to facilitate development of a patient-facing interface[239].

4.4.3 Strengths of the study

Strengths of the current study include the size of the dataset, curated through comprehensive systematic review in accordance with Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data standards[195], and with individual participant data obtained for 18/26 (69%) eligible studies. I conducted population-level analyses using both one- and two-stage IPD-MA approaches in order to present both cumulative TB risk and time-stratified incidence rates, respectively, with consistent results from both. I adhered to TRIPOD[200] standards, using the recommended approach of IECV[212]. The coefficients presented in the model are clinically plausible and have been made publicly available to facilitate further independent external validation. Moreover, the contributing datasets included heterogeneous populations of adults, children, recent TB contacts, migrants from high TB burden countries, and immunocompromised groups from 20 countries across Europe, North America, Asia and Oceania, thus making our results potentially generalisable to settings aiming towards pre-elimination globally.

I also used multi-level multiple imputation to account for missing data in the primary analysis, assuming missingness at random and in keeping with recent guidance[204,211]. This approach facilitated imputation of variables that were systematically missing from some included studies. Previous BCG vaccination and HIV status were noted to be missing from a large proportion of participants and were systematically missing from some contributing studies. This missingness may have reduced our power to detect associations between these variables and incident TB along with interactions between BCG vaccination and HIV status and quantitative LTBI test results. BCG vaccination was notably not included in the final prognostic model. While increasing data support a role for BCG vaccination in reducing sensitisation to *M. tuberculosis*[182,240], additional data are required to further assess the association between BCG vaccination and incident TB risk, after adjustment for other covariates including quantitative T cell responses.

I supported the primary multiple imputation approach using a complete case sensitivity analysis. This sensitivity analysis revealed similar findings to the primary analyses, though effect estimates were noted to be more extreme in the complete case approach, likely owing to a degree of bias in the latter, since complete cases analysis assumes no association between the pattern of missingness and the incident TB outcome after adjusting for all other covariates[206]. Given that TB incidence and predictor missingness both varied according to contributing study, this assumption is unlikely to be valid in the current context.

4.4.4 Study limitations

Individual participant data were not obtained for eight eligible studies. Reasons for this included loss of original data, inability to obtain required governance approvals and failure to engage corresponding authors of eligible studies. This highlights a ubiquitous challenge facing IPD-MA studies, which are known to be resource-intensive and dependent on effective engagement with primary study authors[241]. While systematic differences between included and excluded studies were not obvious, it should be acknowledged that the downstream impact of any subsequent selection bias (in addition to publication bias) is difficult to predict.

While overall quality of the included cohort studies was reasonable, there is a potential risk of differential work-up bias since LTBI test results were likely to be available to clinicians to guide decision-making for participants in most included studies. In addition, there is also a risk of incorporation bias, where the LTBI test result may be used to make a diagnosis of TB disease, particularly in young children among whom obtaining microbiological confirmation is more challenging. These considerations could have led to an exaggerated association between LTBI test results and incident TB in the final model if those with positive results were more likely to have been investigated or followed-up for disease.

A further limitation of the current study is that model calibration was observed to be imperfect during external validation, with evidence of underestimation[190] and overestimation[88] of risk in some studies. This may reflect systematic differences between study populations that are not sufficiently captured by the variables included in the model. For example, a UK study among TB contacts found that

14 of 112 (12.5%) untreated IGRA positive adults developed TB, which may suggest inclusion of a particularly high risk population in this study. However, conventional calibration metrics (such as the calibration slope) may not be entirely appropriate in this context, which has a highly skewed distribution of predicted and observed risk, reflecting the rare occurrence of incident TB events. Reassuringly, in decision curve analysis, which accounts for both discrimination and calibration performance in quantifying net benefit, the model showed clinical utility[215].

Due to a lack of data from contributing studies, other potential predictors that may be associated with incident TB risk (including diabetes, malnutrition, fibrotic chest x-ray lesions and other immunosuppression)[7] were not evaluated. These unmeasured covariates may have contributed to imperfect discrimination and calibration, along with residual heterogeneity in model performance between datasets. As additional studies are published, the prognostic model can be prospectively evaluated and updated as required.

I also note that offer and acceptance of preventative treatment may be more likely among people at higher risk of TB. I therefore accounted for preventative treatment provision in the model by including it as a co-variate along with our other predictors of interest, as widely recommended[242]. However, residual confounding by indication cannot be excluded in observational studies. Reassuringly, discrimination of the final prognostic model was consistent when restricted to untreated participants in my sensitivity analyses.

In addition, the present model is not applicable for patients commencing biologic agents since no datasets were identified that examined the natural history of LTBI in the context of biologic therapy, in the absence of preventative treatment for TB. A ‘hybrid’ modelling approach, with mathematical parameterisation of relative risk for any given biologic agent, may be required to extend its application to these therapies. Since the quantitative LTBI test result is a strong predictor in our model, predictions may also be attenuated in the context of advanced immunosuppression[66]. Reassuringly. performance appeared adequate in a dataset of immunocompromised individuals during validation[225].

Finally, in some circumstances, full quantitative LTBI test results may not be available. In my sensitivity analyses, I evaluated performance of the prognostic model when using a median quantitative value based on binary results. In decision curve analysis, this approach appeared to have higher net benefit than using binary LTBI results alone, but lower net benefit than using full quantitative results. Therefore, median values based on binary results may be used in the prognostic value if full quantitative results are unavailable.

4.4.5 Conclusions

In summary, I have developed a freely available and directly data-driven personalised risk predictor for incident TB (PERISKOPE-TB; periskope.org). This tool may facilitate a more individualised approach for TB prevention services in settings aiming towards pre-elimination, by facilitating shared decision-making between clinicians and patients for preventative treatment initiation.

4.5 Contribution statement

This chapter included a systematic review and pooled IPD-MA of 18 previously reported cohort studies. I led the work from conception to completion and dissemination.

4.6 Outputs relating to this chapter

This study is published in *Nature Medicine*:

Gupta, R.K., Calderwood, C.J., Yavlinsky, A… Lipman M., Noursadeghi M., Abubakar I. (2020) Discovery and validation of a personalized risk predictor for incident tuberculosis in low transmission settings. **Nature Medicine** 26, 1941–1949. https://doi.org/10.1038/s41591-020-1076-0

I also presented this work at the British Thoracic Society Winter Meeting 2020.

5 Objective 4: Concise whole blood transcriptomic signatures for incipient TB: a systematic review and participant-level pooled meta-analysis

5.1 Introduction

Multiple studies have discovered changes in the host transcriptome in association with TB disease, compared to healthy controls, individuals with LTBI or other diseases[104–110,121,243–245]. More recently, perturbation in the transcriptome has been found to predate the diagnosis of TB[118,120,121,246], suggesting that transcriptomic signatures may offer an opportunity to diagnose incipient TB and potentially fulfil the WHO TPP. However, independent validation of each signature is still limited. It remains unclear which of the multiple candidate transcriptomic signatures performs best for the identification of incipient TB, or whether any signatures meet the WHO diagnostic accuracy benchmarks.

To address these knowledge gaps, I performed a systematic literature review to identify concise whole blood transcriptomic signatures for incipient TB, along with whole blood transcriptomic datasets, with sampling prior to TB diagnosis. I then performed an IPD-MA of genome-wide transcriptomic data to compare the diagnostic accuracy of the identified candidate transcriptomic signatures for detection of incipient TB among people at risk of disease over a two-year horizon. Finally, I evaluated the diagnostic accuracy of the best performing transcriptomic signatures, stratified by pre-defined time intervals to TB, in order to assess whether they meet the WHO TPP specifications for incipient TB tests.

5.2 Methods

5.2.1 Overview of systematic review

I hypothesised that any biomarker that distinguishes incipient or active TB from healthy people may detect incipient disease. I therefore performed a systematic review and IPD-MA, in accordance with Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data standards[195], to identify candidate concise whole blood transcriptomic signatures for incipient or active TB. I then examined the performance of eligible signatures in published

whole blood transcriptomic datasets where sampling prior to TB diagnosis was performed and interval time to disease was available. The study was registered at PROSPERO (CRD42019135618).

5.2.2 Search strategy

I searched Medline and Embase on 15/04/2019, with no language or date restrictions. The Medline search strategy is outlined in Table 5-1.I included comprehensive terms for ‘biomarkers’ (terms 1-11); ‘tuberculosis (term 12); ‘transcriptome’ (terms 13-19); and ‘blood’ (terms 20-22). I consolidated the search by also hand-searching reference lists of relevant review articles and consulting experts in the field.

Table 5-1: Medline search strategy for systematic review of concise RNA biomarkers for incipient TB.

1. Biomarkers/
2. Diagnostic Tests, Routine/
3. "Predictive Value of Tests"/
4. diagnostic test*.mp.
5. biomarker*.mp.
6. ppv.mp.
7. npv.mp.
8. sensitivit*.mp.
9. specificit*.mp.
10. signature*.mp.
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. exp TUBERCULOSIS/ or tuberculosis.mp. or exp MYCOBACTERIUM TUBERCULOSIS/ or tb.mp.
13. RNA/
14. Transcriptome/
15. rna.mp.
16. transcript*.mp.
17. gene expression.mp.
18. Gene Expression Profiling/ or RNA, Messenger/ or Transcription, Genetic/ or Gene Expression/
19. 13 or 14 or 15 or 16 or 17 or 18
20. blood/

21. blood.mp.
22. 20 or 21
23. 11 and 12 and 19 and 22
24. remove duplicates from 23
25. limit 24 to "humans only (removes records about animals)"

5.2.3 Eligibility criteria for candidate signatures

I included whole blood messenger RNA signatures that met the following eligibility criteria:

- Discovered with a primary objective of diagnosis of active or incipient TB, compared to controls who were either deemed healthy, or had latent TB infection.
- Signatures should be ‘concise’: In the absence of a standardised definition, I defined concise as signatures that used a defined approach to feature selection to reduce multidimensionality and the number of constituent genes, thus leading to biomarkers that may be more amenable to clinical translation.
- Gene names that comprise the signature, along with the corresponding equation or modelling approach, must be available.
- Signature (including component genes, and modelling approach) must be validated in at least one independent test or validation set, in order to enable reliable signature reconstruction and prioritise the most promising candidate signatures from higher quality studies.
- Signature must be discovered from training sets that included controls who were either deemed healthy, or had latent TB infection, since discriminating incipient TB from healthy or latently infected people is the primary aim of incipient TB diagnostics.
- Where multiple signatures were discovered for the same intended purpose and from the same training dataset, I included the signature with greatest accuracy (as defined by the AUROC in the validation data). Where accuracy was equivalent, I included the signature with fewest number of component genes.

5.2.4 Eligibility criteria for transcriptomic datasets

I included published whole blood transcriptomic datasets (RNA sequencing or microarray) where sampling prior to TB diagnosis was performed and interval time to disease was available. I specified a minimum median duration of follow-up of one year to reduce the risk of outcome misclassification. For studies where preventative TB therapy was offered, individual level data was required to identify the treated cases, who were excluded in the analysis.

5.2.5 Screening and data extraction

Two independent reviewers screened titles and abstracts identified in the search, and determined eligibility for final inclusion following full-text review. Gene lists and corresponding equations or modelling approaches were extracted for each eligible candidate signature and verified by a second reviewer. Disagreements regarding study inclusion or signature calculations were resolved through arbitration by a third reviewer. Quality assessment and risk of bias were assessed for the studies corresponding to included transcriptomic datasets, using modified versions of the Newcastle-Ottawa scale (using the cohort or case-control version as appropriate to each contributing study)[196].

5.2.6 Extension of UK cohort of TB case contacts

In preparation for this meta-analysis the follow-up of a previously published cohort of London TB contacts with RNA sequencing data (described in Chapter 3)[120] was extended by re-linking the full cohort to national TB surveillance records (until 31/12/2017; median follow up increased from 0.9 in original RNA sequencing report to 1.9 years) held at Public Health England using a validated algorithm[184]. An additional 27 samples and individuals were also available for inclusion in the present analysis, compared to the originally reported RNA sequencing analysis[120]. The full updated data set for this study is available in ArrayExpress (Accession number E-MTAB-6845).

The London contacts study was approved by the UK National Research Ethics Service (reference: 14/EM/1208)[120]. No other ethical approvals were sought for this meta-analysis, since all other included patient-level datasets were depersonalised and publicly available.

5.2.7 RNA data processing

Individual level RNA sequencing data were first downloaded for eligible studies, and mapped to the reference transcriptome (Ensembl Human GRCh38 release 95) using Kallisto[247]. The transcript-level output counts and transcripts per million (TPM) values were summed on gene level and annotated with gene symbols using tximport and BioMart[248,249]. Only protein-coding genes were selected for downstream processing, and TPM and counts per million (CPM) values <0.001 were set to 0.001 prior to log2 transformation to act as a lower limit of detection.

I used principal component analysis (PCA) to visualise the TPM data, stratified by source study, in order to examine for between-study technical heterogeneity and determine the need for batch correction. PCA enables visualisation of multidimensional data through dimensionality reduction, while preserving as much information as possible, by creating new uncorrelated variables (principal components) that maximise variance sequentially[250]. PCA also requires no distributional assumptions for descriptive purposes and is therefore a flexible approach to exploratory visualisation of high-dimensional data. My PCAs included (a) genome-wide protein coding genes; (b) selected genes comprising only the candidate signatures included in the analysis; and (c) the intersect of invariant genes that were in the lowest quartile of genes ranked by variance within each of the contributing datasets. The latter PCA focusing on invariant genes was done in order to examine whether observed differences in genome-wide PCAs were likely to be attributable to technical, as opposed to biological, differences between datasets.

Batch correction was performed using the COmbat CO-Normalization Using conTrols (COCONUT) package in R[251]. This approach applies the ComBat function, which adjusts the mean and variance of each gene in each contributing study dataset to minimise batch effect parameters[252]. Unlike the ComBat function, using COCONUT facilitated calculation of batch correction parameters based on the TB-free controls only, which was then applied to those who developed TB disease. I chose this approach, as opposed to applying batch correction regardless of outcome classification, in order to reduce the risk of biasing the distributions of gene expression between datasets, which could

otherwise occur due to differing disease prevalence between the study populations included.

5.2.8 Definitions and sample inclusion

Only samples obtained prior to the diagnosis of TB were included in the analysis. 'Prevalent' TB was defined as a TB diagnosis within 21 days of sample collection, as previously[80]. 'Incipient TB' cases were defined as individuals diagnosed with TB >21 days from blood RNA sample collection. Culture-confirmed and clinically or radiologically diagnosed pulmonary or extra-pulmonary TB cases were included in the primary analysis. 'Non-progressors' were defined as those who remained TB disease free during follow-up; those with less than six months' follow-up from the date of sample collection were excluded due to risk of outcome misclassification. Participants with prevalent TB and those who commenced preventative therapy were excluded. For studies with longitudinal samples from the same individuals, serial samples were included provided that they met these criteria, and that they were collected at least six months apart. Serial samples were handled as being independent in the primary analysis.

5.2.9 Calculation of signature scores

Gene symbols of the original signatures were updated to Ensembl Human GRCh38 release 95. Genes that were not present in the RNA sequencing data, including withdrawn genes and non-coding genes, were omitted from score calculations. Unless otherwise stated, log2-transformed TPM values were used for score calculations.

I sought to use the authors' original described methods to calculate scores from component genes for each signature. 'Disease risk scores' were calculated as difference of sums between upregulated and downregulated signature genes[107]. 'Modified' disease risk scores[121] and 'unsigned sums'[245] were calculated as the sum of expression levels, regardless of the direction of regulation in TB. 'Sum of standardised expression' scores [253] were calculated by standardising expression values for each component gene and then summing. 'Difference in geometric means' scores were calculated by subtracting the mean expression of downregulated genes from the mean of upregulated genes, on the $\log_2$ scale. For regression models where the coefficients were publicly available, scores were

calculated as sum of weighted gene expression values, using the regression coefficients from the original publication.

Random forest models were constructed with the randomForest package in R, using the original standardised training data. Support vector machine (SVM) models were constructed with a linear kernel in the ksvm function of the kernlab package in R, using the original training data for the respective signature. My reconstructed random forest and SVM models were validated against the original data where possible, by comparing AUROCs to the original authors' descriptions.

Batch-corrected signature scores were transformed to Z scores (by subtracting the control mean, and dividing by standard deviation), using a pre-defined 'control' population (including only participants with negative LTBI tests among the pooled dataset), in order to standardise scaling across signatures[120].

5.2.10 Statistical analysis

All analyses were performed using R (version 3.5.1), unless otherwise specified. ROC curves for each signature for the identification of incipient TB were first plotted, for a two-year time horizon. Any data that was originally used to derive specific signatures were excluded from the pooled dataset used to test the performance of the relevant signature to reduce the risk of observing overly optimistic validation performance.

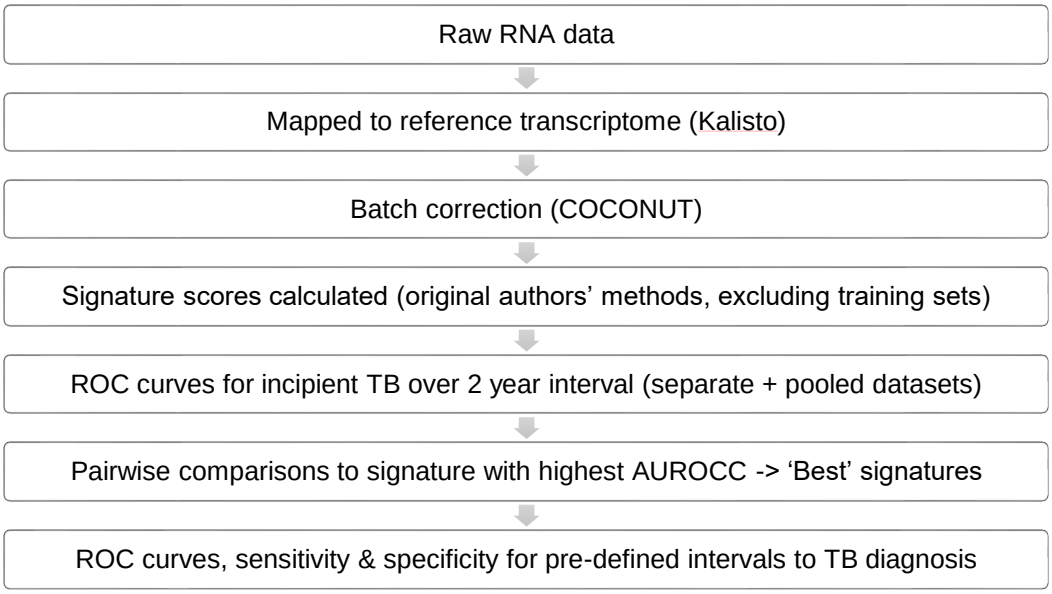
ROC curves and AUROCs for separate contributing study datasets were initially examined to assess the degree of between-study heterogeneity. Since little heterogeneity was observed after stratification by interval to disease, a one-stage IPD-MA, assuming common signature performance across studies, was performed for the primary analysis. AUROCs for each signature were directly compared in a pairwise approach, using the DeLong method[178]. The best performing signature available from all samples in the pooled dataset was used as the reference for comparison with all other signatures; signatures with AUROCs lower than the reference, with $p<0.05$, were deemed statistically inferior.

Correlation between signature scores was assessed using Spearman rank correlation. Pairwise Jaccard similarity indices between signatures were

calculated using lists of their constituent genes. Clustered co-correlation and Jaccard index matrices were generated in Morpheus[254] using average Euclidean distance. Upstream analysis of transcriptomic regulation was performed using Ingenuity Pathway Analysis (Qiagen, Venlo, The Netherlands) and visualized as network diagrams in Gephi v0.9.2, depicting all statistically overrepresented molecules predicted to be upstream of more than two target genes. This upstream analysis was done in order to highlight the predicted regulators shared by the constituents of the transcriptomic signatures.

ROC curves and AUROCs were then assessed for the best performing (statistically equivalent) signatures, using pre-specified intervals to TB of <3 months; <6 months; <1 year; and <2 years from sample collection. Sensitivity and specificity for each of these time intervals were determined at pre-defined cut-offs for each signature, defined as a standardised score of two (Z2), representing the 97.7th percentile of the control population with negative LTBI tests, assuming a Normal distribution, as in previous work[120]. These estimates were used to model the estimated predictive values for incident TB across a range of pre-test probabilities. Sensitivity and specificity for the best performing signatures were also examined using cut-offs defined by the maximal Youden index[179], in order to achieve the highest accuracy within each time interval. The full analysis pipeline is summarised visually in Figure 5-1.

Figure 5-1: Flowchart depicting analysis pipeline for RNA sequencing data.



5.2.11 Sensitivity analyses

Sensitivity analyses included:

- Restricting inclusion of TB cases to those with microbiological confirmation.
- Including only one blood RNA sample per participant by randomly sampling one blood sample per individual, in studies which included serial sampling, in order to evaluate potential bias caused by inclusion of differing numbers of samples per participant.
- Recomputing the ROC curves using mutually exclusive time intervals to TB of 0-3, 3-6, 6-12 and 12-24 months, for each curve excluding participants who had developed TB in an earlier interval.
- Performing a two-stage IPD-MA to complement the primary one-stage IPD-MA. To do this, I calculated AUROCs for each signature, stratified by interval to disease, in each contributing dataset separately, prior to batch correction. I then derived pooled AUCs and 95% CIs for each signature across studies using random-effects meta-analysis of logit-transformed AUROCs, using the metamisc package in R[214]. I also calculated sensitivity, specificity and predictive values at Z2 score cut-offs for each signature within each batch-corrected dataset, and derived pooled estimates using bivariate random-effects meta-analysis in the mada package in R[255].

5.3 Results

5.3.1 Systematic review process and summaries of included datasets and signatures

A total of 643 unique articles were identified in the systematic review (Figure 5-2). Four RNA datasets and 17 signatures met the criteria for inclusion. The RNA datasets included the Adolescent Cohort Study (ACS) of South African adolescents with LTBI[118], the Bill and Melinda Gates Foundation Grand Challenges 6-74 (GC6-74) household TB contacts study in South Africa, the Gambia and Ethiopia[246], a London TB contacts study[120], and a Leicester TB contacts study[121](Table 5-2). All four eligible datasets were publicly available. The ACS and GC6-74 studies were nested case-control designs within larger prospective cohort studies, while the London and Leicester TB contacts studies

were prospective cohort studies, with RNA sequencing performed for all participants. All four studies were done in HIV-uninfected participants. The London TB contacts study included only baseline samples, while the ACS, GC6-74 and Leicester TB contacts studies included serial sampling. All four studies assessed participants for evidence of prevalent TB at enrolment through clinical evaluation, while the London and Leicester TB contacts studies also performed chest radiographs. The GC6-74 and ACS studies excluded participants with TB diagnosed within three or six months of enrolment, respectively. All four studies achieved maximal quality assessment scores (full quality assessments shown in Supplementary Material, Chapter 8.2).

Figure 5-2: Flowchart showing systematic review process for review and meta-analysis of concise RNA signatures for incipient TB.

Material from [256].

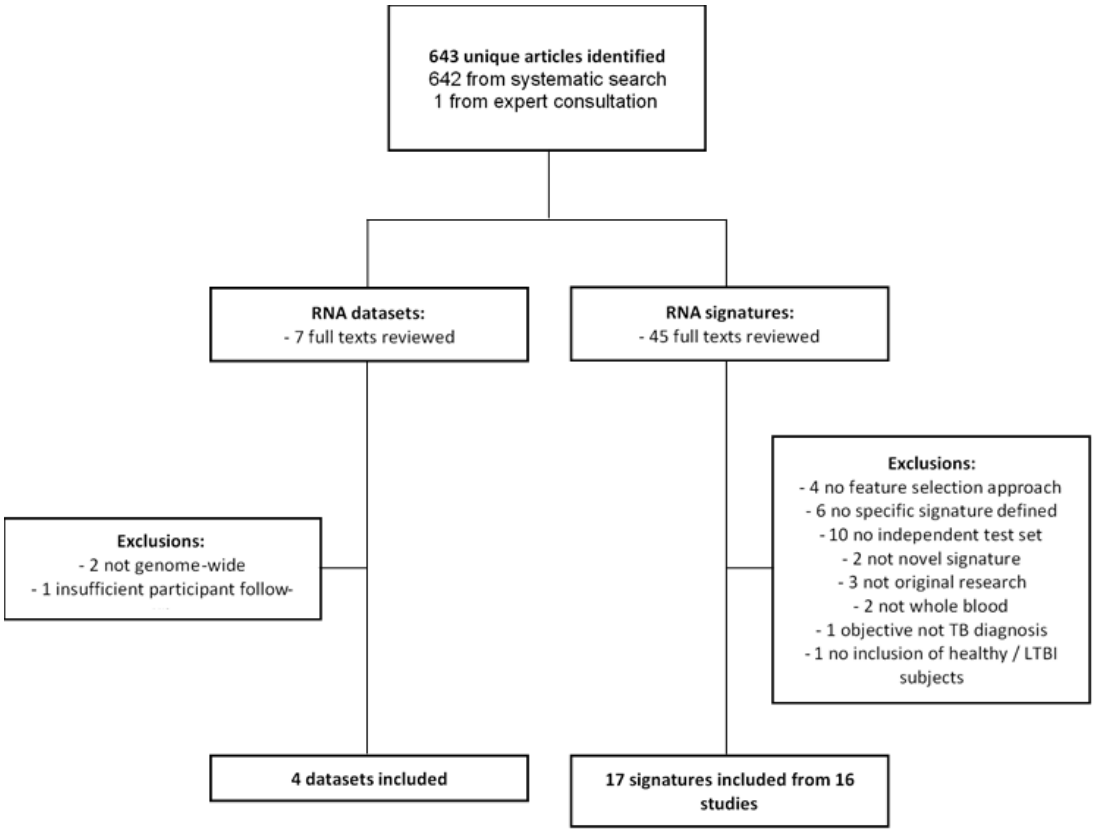


Table 5-2: Characteristics of the datasets included in meta-analysis of candidate RNA signatures for incipient TB.

Study	Samples included	Study design	Population	Setting	HIV status	Sampling	Follow-up duration and method	TB case definition	RNAseq methods	NOS score^	Baseline TB assessment
London TB contacts[120]	324 (8 TB; 316 healthy)	Cohort	Adult TB contacts	London	Negative	Baseline	Median 1.9 years, record linkage	Culture-confirmed, or clinically diagnosed	15-20 million 41 base pair paired end reads	7 (/7)	Clinical evaluation and chest radiograph
Adolescent Cohort Study[118]	287 (73 TB; 214 healthy)	Nested case-control	Adolescents with latent TB infection	South Africa	Negative	Serial (0, 6, 12, 24 months)	2 years, active	Intrathoracic disease with 2 positive smears, or 1 positive culture	30 million 50 base pair paired end reads	9 (/9)	Clinical evaluation. TB <6 months from enrolment excluded. Chest radiograph not specified
Grand Challenges 6-74[246]	412 (98 TB; 314 healthy)	Nested case-control	Adult household pulmonary TB contacts	South Africa, The Gambia, Ethiopia	Negative	Serial (0, 6, 18 months)	2 years, active	Culture-confirmed, or clinically diagnosed	60 million 50 base pair paired end reads	9 (/9)	Clinical evaluation. TB <3 months from enrolment excluded. Chest radiograph not specified
Leicester TB contacts[121]	103 (4 TB; 99 healthy)	Cohort	Adult TB contacts	Leicester	Negative	Baseline + serial for a subset*	2 years, active	Culture- or Xpert MTB/RIF-confirmed	25 million 75 base pair paired end reads	7 (/7)	Clinical evaluation and chest radiograph

RNAseq = RNA sequencing.

**Due to the high frequency of serial sampling (<6-monthly), only baseline samples were included.*

^NOS = Newcastle-Ottawa Scale (denominators shown in brackets).

A total of 1,126 samples from 905 patients met my criteria for inclusion (flowchart shown in Supplementary Material, Chapter 8.2). Characteristics of participants are shown in Table 5-3. I included 183 samples from 127 incipient TB cases, of which 117 (92.1%) were microbiologically confirmed. Only 8/127 TB cases (6.3%) were known to be extra-pulmonary, without pulmonary involvement. Of note, a large proportion of participants in the London (112/324; 34.6%) and Leicester (86/103; 83.5%) contact studies were of South Asian ethnicity.

PCAs revealed clear separation of samples by dataset when including (a) the entire transcriptome; (b) selected genes comprising only the candidate signatures included in the analysis; and (c) invariant genes, indicative of batch effects in the data due to technical variation in RNA sequencing (Figure 5-3a-c)[257]. These batch effects were eliminated after batch correction (Figure 5-3d), while preserving the distributions of expression for target genes within each contributing dataset (Supplementary Material; Chapter 8.2).

Table 5-3: Baseline characteristics of participants in meta-analysis of concise whole blood transcriptomic signatures for incipient TB, stratified by study.

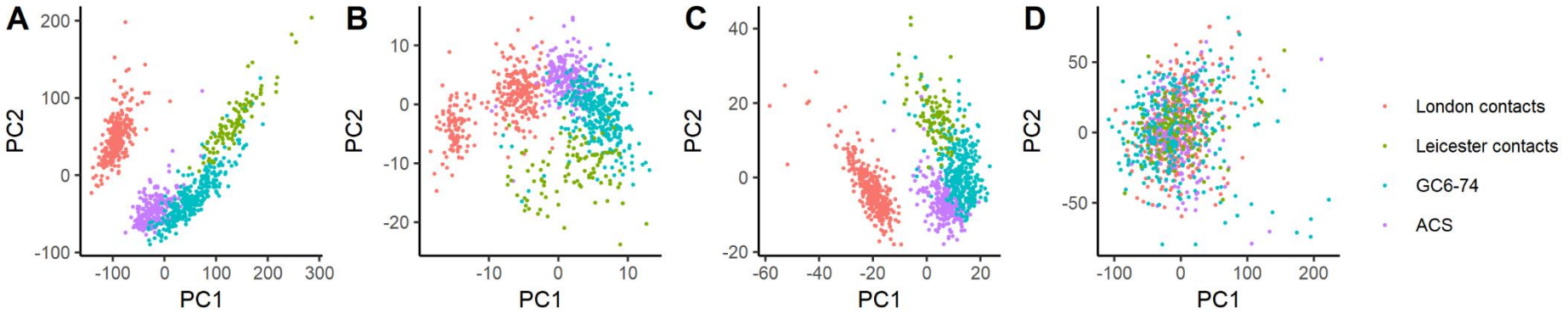
ACS = adolescent cohort study; GC6-74 = Bill and Melinda Gates Foundation Grand Challenges 6-74 TB contacts study; IGRA = interferon gamma release assay; IQR = interquartile range. Data shown as n(%) unless otherwise specified.

Category	Level	London contacts	Leicester contacts	GC6-74	ACS	All
Participants	*n*	324	103	334	144	905
Age	*Median (IQR)*	34 (26, 47)	35 (24, 45.5)	23 (19, 35)	16 (15, 17)	26 (18, 40)
Gender	*Female*	153 (47.2)	43 (41.7)	197 (59.0)	97 (67.4)	490 (54.1)
	Male	166 (51.2)	60 (58.3)	137 (41.0)	47 (32.6)	410 (45.3)
	Missing	5 (1.5)	0 (0.0)	0 (0.0)	0 (0.0)	5 (0.6)
Ethnicity	*White*	75 (23.1)	6 (5.8)	0 (0.0)	0 (0.0)	81 (9.0)
	Black African or Caribbean	66 (20.4)	10 (9.7)	0 (0.0)	12 (8.3)	88 (9.7)
	South Asian	112 (34.6)	86 (83.5)	0 (0.0)	0 (0.0)	198 (21.9)
	Mixed	0 (0.0)	0 (0.0)	0 (0.0)	132 (91.7)	132 (14.6)
	Other	61 (18.8)	1 (1.0)	0 (0.0)	0 (0.0)	62 (6.9)
	Missing	10 (3.1)	0 (0.0)	334 (100.0)	0 (0.0)	344 (38.0)
IGRA	*Negative*	219 (67.6)	50 (48.5)	0 (0.0)	3 (2.1)	272 (30.1)
	Positive	105 (32.4)	53 (51.5)	0 (0.0)	36 (25.0)	194 (21.4)
	Missing	0 (0.0)	0 (0.0)	334 (100.0)	105 (72.9)	439 (48.5)
Country	*Ethiopia*	0 (0.0)	0 (0.0)	36 (10.8)	0 (0.0)	36 (4.0)
	South Africa	0 (0.0)	0 (0.0)	180 (53.9)	144 (100.0)	324 (35.8)

	The Gambia	0 (0.0)	0 (0.0)	118 (35.3)	0 (0.0)	118 (13.0)
	UK	324 (100.0)	103 (100.0)	0 (0.0)	0 (0.0)	427 (47.2)
Outcome	*Non-progressor*	316 (97.5)	99 (96.1)	259 (77.5)	104 (72.2)	778 (86.0)
	Incipient TB	8 (2.5)	4 (3.9)	75 (22.5)	40 (27.8)	127 (14.0)
Months from recruitment to TB	*Median (IQR)*	8.6 (6.4, 11.1)	1.9 (1.0, 3.2)	10.5 (5.5, 17.5)	14.4 (8.8, 18.6)	10.3 (5.5, 17.5)
Microbiological confirmation	*No*	5 (62.5)	0 (0.0)	5 (6.7)	0 (0.0)	10 (7.9)
	Yes	3 (37.5)	4 (100.0)	70 (93.3)	40 (100.0)	117 (92.1)
Pulmonary	*No*	7 (87.5)	1 (25.0)	0 (0.0)	0 (0.0)	8 (6.3)
	Yes	1 (12.5)	3 (75.0)	75 (100.0)	40 (100.0)	119 (93.7)
Samples	*n*	324	103	412	287	1126
	Non-progressor	316 (97.5)	99 (96.1)	314 (76.2)	214 (74.6)	943 (83.7)
	Incipient TB	8 (2.5)	4 (3.9)	98 (23.8)	73 (25.4)	183 (16.3)
Months from sample to TB	*Median (IQR)*	8.6 (6.4, 11.1)	1.9 (1.0, 3.2)	7.5 (5.5, 15.5)	9.3 (6.6, 15.0)	8.5 (5.5, 15.1)
	<3	1 (12.5)	3 (75.0)	6 (6.1)	11 (15.1)	21 (11.5)
	3 to 6	1 (12.5)	1 (25.0)	35 (35.7)	3 (4.1)	40 (21.9)
	6 to 12	5 (62.5)	0 (0.0)	23 (23.5)	28 (38.4)	56 (30.6)
	>12	1 (12.5)	0 (0.0)	34 (34.7)	31 (42.5)	66 (36.1)
Samples per patient	*1*	324 (100.0)	103 (100.0)	262 (78.4)	78 (54.2)	767 (84.8)
	2	0 (0.0)	0 (0.0)	66 (19.8)	22 (15.3)	88 (9.7)
	3	0 (0.0)	0 (0.0)	6 (1.8)	11 (7.6)	17 (1.9)
	4	0 (0.0)	0 (0.0)	0 (0.0)	33 (22.9)	33 (3.6)

Figure 5-3: Principal component analysis of RNA sequencing data before and after batch correction.

Plots pre-batch correction show (a) genome-wide protein-coding transcriptome; (b) selected genes comprising only the candidate signatures included in the analysis; and (c) invariant genes, stratified by source study. Panel (d) shows PCA following batch correction. ACS = adolescent cohort study; GC6-74 = Bill and Melinda Gates Foundation Grand Challenges 6-74 TB contacts study. Material from [256].



The 17 identified eligible signatures are summarised in Table 5-4 and are referred to by combining the first author's name of the corresponding publication as a prefix, with number of constituent genes as a suffix. All 17 signatures were discovered from distinct publications, apart from Suliman4 and Suliman2, which were derived from differing discovery populations within the same study. A total of five studies used existing published datasets for discovery[105,106,120,253,258], with the remainder exploiting novel data. Two signatures were discovered from paediatric populations[108,110]. Four signature discovery datasets included HIV-infected and -uninfected participants[105–108], one was discovered in an exclusively HIV-infected population for the purpose of active case finding[245], while the remainder were discovered in HIV-negative populations. Four signatures were discovered with the intention of diagnosis of incipient TB[118,120,246]. The remaining 13 discovered for diagnosis of active TB disease. Of these, five[105,110,121,253,258] targeted discrimination of TB from other diseases in addition to discriminating people who were healthy or with LTBI. Of the 17 included signatures, only three were not discovered through a genome-wide approach[110,244,253].

Table 5-4: Characteristics of candidate whole blood transcriptomic signatures for incipient TB included in systematic review and meta-analysis.

Signature	Original no. of genes	Model	Discovery population	Discovery HIV status	Discovery setting	Discovery approach	Intended application	Discovery TB cases	Discovery non-TB controls	Eligible signatures discovered[+]
Anderson38[108#]	42	Disease risk score[+]	Children	HIV positive and negative	South Africa, Malawi	Elastic net using genome-wide data	TB vs LTBI	87	43	1
BATF2[109]	1	N/A	Adults	HIV negative	UK	SVM using genome-wide data	TB vs healthy (acute vs convalescent samples)	46	31	1
Gjoen7[110]	7	LASSO regression*	Children	HIV negative	India	LASSO using 198 pre-selected genes	TB vs healthy controls and other diseases	47	36	2
Gliddon3[106]	3	Disease risk score[+]	Adults	HIV positive and negative	South Africa, Malawi[107]	Forward Selection-Partial Least Squares using genome-wide data	TB vs LTBI	285 (TB + non-TB)		1
Huang11[258#]	13	SVM (linear kernel)	Adults	HIV negative	UK[243]	Common genes from elastic net, L1/2 and LASSO models, using genome-wide data	TB vs healthy controls and other diseases	16	79	1
Kaforou25[107#]	27	Disease risk score[+]	Adults	HIV positive and negative	South Africa, Malawi	Elastic net using genome-wide data	TB vs LTBI	285 (TB + non-TB)		1
Maertzdorf4[244]	4	Random forest^	Adults	HIV negative	India	Random forest using 360 selected target genes	TB vs healthy	113	76	2
NPC2[112]	1	N/A	Adults	Not stated	Brazil	Differential expression using genome-wide data	TB vs healthy	6	28	3
Qian17[253]	17	Sum of standardised expression	Adults	HIV negative	UK[243]	Differential expression of nuclear factor, erythroid 2-like 2)-mediated genes	TB vs healthy controls and other diseases	16	69	1
Rajan5[245]	5	Unsigned sums[+]	Adults	HIV positive	Uganda	Differential expression using genome-wide data	TB vs healthy (active case finding among PLHIV)	80 totals (1:2 cases: controls)		1
Roe3[120]	3	SVM (linear kernel)	Adults	HIV negative	UK	Stability selection, using genome-wide data	Incipient TB vs healthy	46	31	1

Singhania20[121]	20	'Modified' disease risk score[%+]	Adults	HIV negative	UK, South Africa	Random forest using modular approach	TB vs healthy controls and other diseases	Discovery set not explicitly stated		1
Suliman2[7]	2	ANKRD22 - OSBPL10	Adults	HIV negative	Gambia, South Africa, Ethiopia	Pair ratios algorithm using genome-wide data	Incipient TB vs healthy	79	328	4
Suliman4[7$]	4	(GAS6 + SEPT4) - (CD1C + BLK)	Adults	HIV negative	Gambia, South Africa	Pair ratios algorithm using genome-wide data	Incipient TB vs healthy	45	141	4
Sweeney3[105]	3	(GBP5 + DUSP3) / 2 – KLF2	Adults	HIV positive and negative	Meta-analysis	Significance thresholding and forward search in genome-wide data	TB vs healthy controls and other diseases	266	931	1
Walter45[111#]	51	SVM (linear kernel)	Adults	HIV negative	USA	SVMs, using genome-wide data	TB vs LTBI	24	24	1
Zak16[118]	16	SVM (linear kernel)	Adolescents	HIV negative	South Africa	SVM-based gene pair models using genome-wide data	Incipient TB vs healthy	37	77	1

For signatures where not all constituent genes were identifiable in the RNAseq data (e.g., due to records being withdrawn), the suffix indicates the number of identifiable genes included in the current analysis.

SVM = support vector machine; LASSO = least absolute shrinkage and selection operator.

[#]Anderson38, Huang11, Kaforou25 and Walter45 included 42, 13, 27 and 51 genes in the original descriptions, respectively (genes not included in current models were either duplicates, or not identifiable in RNAseq data).

[*]Calculated using non-log transformed data using model coefficients from original publication.

[%]Calculated using non-log transformed counts per million data with trimmed mean of M-values normalization, as per original description.

^Required normalisation of the training and test sets. This was performed for each gene by subtracting the mean expression across all samples in the dataset, and dividing by the standard deviation.

[$]Modelling approach was not clear from the original description. I recreated this using two approaches; (1) as a simple equation of gene pairs ((GAS6 + SEPT4) - (CD1C + BLK)); and (2) as an SVM using the four constituent gene pairs, as previously described[124]. Since the former approach achieved marginally better performance that was closer to the authors' original description in their test set, this was included in the final analysis.

[+]Indicates total number of eligible signatures discovered in each study.

Four signatures required reconstruction of SVM models[111,118,120,258], and one required reconstruction of a random forest model[244]. My reconstructed models were validated against the authors' original descriptions by comparing AUROCs in common datasets (Table 5-5). No validation was possible for the Huang11 model as no AUROC was reported by authors in their original training or test set. The Suliman4 signature was reconstructed as a simple equation and as an SVM model using gene pairs. Since performance was marginally better and closer to the authors' AUROC in their own test set using the simple formula ((GAS6 + SEPT4) - (CD1C + BLK); AUROC=0.66), compared to the gene pairs SVM (AUROC=0.65), the simple formula approach was used in the final analysis.

Table 5-5: Validation of reconstructed signature models against the authors' original descriptions in common datasets.

Signature	Original AUROC	Reconstructed AUROC	Common dataset
Zak16	0.69	0.71	Zak test (GSE79362)
Suliman4	0.67	0.66	Suliman test (GSE94438)
Walter45	0.98	0.98	Walter test (GSE73408)
Maertzdorf4	0.98	1.00	Maertzdorf training (GSE74092)

5.3.2 Comparison of candidate signatures for detection of incipient TB over 2-year time-horizon

I first examined ROC curves and corresponding AUROCs for the identification of incipient TB by all 17 signatures over a two-year period in the separate contributing study datasets. This analysis initially suggested overall lower AUROCs in the GC6-74, compared to the ACS dataset (Supplementary Material; Chapter 8.2). However, the distribution of TB events during follow-up differed between these studies. Following stratification by interval to disease, similar AUROCs were observed between studies, suggesting that interval to disease confounded the association between source study and discriminatory performance (Supplementary Material; Chapter 8.2). Since little residual between study heterogeneity was observed and PCAs post-batch correction showed no clustering by study (Figure 5-3), I proceeded to perform a pooled data analysis without further adjustment for source study as the primary analysis.

I omitted scores for the Suliman2, Suliman4 and Zak16 signatures for samples comprising their corresponding original training sets within the GC6-74 and ACS datasets, but included scores for these signatures for all other samples. The signature with highest AUROC for the identification of incipient TB over a two-year period and that was available in pooled data from all 1,126 samples was BATF2 (AUROC 0.74; 95% CI 0.69-0.78; Table 5-6). BATF2 was therefore used as the reference standard for paired comparisons of the other 16 candidate signatures. I found that eight signatures had equivalent discrimination to BATF2. These were Suliman2, Kaforou25, Gliddon3, Sweeney3, Roe3, Zak16 and Suliman4. The remaining nine signatures had significantly inferior AUROCs.

Table 5-6: Receiver operating characteristic areas under the curve showing diagnostic accuracy of candidate transcriptomic signatures for incipient TB, stratified by time interval to disease, among pooled dataset.

P values represent comparisons against the best performing signature available for all participants (BATF2) over two years, using paired DeLong tests.

	Interval to disease (months)				
Signature	**0 to 24**	**p**	**0 to 12**	**0 to 6**	**0 to 3**
Suliman2	0.77 (0.71-0.82)	0.36	0.82 (0.76-0.88)	0.85 (0.74-0.95)	0.91 (0.86-0.96)
BATF2	0.74 (0.69-0.78)	ref	0.77 (0.72-0.82)	0.78 (0.7-0.85)	0.87 (0.79-0.95)
Kaforou25	0.73 (0.69-0.78)	0.85	0.78 (0.73-0.83)	0.79 (0.72-0.86)	0.88 (0.8-0.97)
Gliddon3	0.73 (0.68-0.77)	0.58	0.77 (0.72-0.82)	0.78 (0.71-0.85)	0.85 (0.74-0.96)
Sweeney3	0.72 (0.68-0.77)	0.44	0.77 (0.71-0.82)	0.77 (0.69-0.84)	0.91 (0.84-0.97)
Roe3	0.72 (0.67-0.77)	0.11	0.77 (0.71-0.82)	0.77 (0.7-0.84)	0.88 (0.79-0.97)
Suliman4	0.7 (0.64-0.76)	0.26	0.73 (0.66-0.8)	0.78 (0.68-0.89)	0.82 (0.69-0.94)
Zak16	0.7 (0.64-0.76)	0.94	0.76 (0.69-0.82)	0.79 (0.71-0.86)	0.86 (0.71-1)
NPC2	0.68 (0.64-0.73)	0.012	0.71 (0.66-0.77)	0.75 (0.69-0.82)	0.78 (0.66-0.9)
Maertzdorf4	0.68 (0.63-0.73)	0.001	0.73 (0.68-0.78)	0.71 (0.64-0.79)	0.8 (0.69-0.91)
Gjoen7	0.67 (0.63-0.72)	0.001	0.69 (0.64-0.75)	0.7 (0.62-0.77)	0.75 (0.61-0.88)
Singhania20	0.67 (0.62-0.72)	0.006	0.68 (0.62-0.73)	0.72 (0.65-0.78)	0.74 (0.6-0.87)
Huang11	0.66 (0.61-0.71)	0.007	0.7 (0.65-0.75)	0.67 (0.6-0.75)	0.75 (0.63-0.86)
Qian17	0.66 (0.61-0.71)	<0.0001	0.71 (0.66-0.76)	0.69 (0.62-0.77)	0.79 (0.7-0.88)
Anderson38	0.65 (0.61-0.7)	0.002	0.68 (0.62-0.73)	0.68 (0.6-0.75)	0.74 (0.63-0.85)
Rajan5	0.55 (0.5-0.6)	<0.0001	0.59 (0.53-0.65)	0.57 (0.49-0.66)	0.68 (0.56-0.81)
Walter45	0.55 (0.5-0.6)	<0.0001	0.58 (0.52-0.64)	0.62 (0.54-0.69)	0.47 (0.35-0.6)

5.3.3 Correlation, Jaccard indices and upstream analysis of candidate signatures

Next, I examined the correlation between the 17 candidate signature scores in the pooled dataset, as defined by Spearman rank correlation (Figure 5-4). The eight signatures identified with equivalent performance demonstrated moderate to high correlation. In contrast, Singhania20, Anderson38, Huang11 and Walter45 showed little correlation with any other signature. The correlation matrix dendrogram showed closest relationships between signatures identified by the same research group. To assess whether correlation was driven by overlapping constituent genes, I calculated pairwise Jaccard Indices. There was a weak positive association between Spearman rank correlation and Jaccard Index, suggesting that overlapping constituent genes may partially account for their correlation. The 40 genes comprising the eight signatures with equivalent AUROCs are demonstrated in Figure 5-5a. Upstream analysis predicted that interferon-gamma, interferon-alpha, STAT1 (the canonical mediator of interferon signalling), and tumour necrosis factor were the strongest predicted transcriptomic regulators of these constituent genes (Figure 5-5b).

Figure 5-4: Relationships between 17 candidate transcriptomic signatures for incipient TB.

Displayed as: (a) Spearman rank correlation matrix heatmap; (b) Jaccard similarity index heatmap showing overlapping constituent genes; and (c) Jaccard index vs. Spearman correlation coefficient for pairwise signature comparisons. Material from [256].

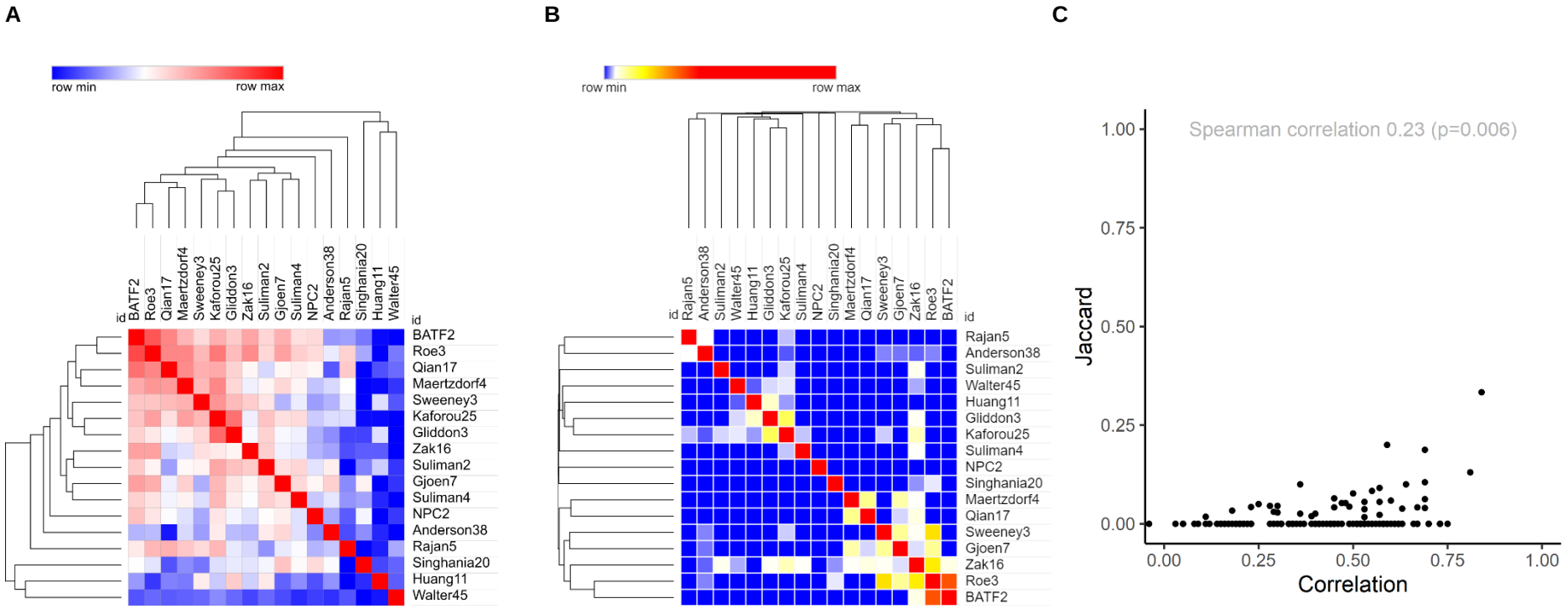
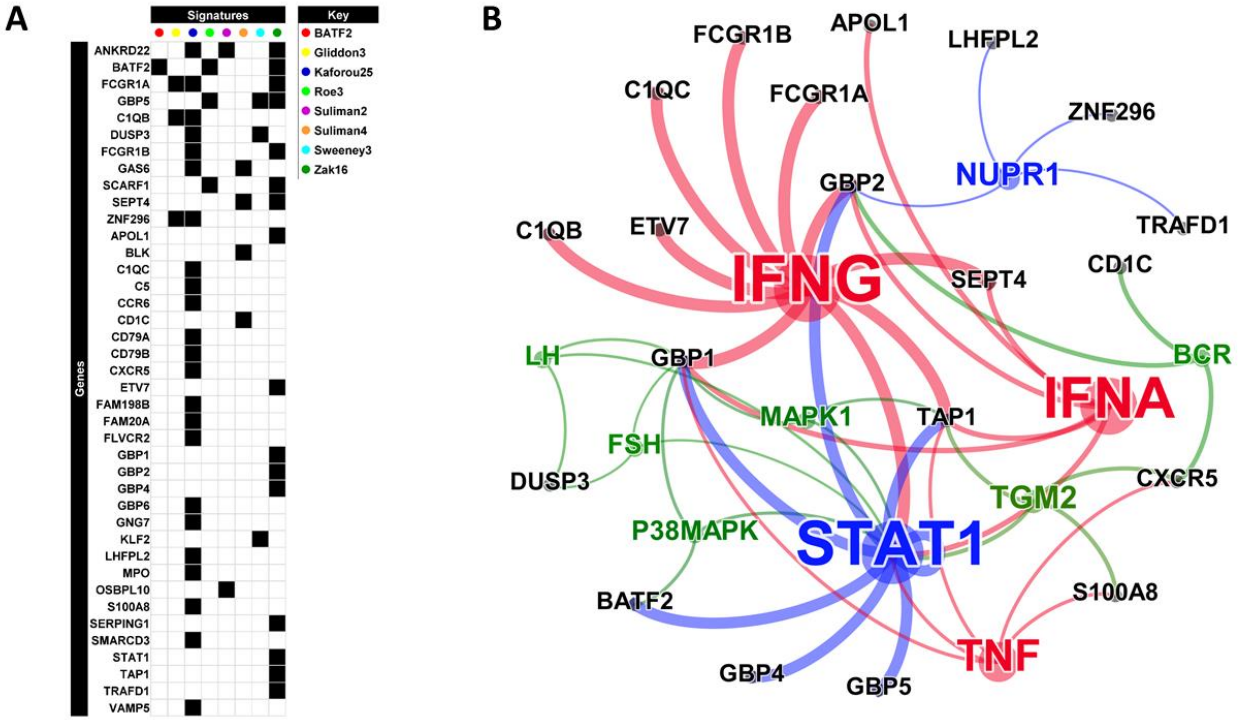


Figure 5-5: Genes comprising the top eight blood transcriptomic signatures for incipient TB.

Shown as (a) chessboard matrix; and (b) network diagram. Network diagram shows predicted upstream regulators of the 40 genes. Coloured nodes represent the predicted upstream regulators, grouped by function (red=cytokine, blue=transcription factor, green=other). Grey nodes represent the transcriptomic biomarkers downstream of these regulators. Size of the nodes is proportional to the number of downstream genes associated with each regulator and the thickness of the edges is proportional to the -log10 P value for enrichment of each of the upstream regulators. Material from [256].



5.3.4 Diagnostic accuracies of candidate signatures decline with interval to disease

The distributions of the eight best performing signatures among the IGRA-negative control population followed an approximately Normal distribution prior to Z score transformation (Supplementary Material; Chapter 8.2). Z scores for the eight best performing signatures, stratified by interval to disease, are shown in Figure 5-6 and Figure 5-7. AUROCs of these signatures declined with increasing interval to disease (Figure 5-8), ranging 0.82-0.91 for 0-3 months vs. 0.70-0.77 for 0-24 months.

Figure 5-6: Scatterplots showing scores of eight best performing transcriptomic signatures for incipient TB, stratified by interval to disease.

Dashed horizontal lines indicate Z2 thresholds for each signature. NP = non-progressors, who remained healthy during follow-up. Material from [256].

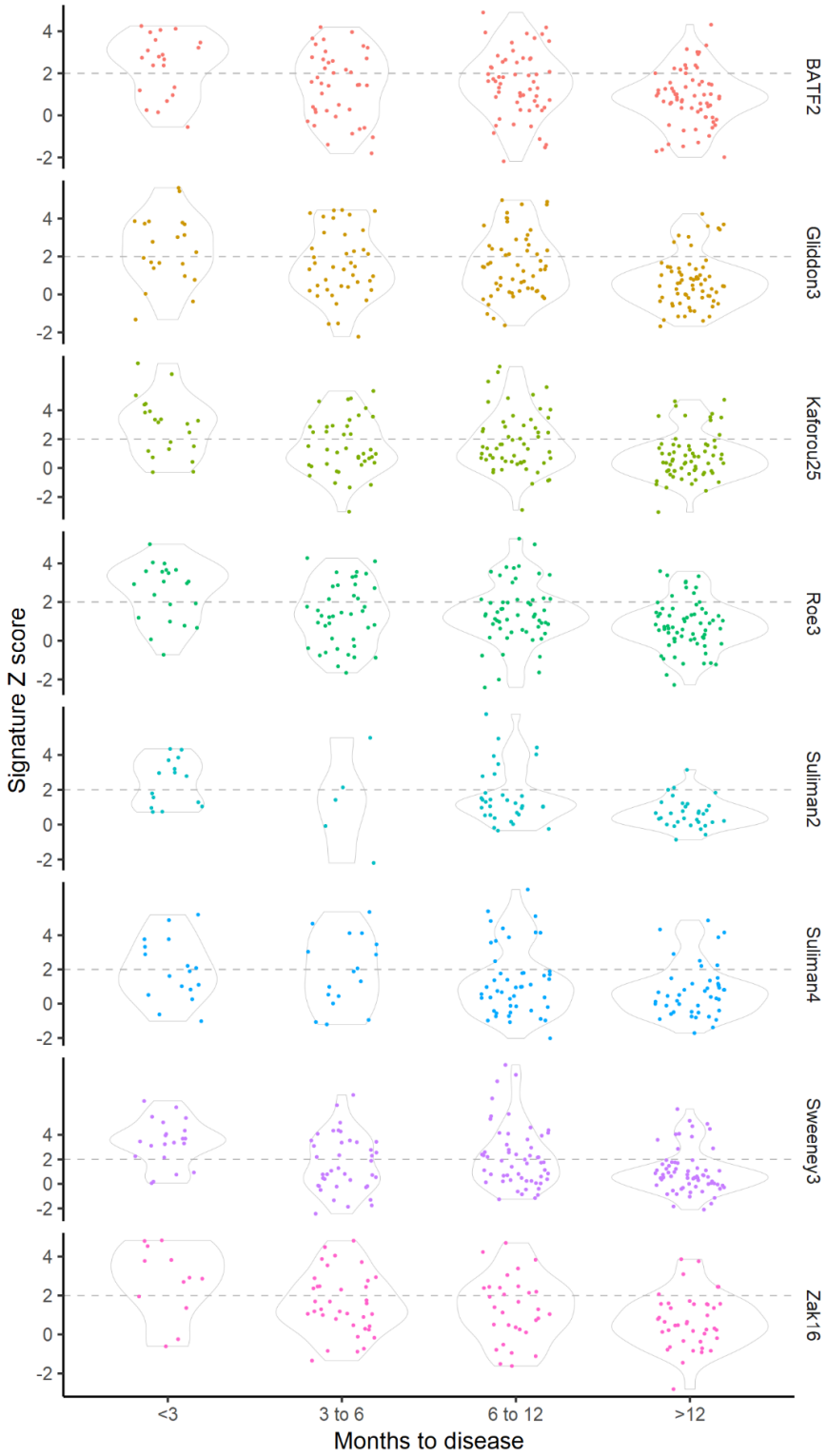


Figure 5-7: Expression of eight best performing transcriptomic signatures among participants progressing to TB by days from sampling to disease.

Horizontal dashed line indicates cut-off defined by two standard deviations above the mean of control population. Vertical grey dashed line indicates the initial 90 day interval after sampling. Material from [256].

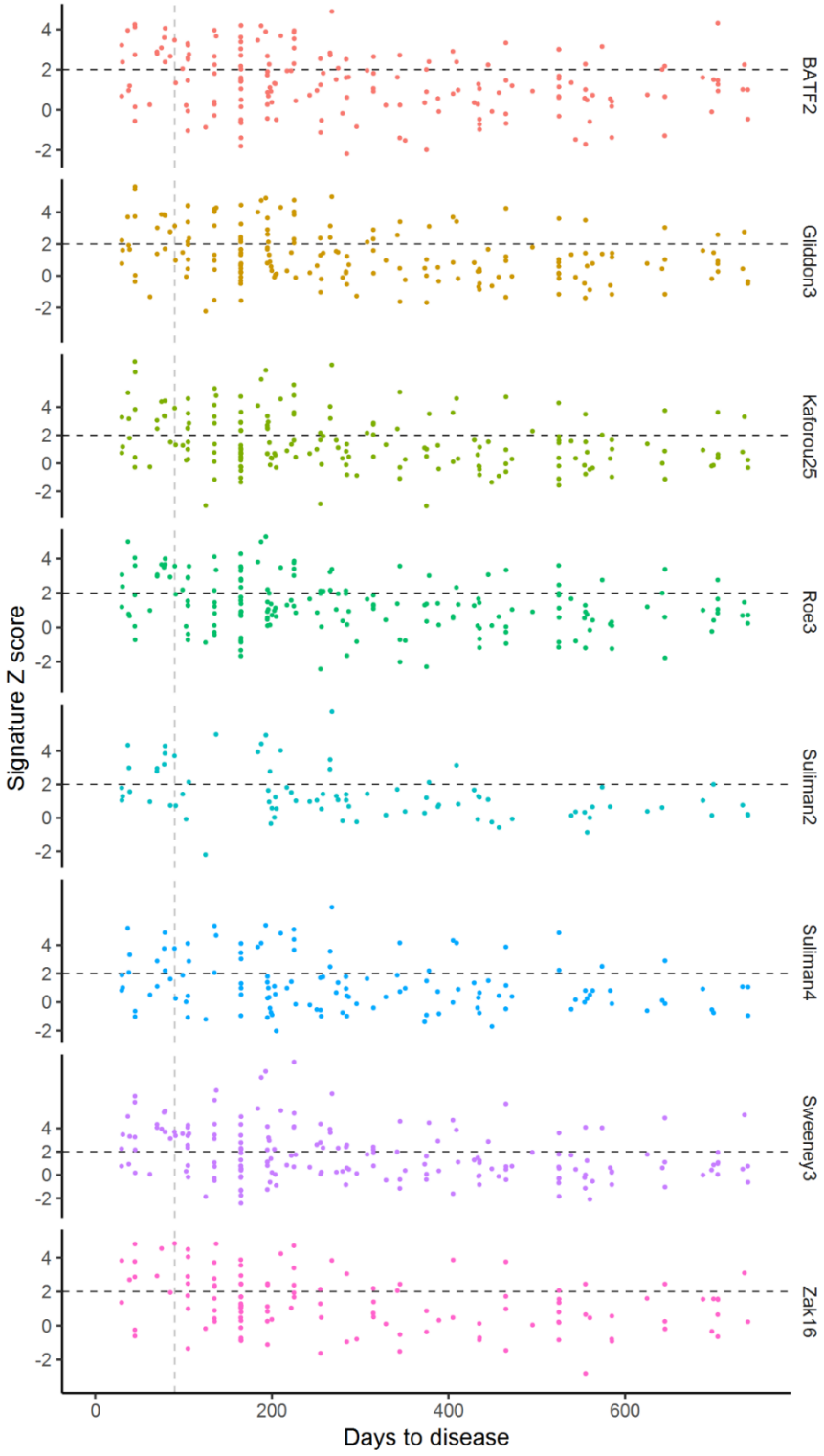


Figure 5-8: ROC curves showing diagnostic accuracy of eight best performing transcriptomic signatures for incipient TB.

Curves are shown stratified by months from sample collection to disease. Material from [256].

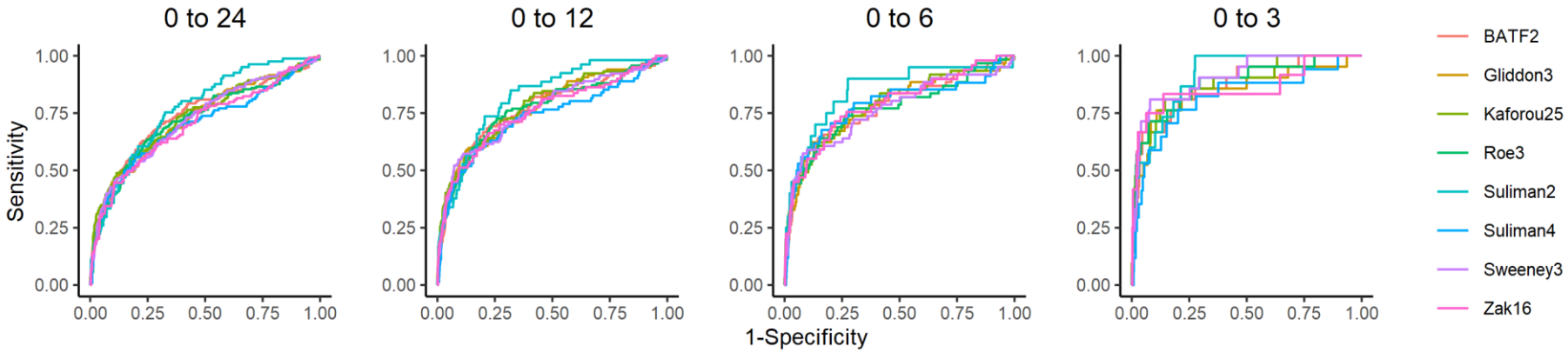
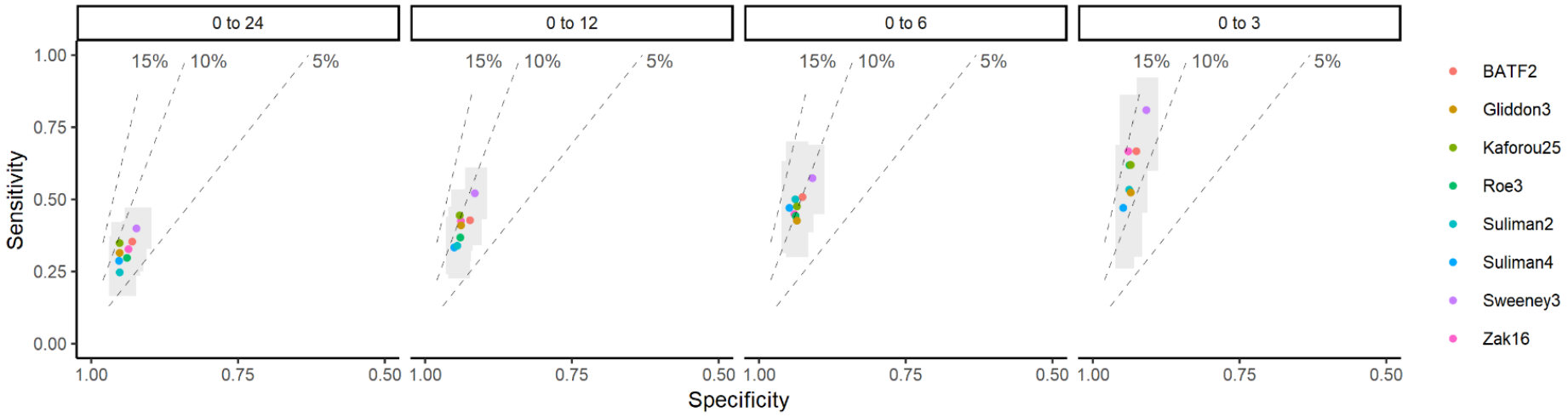


Figure 5-9a demonstrates diagnostic accuracy of the eight best performing candidates using pre-specified Z2 cut-offs based on the 97.7th percentile of the IGRA-negative control population, stratified by interval to disease and benchmarked against positive predictive value estimates based on a pre-test probability of 2%. At this threshold, test sensitivities over 0-24 months of the eight best performing signatures ranged from 24.7% (16.6-35.1) to 39.9% (33.0-47.2) for the Suliman2 and Sweeney3 signatures, respectively, while corresponding specificities ranged from 92.3% (89.8-94.2) to 95.3% (92.3-96.9). In contrast, over a 0-3 month interval, sensitivities ranged from 47.1% (26.2-69.0) for the Suliman4 signature to 81.0% (60.0-92.3) for the Sweeney3 signature, with corresponding specificities of 90.9% (88.9-92.6) to 94.8% (93.0-96.2). For each of the time points, the eight signatures had overlapping confidence intervals, and largely fell in the same positive predictive value plane (5-10% over 0-24 months *vs.* 10-15% over 0-3 months). Using cut-offs defined by the maximal Youden index for each time interval, sensitivity and specificity estimates fell below the minimum WHO TPP criteria for incipient TB tests over a 0-24 month period, but met or approximated the minimum criteria over 0-3 months (Figure 5-9b).

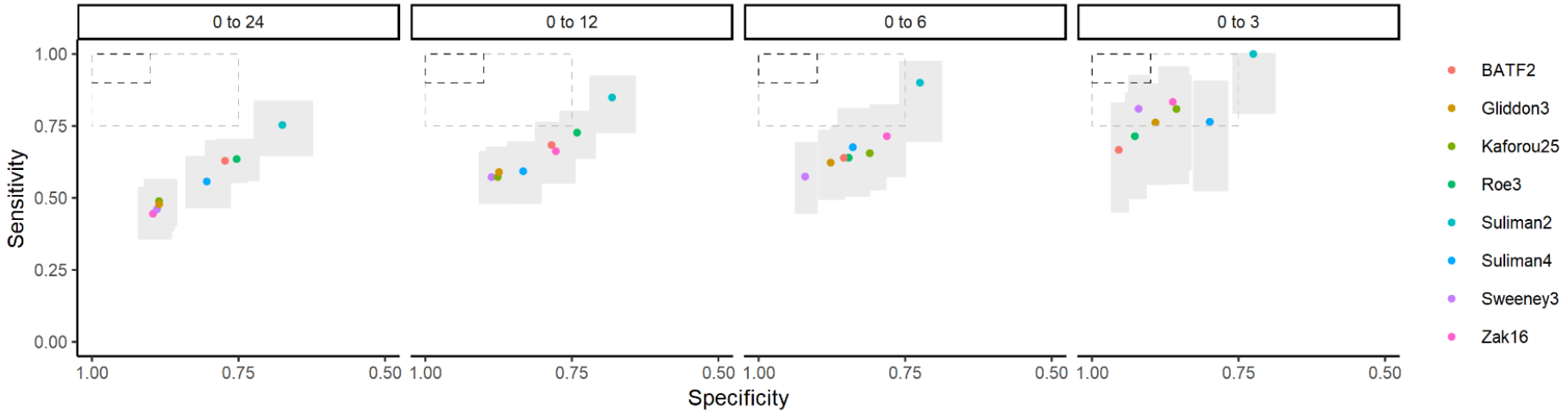
Figure 5-9: Diagnostic accuracy of eight best performing transcriptomic signatures for incipient TB.

Shown in receiver operating characteristic space, stratified by months to disease. Grey shaded zones indicate 95% CIs for each signature. (a) Dashed lines represent positive predictive value planes of 5, 10 and 10%, respectively, based on 2% pre-test probability, and using Z2 cut-offs. (b) Secondary analysis presented using biomarkers cut-offs defined by the maximal Youden indices for each time period, benchmarked against minimal (grey dashed box) and optimal (black dashed box) criteria from the WHO Target Product Profile for incipient TB biomarkers. Material from [256].

(a)



(b)

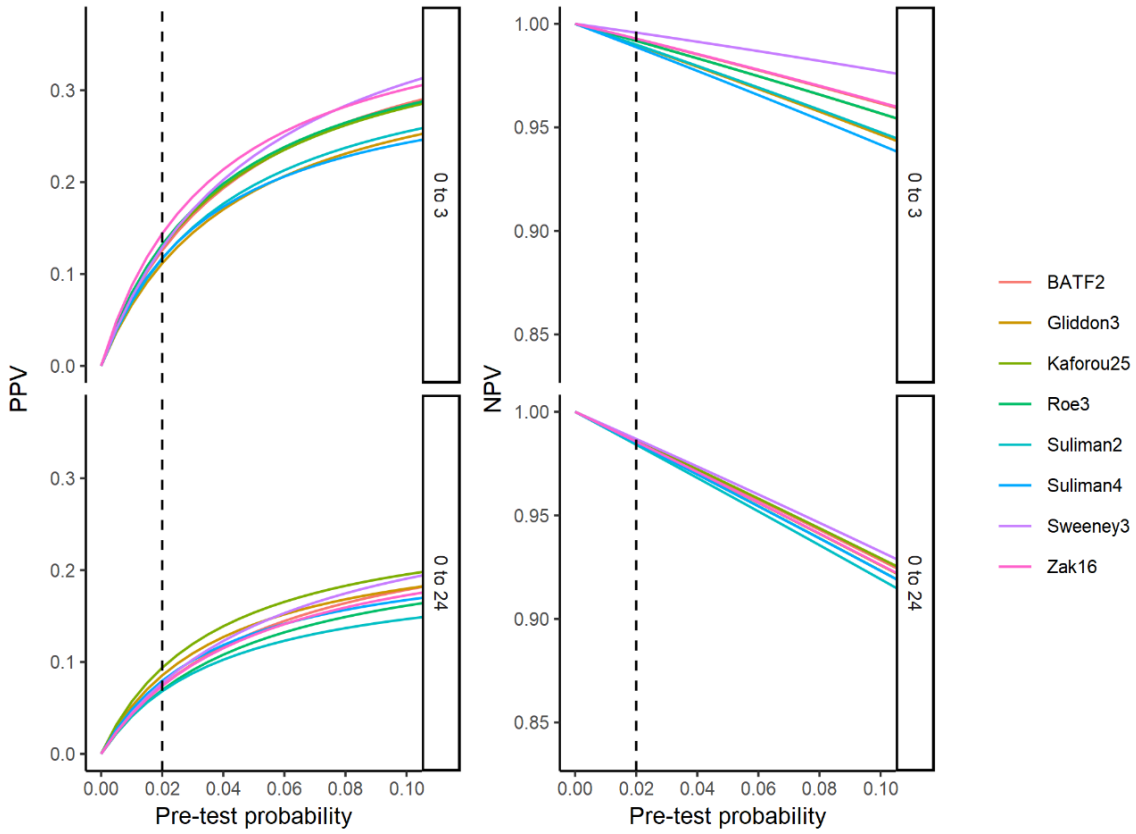

0 to 24
0 to 12
0 to 6
0 to 3
Sensitivity
1.00
0.75
0.50
0.25
0.00
Specificity
1.00
0.75
0.50
BATF2
Gliddon3
Kaforou25
Roe3
Suliman2
Suliman4
Sweeney3
Zak16

5.3.5 Predictive values for incipient TB

Positive- and negative-predictive values, using Z2 cut-offs and modelled across a range of pre-test probabilities, are shown in Figure 5-10. Based on pre-test probability of 2% all eight top performing signatures achieved a positive predictive value marginally above the WHO benchmark of 5.8% for a 0-24 month period, ranging from 6.8% for Suliman2 to 9.4% for Kaforou25, with corresponding negative predictive values of 98.4% and 98.6%, respectively. For the 0-3 month time period, positive predictive values ranged from 11.2% for Gliddon3 to 14.4% for Zak16, with corresponding negative predictive values of 99.0% and 99.3%, respectively.

Figure 5-10: Positive- and negative-predictive values (PPVs/NPVs) for best performing transcriptomic signatures for incipient TB.

Shown stratified by months to disease, using pre-specified Z2 cut-offs based on the 97.7th percentile of the control population, across a range of pre-test probabilities. Dashed line indicates 2% pre-test probability. Material from [256].



5.3.6 Sensitivity analyses

Restricting inclusion of incipient TB cases to those with documented microbiological confirmation and including only one blood RNA sample per participant (by random sampling) produced no significant change to the main results (Figure 5-11, Figure 5-12). Re-analysis of the ROC curves using mutually exclusive time periods of 0-3, 3-6, 6-12 and 12-24 months magnified the difference in performance between the intervals, with performance declining more markedly with increasing interval to disease (Table 5-7). AUROCs in the 12-24 month interval ranged from 0.60 (0.50-0.70) to 0.67 (0.60-0.75) for the top eight equivalent signatures. Finally, the two-stage meta-analysis approach (using data without batch correction) showed similar findings to the primary analysis (Figure 5-13).

Figure 5-11: ROC curves showing diagnostic accuracy of eight best performing transcriptomic signatures for incipient TB - sensitivity analysis restricting incipient TB cases to microbiologically confirmed.

Stratified by months from sample collection to disease. Material from [256].

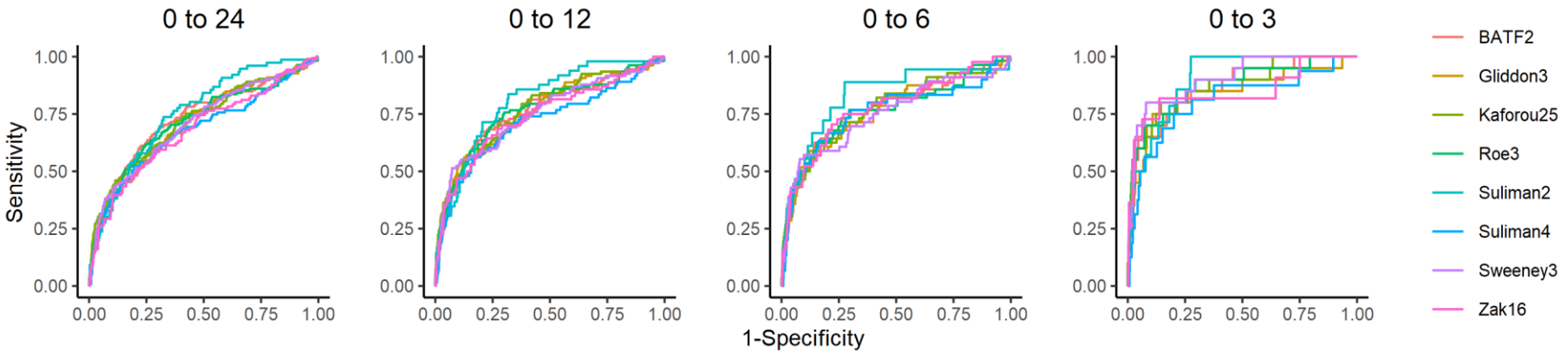


Figure 5-12: ROC curves showing diagnostic accuracy of eight best performing transcriptomic signatures for incipient TB - sensitivity analysis including only one blood RNA sample per participant.

One sample per participant selected using random sampling. Performance stratified by months from sample collection to disease. Material from [256].

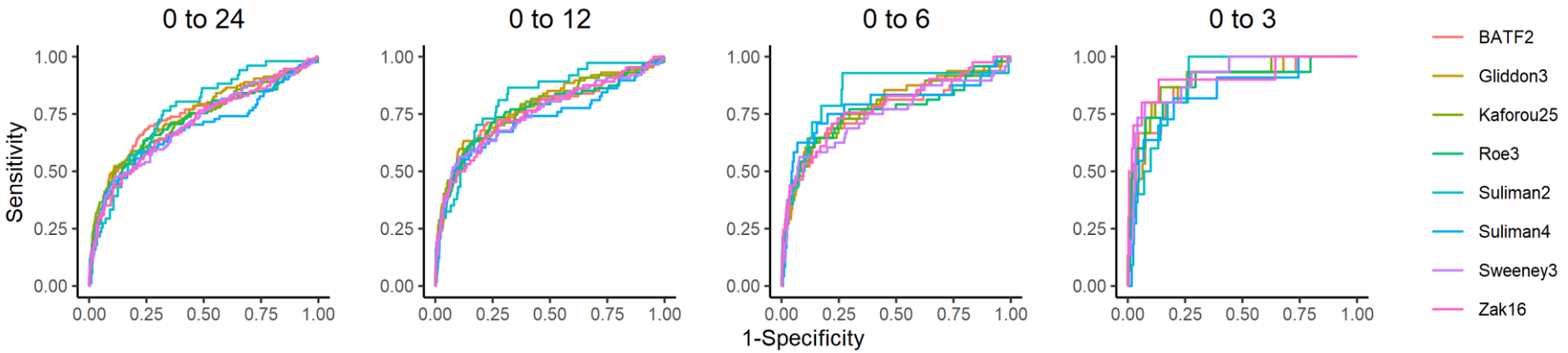


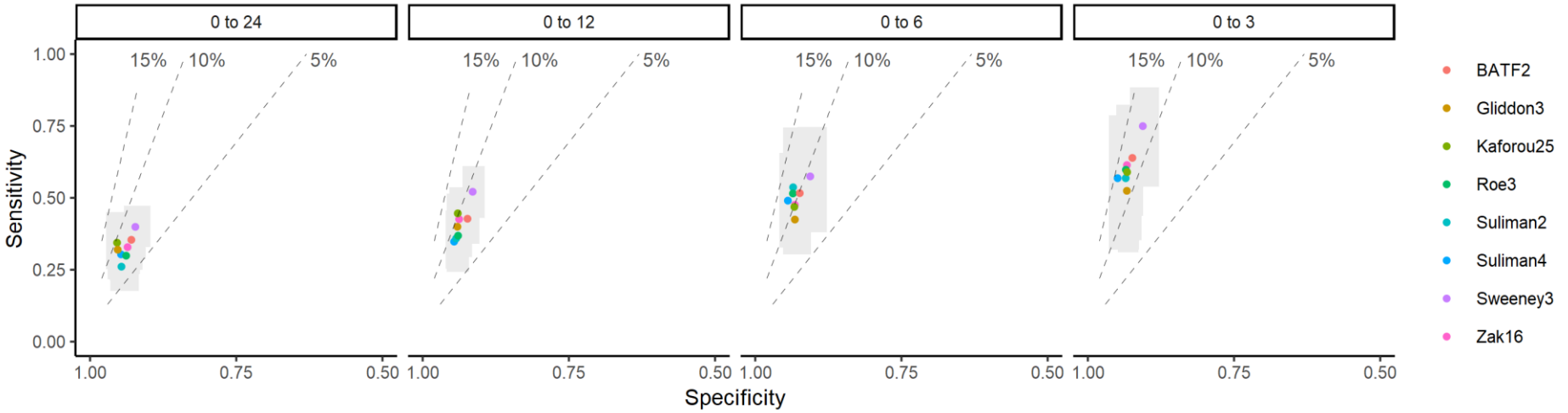
Table 5-7: AUROCs (95% CIs) showing diagnostic accuracy of eight best performing transcriptomic signatures for incipient TB using mutually exclusive time periods.

Stratified by months from sample collection to disease.

Signature	0 to 3	3 to 6	6 to 12	12 to 24
Suliman2	0.91 (0.86 - 0.96)	0.66 (0.29 - 1)	0.8 (0.74 - 0.87)	0.67 (0.58 - 0.76)
Sweeney3	0.91 (0.84 - 0.97)	0.69 (0.59 - 0.8)	0.76 (0.69 - 0.83)	0.63 (0.56 - 0.71)
Kaforou25	0.88 (0.8 - 0.97)	0.74 (0.65 - 0.83)	0.78 (0.71 - 0.85)	0.63 (0.55 - 0.71)
Roe3	0.88 (0.79 - 0.97)	0.71 (0.61 - 0.81)	0.77 (0.69 - 0.84)	0.64 (0.56 - 0.72)
BATF2	0.87 (0.79 - 0.95)	0.73 (0.63 - 0.82)	0.76 (0.69 - 0.84)	0.67 (0.6 - 0.75)
Zak16	0.86 (0.71 - 1)	0.77 (0.68 - 0.85)	0.71 (0.6 - 0.82)	0.6 (0.5 - 0.7)
Gliddon3	0.85 (0.74 - 0.96)	0.75 (0.66 - 0.84)	0.76 (0.69 - 0.83)	0.62 (0.55 - 0.7)
Suliman4	0.82 (0.69 - 0.94)	0.75 (0.59 - 0.91)	0.69 (0.6 - 0.78)	0.63 (0.53 - 0.73)
Maertzdorf4	0.8 (0.69 - 0.91)	0.67 (0.57 - 0.76)	0.75 (0.67 - 0.82)	0.59 (0.51 - 0.67)
Qian17	0.79 (0.7 - 0.88)	0.64 (0.54 - 0.74)	0.73 (0.66 - 0.8)	0.57 (0.48 - 0.65)
NPC2	0.78 (0.66 - 0.9)	0.74 (0.66 - 0.82)	0.67 (0.59 - 0.76)	0.62 (0.55 - 0.7)
Huang11	0.75 (0.63 - 0.86)	0.63 (0.54 - 0.73)	0.73 (0.67 - 0.79)	0.55 (0.47 - 0.63)
Gjoen7	0.75 (0.61 - 0.88)	0.67 (0.59 - 0.76)	0.69 (0.6 - 0.77)	0.62 (0.55 - 0.69)
Anderson38	0.74 (0.63 - 0.85)	0.64 (0.55 - 0.74)	0.68 (0.6 - 0.76)	0.6 (0.52 - 0.68)
Singhania20	0.74 (0.6 - 0.87)	0.7 (0.63 - 0.77)	0.64 (0.56 - 0.72)	0.64 (0.56 - 0.72)
Rajan5	0.68 (0.56 - 0.81)	0.52 (0.42 - 0.62)	0.6 (0.52 - 0.68)	0.5 (0.43 - 0.58)
Walter45	0.47 (0.35 - 0.6)	0.66 (0.58 - 0.75)	0.54 (0.46 - 0.62)	0.49 (0.41 - 0.57)

Figure 5-13: Diagnostic accuracy of eight best performing transcriptomic signatures for incipient TB from two-stage IPD-MA sensitivity analysis.

Calculated using bivariate random effects meta-analysis. Shown in receiver operating characteristic space, stratified by months to disease. Grey shaded zones indicate 95% Cis for each signature. Cut-offs derived from two standard scores above the mean of control population. Dashed lines represent positive predictive value planes of 5, 10 and 10%, respectively, based on 2% pre-test probability. Material from [256].



5.4 Discussion

5.4.1 Summary of key findings

In a pooled dataset from four published genome-wide RNA sequencing studies, I have demonstrated that eight candidate transcriptomic signatures performed with equivalent diagnostic accuracy for detection of incipient TB over a two-year time horizon. These signatures ranged from a single transcript (BATF2) to 25 genes (Kaforou25). The accuracy of all eight signatures declined markedly with increasing intervals to disease. These signatures only marginally surpassed the WHO target positive predictive value of 5.8% over two years, assuming 2% pre-test probability and using a cut-off of two standard scores (Z2). However, sensitivity at the Z2 threshold was only 24.7-39.9%, missing the majority of cases. No signature achieved the WHO target sensitivity and specificity of ≥75% over two years, even when using the cut-off with maximal accuracy. In contrast, using Z2 cut-offs over a 0-3 month period, the eight best performing signatures achieved sensitivities of 47.1-81.0% and specificities >90%. This led to positive- and negative-predictive values of 11.2-14.4% and >98.9%, respectively, when assuming 2% pre-test probability, suggesting that the minimum WHO TPP may be achieved over shorter time intervals. These findings are consistent with the recently reported CORTIS trial, which showed that the performance of a transcriptomic signature was highly time-dependent when implemented for general population screening[123].

5.4.2 Policy implications

In order to achieve the WHO TPP, a screening strategy that incorporates serial testing on a 3-6 monthly basis may therefore be required for transcriptomic signatures. Such a strategy, however, is unlikely to be feasible or cost-effective at general population level, where the pre-test probability of incident TB is relatively low. Instead, high-risk groups such as household contacts could be targeted. However, even this may be challenging in high transmission settings, given the limited global coverage of contact-tracing programmes. In lower transmission, higher resource settings, serial blood transcriptomic testing for risk-stratification over a defined 1-2 year period may be more achievable, particularly among recent contacts or new entry migrants from high transmission countries,

for whom risk of disease is highest within an initial two year interval[16,80,120]. Integral to scale-up of these biomarkers is translation of transcriptomic measurements from genome-wide approaches to the reproducible quantification of selected signature genes, with appropriately defined cut-offs. This is underway for the Sweeney3 signature, using the GeneXpert platform[259].

5.4.3 Characteristics of best performing signatures and implications for biomarker discovery

The eight signatures that achieved equivalent performance were discovered with the primary intention of diagnosis of incipient TB[118,120,246], or differentiating active TB from people who are healthy or with LTBI[105–107,109]. Discovery populations for these eight signatures included adults or adolescents from the UK or sub-Saharan Africa[106,107,109,118,120,246], or a meta-analysis of microarray data from multiple studies[105], including a minimum of 37 incipient or active TB cases. All eight signatures were discovered using genome-wide approaches. In contrast, the nine signatures with inferior performance included two discovered from children[108,110], one study that prioritised discrimination of active TB from other bacterial and viral infections[121], and one study that conducted active case-finding for TB among people living with HIV[245]. The differences in primary intended applications, which are reflected in the study populations used for biomarker discovery, may account for their inferior performance when evaluated solely for identification of incipient TB in a predominantly healthy, HIV-negative adult population. The signatures with inferior performance also included three discovered from panels of pre-selected candidate genes, rather than a genome-wide approach[110,244,253], and four with only 6-24 TB cases in the discovery sets[111,112,253,258]. Taken together, these findings suggest that using a genome-wide approach and including adequate numbers of diseased cases are important considerations during signature discovery.

Previous reviews have also highlighted that original biomarker discovery studies often produce optimistic discrimination measures, which cannot be replicated in independent validation studies, thus meaning that the vast majority of biomarkers never reach clinical implementation[103]. In addition to ensuring adequate sample size, future biomarker studies should also ensure that study cohorts included in discovery and validation are representative of target populations for

implementation. Such studies should also ideally be performed using a cohort study, as opposed to case-control, design to reduce risk of spectrum bias caused by the inclusion of more extreme healthy and diseased phenotypes[122].

The eight best performing signatures were derived from the application of different computational approaches, but showed moderate to high levels of co-correlation, with closest relationships between signatures identified by the same research group. This likely reflects common discovery datasets and modelling approaches used within research groups. Overlapping constituent genes only partially accounted for correlation between signatures, suggesting that they reflect different dimensions of a common host response to infection with *M. tuberculosis*. This hypothesis was strongly supported by the identification of interferon and tumour necrosis factor signalling pathways as statistically-enriched upstream regulators of the genes across the eight signatures. Whilst these host response pathways are not likely to be specific to TB, the application of these biomarkers for incipient TB mitigates against the limitations of imperfect specificity by focusing on asymptomatic individuals in which the prior probability of other diseases is low. The time-dependent sensitivity of the signatures suggests that the duration of the incipient phase of TB is typically 3-6 months. However, even within the <3 month time interval, the sensitivity of the best performing transcriptomic signatures ranged from 47.1-81.0%, indicating that the biomarkers may genuinely have imperfect sensitivity for incipient TB, or that the incipient phase can progress very rapidly among a subset of cases. Importantly, each signature did exhibit an AUROC >0.5 for discriminating incipient TB from non-progressors even 12-24 months after sampling, suggesting that the incipient phase may be more prolonged in some cases. This may reflect cases in which the host response initially achieves mycobacterial control in dynamic host-pathogen interactions[260]. Of interest, these findings are generally mirrored in proteomic and metabolomic data from similar cohorts[126,127].

5.4.4 Strengths of this study

Strengths of this study include the size of the pooled dataset, including 1,126 samples from 905 patients, and 183 samples from 127 incipient TB cases. Participant-level data were available for all four eligible studies, all of which achieved maximal quality assessment scores and were performed in relevant

target populations of either recent TB contacts, or people with LTBI. This facilitated a robust analysis of diagnostic accuracy of the candidate signatures, stratified by interval to disease. Second, I performed a comprehensive systematic review, and identified 17 candidate signatures. For each of these signatures, gene lists and modelling approaches were extracted and verified by independent reviewers. Moreover, for signatures that required model reconstruction, my models were cross-validated against original models by comparing AUROCs using the same dataset wherever possible. This allowed us to perform a comprehensive, head-to-head analysis of candidate signatures for incipient TB, ensuring that each head-to-head comparison was performed on paired data. Finally, my meta-analytic methods ensured a standardised approach to processing of raw RNA sequencing data, with consistent results obtained from one-stage and two-stage IPD-MA approaches.

5.4.5 Study limitations

A weakness of my analysis is that the total sample size included 183 samples from 127 incipient TB cases and I was unable to perform subgroup analyses by age, ethnicity or country, since the contributing studies largely defined these strata. Reassuringly, there were no clear differences in performance by study, supporting the generalisability of the results, although additional validation studies are required. Targeted PCR-based approaches to quantity the most promising candidate signatures could supplant use of genome-wide RNA sequencing to reduce costs and increase scalability. The accuracy of transcript measurements using near-patient technology, such as the GeneXpert platform, could also be evaluated[259]. I was also unable to account for prior BCG vaccination status, though I anticipate that BCG coverage would likely be very high among the study populations included.

Having observed little heterogeneity between studies, I conducted a pooled analysis, assuming common diagnostic accuracy between studies. The precision of my estimates therefore may be slightly overstated and statistical tests may be anti-conservative. However, sensitivity analysis using a two-stage meta-analysis approach with random effects yielded similar findings, supporting the robustness of my results. Likewise, treating serial samples as independent was anti-conservative and may have led to bias if participants with serial samples were

systematically different from those with single samples. However, my findings were similar in my sensitivity analysis when including only one sample per individual at random.

All included datasets were from sub-Saharan Africa and the UK, where a significant proportion of participants were of Asian ethnicity. No data were available for PLHIV or children under 10, among whom different blood transcriptomic perturbations may occur in TB[108,131]. Prospective validation studies in other world regions and among these specific target populations are needed, and could be used to periodically update the current meta-analysis to further increase generalisability.

There were only eight TB cases known to be extra-pulmonary, thus precluding assessment of diagnostic accuracy stratified by TB disease site. In addition, the majority of incipient TB cases were contributed from the African datasets, with 12 cases from the UK studies. Nevertheless, the UK studies were done in appropriate target populations of close contacts of TB index cases and were performed as cohort studies, as opposed to the African study case-control designs. High specificity for correctly identifying non-progressors among contacts is a key attribute in improving positive predictive value, compared to existing tests. Hence, these UK datasets were useful additions to the pooled meta-analysis, though further data are required to confirm that signature sensitivities observed in the African case-control studies generalises to settings with lower TB incidence.

Non-progressor samples with less than 6 months' follow-up were excluded due to risk of outcome misclassification, since I considered 6 months to be an insufficient follow-up duration for TB development. Finally, participants who received preventative treatment were also excluded, which may have led to selection bias if they were systematically different from those who did not. However, these exclusions applied to only 30 and 35 samples, respectively. The impact of any selection bias is likely small, although the potential directions of these biases are difficult to predict and could have led to under- or over-estimation of accuracy for detecting incipient TB.

5.4.6 Conclusion

In summary, I have demonstrated that eight transcriptomic signatures, including a single transcript (BATF2), have equivalent diagnostic accuracy for identification of incipient TB. Performance appeared similar across studies, including participants from the UK and sub-Saharan Africa. Signature performance was highly time-dependent, with lower accuracy at longer intervals to disease. A screening strategy that incorporates serial testing on a 3-6 monthly basis among carefully selected high-risk groups may be required for these biomarkers to surpass the WHO TPP benchmarks.

5.5 Contribution statement

This chapter included a systematic review and pooled analysis of four publicly available RNA sequencing datasets. I led the work from conception to completion and dissemination.

5.6 Outputs relating to this chapter

This study is published in *The Lancet Respiratory Medicine*:

Gupta RK, Turner CT, Venturini C, Esmail H, Rangaka MX, Copas A, Lipman M, Abubakar I, Noursadeghi M (2020). Concise whole blood transcriptional signatures for incipient tuberculosis: A systematic review and patient-level pooled meta-analysis. **The Lancet Respiratory Medicine.** https://doi.org/10.1016/S2213-2600(19)30282-6

I also presented this work at the:

- TBScience Symposium (pre-conference to the Union World TB Conference) in Hyderabad, India, October 2019 - receiving a Gates Foundation Award.
- British Thoracic Society Winter Meeting, London, December 2019.

In addition, the systematic review and analysis pipeline underpinning this work directly led to a second paper evaluating the performance of candidate blood transcriptomic signatures for diagnosis of TB disease in the publication below in *The Lancet Respiratory Medicine*:

Turner CT, **Gupta RK**, Tsaliki E, Roe J, Mondal P, Nyawo G, Palmer Z, Miller RF, Reeve B, Theron G, Noursadeghi M (2020). Systematic validation of blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective diagnostic accuracy study. **The Lancet Respiratory Medicine**. https://doi.org/10.1016/S2213-2600(19)30469-2

6 Discussion

6.1 Summary of key findings in context of wider literature

In this thesis, I evaluated a range of methods to facilitate more precise targeting of preventative treatment for TB, using existing and next generation biomarkers. The key findings are summarised in the paragraphs below. Detailed critique of the strengths and limitations of these analyses were discussed in the relevant chapters.

6.1.1 Prognostic value of TST and IGRAs for incident TB

In Chapters 2 and 4, I demonstrated that the TST and IGRAs (including QFT-GIT and T-SPOT.TB) have similar predictive ability for incident TB when using either binary cut-offs (particularly for TST_{15} or $TST_{5/15}$) and quantitative results in settings with low TB incidence. For both TST and IGRAs, stronger T cell responses are associated with higher risk of incident TB, in keeping with previous data[84,85]. However, implementation of higher cut-offs is not likely to be a tractable approach in settings aiming towards pre-elimination due to a marked loss of sensitivity, with only modest gain in positive predictive value when compared to standard thresholds. The newer generation QFT-Plus assay is also likely to perform similarly to existing iterations, in view of strong agreement and correlation between the two TB antigen response tubes (Chapter 3).

It is widely accepted that both TSTs and IGRAs do not discriminate persistent from cleared *M. tuberculosis* infection (Chapter 1.4.7); both tests therefore have limited specificity for detecting people at highest risk of progression to disease. In addition, they also have imperfect sensitivity for detecting people who progress to incident TB, as demonstrated in Chapters 2 and 3. This finding is consistent with previous data describing TST and IGRA sensitivities ranging from 65-90% even in the context of prevalent TB disease (Chapter 1.4.3). Negative TST and IGRAs in these examples may reflect susceptibility to TB disease among a subgroup of exposed individuals who either fail to mount any adaptive immune response to *M. tuberculosis*, or develop an immune response that is independent of interferon-gamma[75]. Taken together, the evidence suggests that TST and IGRA should be considered as useful but imperfect correlates of risk when used

for TB diagnosis, prognostication, or as outcome measures in vaccine efficacy studies[240].

6.1.2 Interpreting the quantitative T cell response in context

Consistent with findings from previous systematic reviews and meta-analyses[76–78], I found that the risk of TB among people with a conventional diagnosis of LTBI was highly variable, even when stratified by indication for screening (Chapter 4). By integrating quantitative T cell responses (measured by TST or IGRA) with demographics, history of exposure to *M. tuberculosis* and measures of immunocompromise, the PERISKOPE-TB multivariable prognostic model facilitates interpretation of LTBI test results within the context of individual patients in settings with low TB incidence. Despite differences in baseline risk across studies resulting in imperfect calibration estimates, decision curve analyses demonstrate that the tool is likely to have clinical utility to guide shared decision making regarding the initiation of preventative treatment, across a broad range of clinician and patient preferences.

6.1.3 Blood transcriptomic signatures for incipient TB

In Chapter 5, I showed that eight blood transcriptomic signatures have similar accuracy for detection of incipient TB. These signatures reflect a common underlying host response in TB, predominantly driven by interferon and tumour necrosis factor-inducible gene expression, suggesting that there are multiple transcriptomic targets that could be amenable to translation to clinical diagnostics. However, since these signatures reflect the host immune response in early TB, they also reflect short-term risk of disease and only meet the WHO TPP benchmarks for an incipient TB test over a 3-6 month interval (as opposed to the 2-year time horizon specified in the original TPP[11]).

In addition, two of the contributing studies in my IPD-MA were case-control designs, which may be prone to spectrum bias due to inclusion of severe phenotypes, thus leading to optimistic accuracy estimates[122]. Additional cohort studies with study populations that are representative of target populations for implementation have been required. Reassuringly, my results are supported by recently published data from the CORTIS study, which demonstrated that the WHO incipient test TPP were only met for a six-month time horizon when

implemented for general population screening[123]. Taken together, these data suggest that optimal implementation of these biomarkers will require careful selection of target populations who are at highest risk of disease in the short-term, to maximise prior probability and thus positive predictive value for incident disease.

Translation of blood transcriptomic signatures to near-patient diagnostic platforms is underway and one prototype is under evaluation, although further validation is required and appropriate test thresholds must be defined[259]. An additional challenge to implementation is the cost of these assays. This is likely to far exceed the $2 target specified by the WHO TPP for a non-sputum triage test for TB disease[101], but may achieve the WHO target price to identify incipient TB for <$100, using the price of IGRAs as an initial benchmark[11]. Importantly, the fact that a number of different signatures show similar performance may encourage market competition to drive down costs.

6.1.4 Implications for global TB control

My findings further highlight the limitations of current LTBI assays and RNA biomarkers as next generation incipient TB tests for predicting TB development. In countries with low TB incidence, TB prevention through LTBI screening is widely implemented among risk groups such as recent TB contacts, migrants from countries with high TB incidence and immunocompromised groups (Chapter 1.8). My analyses in Chapter 2 suggest that current TST and IGRA thresholds are appropriate and should not be increased programmatically. The PERISKOPE-TB prognostic tool developed in Chapter 4 may allow more precise targeting of preventative treatment in such settings by facilitating individualised risk: benefit assessments. In addition, the utility of RNA biomarkers to stratify preventative treatment in risk groups such as recent contacts could be evaluated in future studies (discussed in Chapter 6.2.4). However, it should be noted that approaches that facilitate intentional targeting of preventative treatment towards people at higher risk of incident TB may perversely undermine TB pre-elimination efforts in low incidence settings, when compared to a utilitarian approach that is willing to accept the risk of individual-level net harm in order to pursue the long-term goals of reducing TB incidence[261]. A utilitarian approach may be more acceptable when individual-level harms are minimised through discovery and

implementation of ultra-short preventative treatment regimens that are safe and effective.

In higher transmission settings, LTBI screening is much less widely implemented; scale-up of preventative treatment therefore represents an important component of the End TB Strategy (Chapter 1.2.3). In these settings, TST remains the most widely available prognostic test, due to the higher cost and requirement of laboratory infrastructure for IGRAs. As discussed in Chapter 6.1.3, blood RNA biomarkers are being commercialised using near-patient platforms, but their cost may be prohibitive for widespread implementation as screening tests in low- and middle-income countries, particularly given that they only reflect short-term risk of disease. This means that future investments in screening using prognostic tests for TB may require careful consideration of potential opportunity costs when compared to alternative investments in emerging vaccines and truncated preventative treatment regimens.

Moreover, even in 2019, a large diagnosis and notification gap of “missing millions” (estimated as 2.9 million cases) remained between annual TB notifications and the true number estimated to have developed TB, while only 38% of multi-drug resistant cases were estimated to be notified and enrolled in treatment[21]. When combined with ongoing gaps in the cascade of TB care, these sobering statistics highlight a critical need to improve detection and treatment of global TB disease. In the absence of a widely available, affordable, accurate and rapid prognostic test for future TB development, strengthening of TB disease detection and treatment programmes may currently represent a more urgent public health priority, when compared to preventative treatment scale-up, in the highest burden settings. As progress is made towards closing the case notification gap and strengthening cascades of care, increasing investments may follow in TB prevention, focusing on highest risk groups such as recent household contacts (particularly young children) and PLHIV. Nevertheless, investments in TB prevention should ideally occur alongside strengthening of TB detection and treatment programmes where possible.

6.2 Future research priorities

My findings give rise to a number of subsequent research questions that are summarised by the hypotheses below.

6.2.1 Hypothesis 1: Binary and quantitative TST and IGRAs have equivalent prognostic ability for incident TB in high-incidence settings

The prognostic ability of TST and IGRAs from head-to-head studies has been consistently observed to be lower in settings with high TB incidence, with pooled incidence rate ratios of 1.35-2.16[79], compared to 5.4-8.8 in the UK PREDICT study[80]. This is likely to be explained by lower TST and IGRA specificity in high *M. tuberculosis* transmission settings, where the majority of the population demonstrate evidence of sensitisation, along with higher risk of future infection among people without sensitisation[62,79]. While I have found that quantitative TST and IGRA results have equivalent prognostic ability in head-to-head analyses in low incidence settings, it therefore remains unclear whether this observation extends to countries with higher TB burdens. This important question will be addressed by an underway systematic review and IPD-MA that seeks to compare the predictive ability of the TST and IGRAs for incident TB globally (PROSPERO CRD42020205667), as an extension of my work in Chapter 4.

6.2.2 Hypothesis 2: The PERISKOPE-TB prognostic model can be tailored to improve prognostic ability of TST and IGRA in moderate- and high-burden settings

A limitation of PERISKOPE-TB is that its generalisability as currently configured is restricted to low transmission settings (annual incidence ≤20/100,000 persons). The rationale for limiting to such settings was, firstly, to examine progression from LTBI to TB disease with a low risk of re-infection with *M. tuberculosis* during follow-up. Secondly, the majority of the population in high transmission settings are likely to have a positive LTBI test result[62]. Since the quantitative LTBI test result is a strong predictor in PERISKOPE-TB, a tailored prediction model may therefore be required in such settings. Future studies could evaluate PERISKOPE-TB for use in high transmission settings, updating the parameters as necessary, in order to assess whether a multivariable approach may improve

the discrimination and clinical utility of TST and IGRAs. A particular challenge when updating the model is likely to be incorporating the quantitative LTBI test result, since the distribution of quantitative responses is likely to be very different in higher TB burden settings, where the majority of population may be sensitised to *M. tuberculosis*. Thus, a large-scale multi-site dataset is likely to be required in order to update and validate the model for use in such settings.

6.2.3 Hypothesis 3: PERISKOPE-TB can facilitate more precise delivery of preventative treatment, compared to TST or IGRAs alone

As shown in Chapter 4, the PERISKOPE-TB demonstrates potential clinical utility for guiding decisions to initiate preventative treatment. The aim of this approach would be to target treatment to people at higher risk of TB disease, while reducing the burden of unnecessary treatment among people at lower risk. This hypothesis could be evaluated in a future RCT. Such a trial could compare systematically offering preventative treatment in an intervention arm to people with two-year incident TB risk (as determined by PERISKOPE-TB) above a set threshold probability, to a standard care arm where treatment is offered to all people with a positive TST or IGRA result. The outcome of this evaluation should consider whether the PERISKOPE-TB approach is non-inferior to using binary TST or IGRA cut-offs with respect to incident TB outcomes, while being superior in reducing the number of people receiving preventative antimicrobials. Alternatively, a composite outcome related to decision theory could be considered, similar to the net benefit metric used in decision curve analysis in Chapter 4. Specifically, a composite 'net harm' outcome could be incorporated, reflecting the cumulative incident TB risk *plus* the number of people receiving preventative treatment, weighted by *w*, where *w* is a weighting factor that considers the relative importance of each adverse outcome[262]. Qualitative work is likely to be required to define an acceptable range of weighting factors.

Future trials may also evaluate the potential health economic impact of programmatic implementation of the model, along with the impact of this approach on the pursuit of TB pre-elimination.

6.2.4 Hypothesis 4: Blood transcriptomic biomarkers can facilitate more precise delivery of preventative treatment, compared to TST or IGRAs, among recent TB contacts

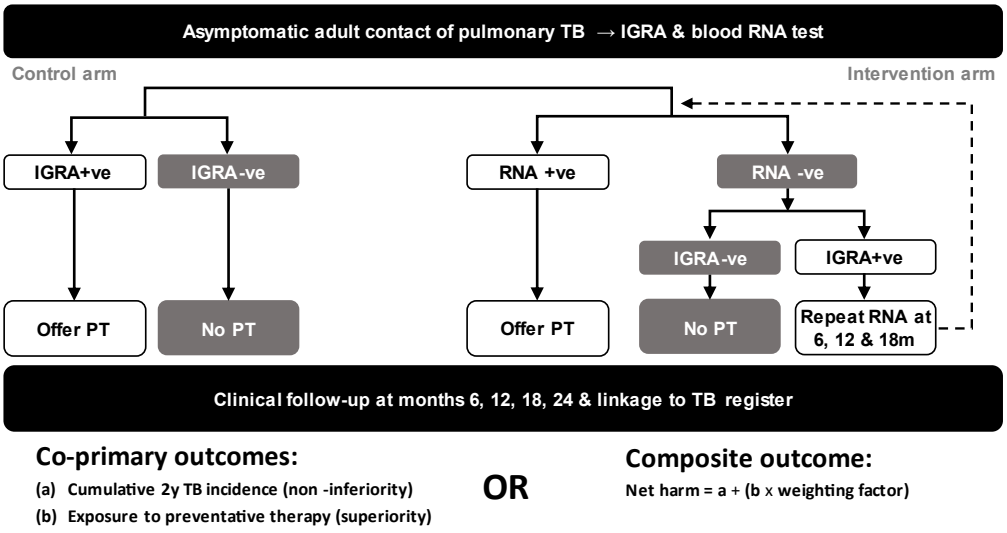
Blood transcriptomic biomarkers also show promise as a method to facilitate more precise administration of preventative treatment, when compared to existing LTBI tests. However, since they only reflect short-term risk of disease, careful selection of target populations is likely to be required in order to achieve impact. Recent TB contacts represent a tractable target population, since the risk of TB decreases with increasing time since infection, with approximately 80% of cumulative five-year risk accrued in the first two years (Chapter 4). Future cohort studies could assess the performance of the best performing transcriptomic signatures, with direct head-to-head comparison to the other most promising candidate biomarkers (e.g., proteomic, metabolomic and antigen candidates), alone and in combination with TST or IGRA to stratify preventative treatment among recent TB contacts. Such studies should include study populations representative of target populations for implementation and could assess clinical utility using decision curve analysis, in addition to standard assessments of diagnostic accuracy and prognostic ability.

An RCT could also be conducted to assess whether implementing blood transcriptomic biomarker-stratified preventative treatment is non-inferior to standard care (using TST or IGRA) with respect to incident TB outcomes, while achieving superiority in reducing the number of people receiving preventative treatment and improving cost-effectiveness. Similarly to the RCT theme proposed in Chapter 6.2.3, a composite 'net harm' endpoint could also be considered. Such a trial design would also have to take into account the need for serial transcriptomic testing, given the time-dependent accuracy of these biomarkers. Since TST and IGRAs are less affected by interval to disease, one potential approach would be a combined testing strategy with both IGRA and RNA biomarkers; serial RNA testing on a six-monthly basis for 1-2 years could then be considered for people who are IGRA positive but RNA signature negative at baseline (Figure 6-1). Such a trial could take place in low TB incidence settings, where there is likely to be a low risk of re-exposure to *M. tuberculosis* during follow-up; though also in higher burden settings where coverage of contact tracing is generally poor[21]. In addition, the feasibility of implementing blood

transcriptomic biomarkers (including acceptability of serial testing) and their impact on preventative treatment acceptance and completion could be evaluated in parallel to assess the likely downstream effects on the cascade of LTBI care.

Figure 6-1: Proposed trial design for RCT comparing blood transcriptomic (RNA) biomarker-stratified preventative treatment to standard care among recent contacts.

PT = preventative treatment; m = month; y = year.



6.2.5 Hypothesis 5: The diagnostic and prognostic ability of blood transcriptomic biomarkers can be improved by integration into multivariable prediction models

My work in Chapter 5 demonstrates that transcriptomic biomarkers meet or approximate the minimum WHO TPP benchmarks over a 3-6 months interval, but do not meet optimal performance targets. These findings were mirrored in a parallel analysis to which my work contributed, comparing the diagnostic accuracy of candidate transcriptomic biomarkers among adults presenting with symptoms compatible with TB disease in South Africa[263]. In this analysis, four candidate signatures had equivalent accuracy and met or approximated the WHO minimum, but not optimum, targets for a triage test. As demonstrated in Chapter 4, the discrimination and clinical utility of TST and IGRA results can be improved by integration into a multivariable prognostic model. Thus, future studies could consider integration of blood transcriptomic biomarkers with other clinical

variables in order to assess whether a multivariable approach improves discrimination and clinical utility for detection of incipient TB and triaging risk of prevalent TB disease.

6.2.6 Hypothesis 6: Incipient TB requires a stratified treatment approach

With increasing recognition of the continuum of *M. tuberculosis* infection, the previously binary distinctions between prevalent *vs* incident, and latent *vs* active TB are becoming increasingly blurred, and appropriate treatment regimens for incipient TB are yet to be defined (as discussed in Chapter 1.7.4). Recent evidence from the CORTIS trial found that 3HP did not reduce incident TB over a 15-month interval among people with a positive blood transcriptomic biomarker[123]. However, the trial was limited by power, with only six incident TB events occurring in the 3HP+ arm, and no cases occurring before nine months' follow-up among adherent participants suggesting possible contributions of non-adherence and re-infection with *M. tuberculosis* for incident cases.

The incipient phase, between latent infection and TB disease, is likely to be heterogeneous, with increasing mycobacterial burden as it progresses (Figure 1-2). Notably, previous studies have shown that shorter treatment regimens are non-inferior to standard six-month treatment for people with non-severe, paucibacillary TB disease (defined by low grade positive or negative sputum smear status and absence of cavitation)[139,140]. Thus, it follows that a stratified approach to incipient TB treatment may be required. Future RCTs could evaluate such an approach (Table 6-1), possibly in combination with the diagnostic trial design proposed in Figure 6-1.

Table 6-1: Possible stratified treatment approach for people with incipient TB or drug-susceptible pulmonary TB disease.

Incipient TB test	Symptoms	Sputum smear	Xpert / Culture	Chest radiograph abnormalities*	Proposed treatment
+	-	-	-	-	Preventative therapy
+	-	-	Either but not both +		2-4 month regimen
+	-	-	+	+	4 month regimen
+	+	-	+	+	4 month regimen

+	+	+ (or CXR cavities)	+	+	6 month regimen

CXR = chest x-ray

**Excludes cavitatory disease.*

Preventative treatment may include any regimen approved in WHO LTBI guidance[81].

Two month regimen could consist of intensive phase treatment only, or regimens under evaluation in treatment shortening trials[143].

Four month regimen could include fluoroquinolones[139,141].

6.3 Wider relevance and applications of methodology used in this thesis

In this thesis, I have used a range of methodological approaches - spanning epidemiological analyses, evaluation of prognostic tests, IPD-MA, multivariable risk prediction, whole blood transcriptomics and systematic head-to-head evaluation of candidate biomarkers. These methods are likely to be transferrable to the study of a range of infectious diseases, as demonstrated in the examples below for COVID-19.

6.3.1 Systematic head-to-head external validation

In Chapter 5, I performed a head-to-head external validation study of 17 systemically-identified candidate transcriptomic signatures for TB for detection of incipient TB. In parallel, my systematic review and head-to-head analysis contributed towards a second study, which compared the diagnostic accuracy of transcriptomic signatures for TB disease among symptomatic adults in South Africa, as described in Chapter 6.2.5[263].

I have subsequently applied similar methods in COVID-19 by systematically evaluating the diagnostic accuracy of candidate transcriptomic signatures for viral infection for detection of early and pre-symptomatic SARS-CoV-2 infection, in a nested case-control study among healthcare workers[264]. I found that four signatures reflecting type 1 interferon signalling (including a single transcript IFI27) demonstrated high accuracy for detecting contemporaneous PCR-positivity, with AUROCs 0.91-0.95. These signatures also showed some, albeit lower, discrimination a week prior to PCR positivity, with AUROCs 0.75-0.80. My findings suggest that host response blood biomarkers warrant further evaluation for potential application alongside viral PCR to facilitate early SARS-CoV-2 case detection, particularly if translated to near-patient platforms.

I have also used the framework of systematic head-to-head evaluation to evaluate the performance of a range of clinical prognostic models to predict clinical outcomes among adults hospitalised with COVID-19. Upon evaluation of the discrimination, calibration and clinical utility of 22 candidate prognostic models early in the pandemic, I found that none offered incremental value over and above simple univariable predictors of age to predict in-hospital mortality and admission peripheral oxygen saturation on air to predict in-hospital deterioration[265]. Thus, none of the models evaluated could be recommended for clinical use.

6.3.2 Prediction model development and validation

Prediction models seek to make inherently subjective medical decision-making, which may vary markedly between clinicians, more objective by providing data-driven risk estimates. In Chapter 4, I developed and validated a prognostic model for incident TB, and sought to align with best practice standards in prediction modelling throughout. This included: adherence to TRIPOD reporting; using multiple imputation to deal with missing data; accounting for non-linear associations with restricted cubic splines; assessing validation through discrimination and calibration parameters; and evaluating clinical utility using decision curve analysis[197,200]. In addition, I used an IPD-MA approach with the IECV framework in order to assess between-study heterogeneity and further evaluate potential generalisability.

I have subsequently applied these methods in response to the COVID-19 pandemic by leading development and validation of the International Severe Acute Respiratory and Emerging Infections Consortium Coronavirus Clinical Characterisation Consortium (ISARIC4C) Deterioration model for adults hospitalised with COVID-19. In this work, I sought to address the weaknesses of previously derived clinical prognostic models for COVID-19, which were often limited to small, inadequately powered samples from single centres and were deemed to be at high risk of bias with overly optimistic performance metrics in original reports during comprehensive quality assessment[266]. I used ISARIC4C data from approximately 75,000 adults from >250 hospitals across the UK to develop and validate a prognostic model for in-hospital clinical deterioration (defined as any requirement of ventilatory support or critical care, or death). I

used the IECV approach to explore heterogeneity in model performance across nine NHS regions, and found strong evidence of consistent performance and greater clinical utility to inform medical decision-making than other prognostic scores across all regions. The prognostic tool is now implemented for NHS use at https://isaric4c.net/risk/.

6.4 Conclusion

In this thesis, I have developed and evaluated a range of approaches to facilitate more precise targeting of TB preventative treatment. While TST and IGRAs have limitations, their prognostic value can be optimised for use in settings with low TB incidence by incorporating quantitative results into a multivariable prognostic model along with routinely measured clinical covariates. Multiple blood transcriptomic signatures appear promising as biomarkers for incipient TB but only reflect short-term disease. Carefully considered implementation among selected high-risk populations, such as recent TB contacts, is therefore likely to be required to achieve impact. I have highlighted future research priorities in the context of specific hypotheses arising from this work, including a need to evaluate the impact of implementing the PERISKOPE-TB prognostic model and blood transcriptomic biomarkers on individual-level clinical outcomes, along with population-level trends in TB incidence, through interventional trials. Future approaches could also consider integration of next generation biomarkers and clinical variables into multivariable models to assess whether their diagnostic and prognostic value may be further improved. In parallel, I have shown that the methods applied in this thesis are transferable to the study of other infectious diseases, as demonstrated through specific applications in COVID-19.

6.5 Outputs relating to this chapter

I applied the methodology used in this thesis in the following publications to study COVID-19:

- **Gupta RK**, Harrison EM, …, Noursadeghi M on behalf of the ISARIC4C Investigators (2021). Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. **The Lancet Respiratory Medicine.** https://doi.org.10.1016/S2213-2600(20)30559-2

- **Gupta RK**, Marks M, Samuels T, Luintel A, Rampling T, Chowdhury H, Quartagno M, Nair A, Lipman M, Abubakar I, van Smeden M, Wong WK, Williams B, Noursadeghi M (2020). Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: An observational cohort study. **European Respiratory Journal.** https://doi.org.10.1183/13993003.03498-2020
- **Gupta RK**, Rosenheim J, Bell LC, …, Noursadeghi M (2021). Blood transcriptional biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection. **medRxiv** 2021.01.18.21250044; https://doi.org/10.1101/2021.01.18.21250044
- Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, Dunning J, Fairfield CJ, Gamble C, Green CA, **Gupta RK**, …, Openshaw PJ, Baillie JK, Semple MG, Docherty AB, Harrison EM (2020). Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. **BMJ**. https://doi.org.10.1136/bmj.m3339

7 References

1. Pai, M. *et al.* Tuberculosis. *Nature Reviews Disease Primers* vol. 2 1–23 (2016).
2. WHO policy on collaborative TB/HIV activities. *WHO* https://www.who.int/tb/publications/2012/tb_hiv_policy_9789241503006/en/ (2015).
3. Keane, J. *et al.* Tuberculosis Associated with Infliximab, a Tumor Necrosis Factor α–Neutralizing Agent. *N. Engl. J. Med.* **345**, 1098–1104 (2001).
4. Bustamante, J., Boisson-Dupuis, S., Abel, L. & Casanova, J.-L. Mendelian susceptibility to mycobacterial disease: Genetic, immunological, and clinical features of inborn errors of IFN-γ immunity. *Semin. Immunol.* **26**, 454–470 (2014).
5. Feldman, W. H. & Baggenstoss, A. H. The residual infectivity of the primary complex of tuberculosis. *Am. J. Pathol.* **14**, 473-490.3 (1938).
6. Opie, E. Tubercle bacilli in latent tuberculous lesions and in the lung tissue without tuberculous lesions. *Arch. Pathol. Lab. Med.* **4**, 1–21 (1927).
7. Getahun, H., Matteelli, A., Chaisson, R. E. & Raviglione, M. Latent Mycobacterium tuberculosis Infection. *N. Engl. J. Med.* **372**, 2127–2135 (2015).
8. Frascella, B. *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. The discovery of symptomless pulmonary tuberculosis: The policy of the national association for the prevention of tuberculosis. *Tubercle* vol. 23 56–57 (1942).
10. Esmail, H., Barry, C. E., Young, D. B. & Wilkinson, R. J. The ongoing challenge of latent tuberculosis. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130437–20130437 (2014).
11. WHO. Development of a Target Product Profile (TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease 2017 WHO collaborating centre for the evaluation of new diagnostic technologies. https://apps.who.int/iris/bitstream/handle/10665/259176/WHO-HTM-TB-2017.18-eng.pdf?sequence=1 (2017).
12. Drain, P. K. *et al.* Incipient and subclinical tuberculosis: A clinical review of early stages and progression of infection. *Clin. Microbiol. Rev.* **31**, e00021-18 (2018).
13. Ritter, J. So-Called “Incipient” Tuberculosis. *Journal of the American Medical Association* vol. LXVI 592 (1916).
14. Ferebee, S. H. Controlled chemoprophylaxis trials in tuberculosis. A general review. *Bibl. Tuberc.* **26**, 28–106 (1970).

15. Efficacy of various durations of isoniazid preventive therapy for tuberculosis: five years of follow-up in the IUAT trial. International Union Against Tuberculosis Committee on Prophylaxis. *Bull. World Health Organ.* **60**, 555–64 (1982).

16. Behr, M. A., Edelstein, P. H. & Ramakrishnan, L. Revisiting the timetable of tuberculosis. *BMJ* **362**, k2738 (2018).

17. Aldridge, R. W. *et al.* Tuberculosis in migrants moving from high-incidence to low-incidence countries: a population-based cohort study of 519 955 migrants screened before entry to England, Wales, and Northern Ireland. *Lancet* **388**, 2510–2518 (2016).

18. Dale, K. D., Trauer, J. M., Dodd, P. J., Houben, R. M. G. J. & Denholm, J. T. Estimating Long-term Tuberculosis Reactivation Rates in Australian Migrants. *Clin. Infect. Dis.* **70**, 2111–2118 (2020).

19. Tuberculosis in England: annual report 2020. https://www.gov.uk/government/publications/tuberculosis-in-england-annual-report.

20. Hermans, S., Horsburgh Jr., C. R. & Wood, R. A Century of Tuberculosis Epidemiology in the Northern and Southern Hemisphere: The Differential Impact of Control Interventions. *PLoS One* **10**, e0135179 (2015).

21. WHO Global tuberculosis report 2020. https://www.who.int/publications/i/item/9789240013131.

22. WHO. The End TB Strategy. http://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1 (2015).

23. Comstock, G. W., Ferebee, S. H. & Hammes, L. M. A Controlled Trial of Community-wide Isoniazid Prophylaxis in Alaska. *Am. Rev. Respir. Dis.* **95**, (1967).

24. UN General Assembly High-Level Meeting on Ending TB 2018. https://www.who.int/news-room/events/un-general-assembly-high-level-meeting-on-ending-tb.

25. Aronson, N. E. *et al.* Long-term Efficacy of BCG Vaccine in American Indians and Alaska Natives: A 60-Year Follow-up Study. *Journal of the American Medical Association* vol. 291 2086–2091 (2004).

26. Trunz, B. B., Fine, P. & Dye, C. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet* **367**, 1173–1180 (2006).

27. Zwerling, A. *et al.* The BCG World Atlas: A Database of Global BCG Vaccination Policies and Practices. *PLoS Med.* **8**, e1001012 (2011).

28. Mangtani, P. *et al.* Protection by BCG Vaccine Against Tuberculosis: A Systematic Review of Randomized Controlled Trials. *Clin. Infect. Dis.* **58**, 470–480 (2014).

29. Roy, A. *et al.* Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *Bmj* **349**, g4643–g4643 (2014).

30. Nemes, E. *et al.* Prevention of *M. tuberculosis* Infection with H4:IC31 Vaccine or BCG Revaccination. *N. Engl. J. Med.* **379**, 138–149 (2018).

31. Tait, D. R. *et al.* Final Analysis of a Trial of M72/AS01 $_E$ Vaccine to Prevent Tuberculosis. *N. Engl. J. Med.* **381**, 2429–2439 (2019).

32. WHO | WHO Guidelines on tuberculosis infection prevention and control, 2019 update. *WHO* http://www.who.int/tb/publications/2019/guidelines-tuberculosis-infection-prevention-2019/en/ (2019).

33. WHO | TB comorbidities and risk factors. https://www.who.int/tb/areas-of-work/treatment/risk-factors/en/.

34. Luo, W., Lin, Y., Li, Z., Wang, W. & Shi, Y. Comparison of sputum induction and bronchoscopy in diagnosis of sputum smear-negative pulmonary tuberculosis: A systemic review and meta-analysis. *BMC Pulmonary Medicine* vol. 20 146 (2020).

35. Gupta, R. K. *et al.* Impact of human immunodeficiency virus and CD4 count on tuberculosis diagnosis: analysis of city-wide data from Cape Town, South Africa. *Int. J. Tuberc. Lung Dis.* **17**, 1014–22 (2013).

36. Chihota, V. N. *et al.* Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting. *Int. J. Tuberc. Lung Dis.* **14**, 1024–31 (2010).

37. WHO. Automated Real-time Nucleic Acid Amplification Technology for Rapid and Simultaneous Detection of Tuberculosis and Rifampicin Resistance : Xpert MTB / RIF System Policy Statement. *Geneva, Switzerland* http://whqlibdoc.who.int/publications/2011/9789241501545_eng.pdf?ua=1 (2011).

38. WHO Meeting Report of a Technical Expert Consultation: Non-inferiority analysis of Xpert MTB/RIF Ultra compared to Xpert MTB/RIF. http://apps.who.int/iris/bitstream/10665/254792/1/WHO-HTM-TB-2017.04-eng.pdf?ua=1.

39. Chakravorty, S. *et al.* The New Xpert MTB/RIF Ultra: Improving Detection of *Mycobacterium tuberculosis* and Resistance to Rifampin in an Assay Suitable for Point-of-Care Testing. *MBio* **8**, e00812-17 (2017).

40. Mishra, H. *et al.* Xpert MTB/RIF Ultra and Xpert MTB/RIF for diagnosis of tuberculosis in an HIV-endemic setting with a high burden of previous

tuberculosis: a two-cohort diagnostic accuracy study. *Lancet Respir. Med.* **8**, 368–382 (2020).

41. Kohli, M. *et al.* Xpert MTB/RIF Ultra and Xpert MTB/RIF assays for extrapulmonary tuberculosis and rifampicin resistance in adults. *The Cochrane database of systematic reviews* vol. 1 CD012768 (2021).

42. Kay, A. W. *et al.* Xpert MTB/RIF and Xpert MTB/RIF Ultra assays for active tuberculosis and rifampicin resistance in children. *Cochrane Database of Systematic Reviews* vol. 2020 (2020).

43. Horne, D. & Steingart, K. Cochrane Library Cochrane Database of Systematic Reviews Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults (Review). (2019) doi:10.1002/14651858.CD009593.pub4.

44. WHO | Rapid Communication: Molecular assays as initial tests for the diagnosis of tuberculosis and rifampicin resistance. *WHO* http://www.who.int/tb/publications/2020/rapid-communications-molecular-assays/en/ (2020).

45. Small, P. M. & Pai, M. Tuberculosis diagnosis--time for a game change. *N Engl J Med* **363**, 1070–1 (2010).

46. Di Tanna, G. L. *et al.* Effect of Xpert MTB/RIF on clinical outcomes in routine care settings: individual patient data meta-analysis. *Lancet Glob. Heal.* **7**, e191–e199 (2019).

47. WHO | The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV. http://www.who.int/tb/publications/use-of-lf-lam-tb-hiv/en/ (2015).

48. WHO | Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV. *WHO* http://www.who.int/tb/publications/2019/LAMPolicyUpdate2019/en/ (2019).

49. Gupta-Wright, A. *et al.* Rapid urine-based screening for tuberculosis in HIV-positive patients admitted to hospital in Africa (STAMP): a pragmatic, multicentre, parallel-group, double-blind, randomised controlled trial. *Lancet* **392**, 292–301 (2018).

50. Peter, J. G. *et al.* Effect on mortality of point-of-care, urine-based lipoarabinomannan testing to guide tuberculosis treatment initiation in HIV-positive hospital inpatients: A pragmatic, parallel-group, multicountry, open-label, randomised controlled trial. *Lancet* **387**, 1187–1197 (2016).

51. Broger, T. *et al.* Diagnostic accuracy of a novel tuberculosis point-of-care urine

lipoarabinomannan assay for people living with HIV: A meta-analysis of individual in- and outpatient data. *PLOS Med.* **17**, e1003113 (2020).

52. Broger, T. *et al.* Diagnostic accuracy of 3 urine lipoarabinomannan tuberculosis assays in HIV-negative outpatients. *J. Clin. Invest.* **130**, 5756–5764 (2020).
53. Xie, Y. L. *et al.* Evaluation of a Rapid Molecular Drug-Susceptibility Test for Tuberculosis. *N. Engl. J. Med.* **377**, 1043–1054 (2017).
54. Walker, T. M. *et al.* Tuberculosis is changing. *Lancet. Infect. Dis.* **17**, 359–361 (2017).
55. CRyPTIC – Comprehensive Resistance Prediction for Tuberculosis: an International Consortium. http://www.crypticproject.org/.
56. WHO | WHO treatment guidelines for multidrug- and rifampicin-resistant tuberculosis, 2018 update. *WHO* http://www.who.int/tb/areas-of-work/drug-resistant-tb/guideline-update2018/en/ (2019).
57. Silverman, C. An appraisal of the contribution of mass radiography in the discovery of pulmonary tuberculosis. *Am. Rev. Tuberc.* **60**, 466–482 (1949).
58. WHO | Chest radiography in tuberculosis detection. *WHO* http://www.who.int/tb/publications/chest-radiography/en/ (2017).
59. Pande, T., Pai, M., Khan, F. A. & Denkinger, C. M. Use of chest radiography in the 22 highest tuberculosis burden countries. *European Respiratory Journal* vol. 46 1816–1819 (2015).
60. Qin, Z. Z. *et al.* Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Sci. Rep.* **9**, 1–10 (2019).
61. Chen, R. Y. *et al.* PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis. *Sci. Transl. Med.* **6**, 265ra166 (2014).
62. WHO | Latent TB Infection : Updated and consolidated guidelines for programmatic management. *WHO* http://www.who.int/tb/publications/2018/latent-tuberculosis-infection/en/ (2018).
63. Menzies, D. Interpretation of repeated tuberculin tests: Boosting, conversion, and reversion. *American Journal of Respiratory and Critical Care Medicine* vol. 159 15–21 (1999).
64. QuantiFERON®-TB Gold Plus (QFT®-Plus) ELISA Package Insert. https://www.quantiferon.com/wp-content/uploads/2020/01/L1083163-R06-QF-TB-Gold-Plus-ELISA-IFU-CE.pdf.
65. T-SPOT.TB package insert. https://www.tspot.com/uk/wp-content/uploads/sites/3/2020/12/PI-TB-IVD-UK-v3.pdf.

66. Pai, M. *et al.* Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. *Clin. Microbiol. Rev.* **27**, 3–20 (2014).

67. Dorman, S. E. *et al.* Interferon-gamma release assays and tuberculin skin testing for diagnosis of latent tuberculosis infection in healthcare workers in the United States. *Am. J. Respir. Crit. Care Med.* **189**, 77–87 (2014).

68. Nemes, E. *et al.* Optimization and interpretation of serial QuantiFERON testing to measure acquisition of mycobacterium tuberculosis infection. *Am. J. Respir. Crit. Care Med.* **196**, 638–648 (2017).

69. Pai, M., Zwerling, A. & Menzies, D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Ann. Intern. Med.* **149**, 177–84 (2008).

70. Sester, M. *et al.* Interferon-gamma release assays for the diagnosis of active tuberculosis: a systematic review and meta-analysis. *Eur. Respir. J.* **37**, 100–111 (2011).

71. Metcalfe, J. Z. *et al.* Interferon-gamma release assays for active pulmonary tuberculosis diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. *J. Infect. Dis.* **204 Suppl**, S1120-9 (2011).

72. Whitworth, H. S. *et al.* Clinical utility of existing and second-generation interferon-γ release assays for diagnostic evaluation of tuberculosis: an observational cohort study. *Lancet Infect. Dis.* **19**, 193–202 (2019).

73. Du, F. *et al.* Prospective Comparison of QFT-GIT and T-SPOT.TB Assays for Diagnosis of Active Tuberculosis. *Sci. Rep.* **8**, 5882 (2018).

74. National Institute for Health and Care Excellent Guidance and Guidelines. Tuberculosis. https://www.nice.org.uk/guidance/ng33 (2016).

75. Lu, L. L. *et al.* IFN-γ-independent immune markers of Mycobacterium tuberculosis exposure. *Nat. Med.* **25**, 977–987 (2019).

76. Campbell, J. R., Winters, N. & Menzies, D. Absolute risk of tuberculosis among untreated populations with a positive tuberculin skin test or interferon-gamma release assay result: systematic review and meta-analysis. *BMJ* **368**, (2020).

77. Zhou, G. *et al.* Interferon-γ release assays or tuberculin skin test for detection and management of latent tuberculosis infection: a systematic review and meta-analysis. *Lancet. Infect. Dis.* **0**, (2020).

78. Rangaka, M. X. *et al.* Predictive value of interferon-gamma release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet. Infect. Dis.* **12**, 45–55 (2012).

79. Hamada, Y., Gupta, R. K., Matteelli, A., Abubakar, I. & Rangaka, M. X. Predictive

performance of interferon-γ release assays and tuberculin skin tests. *The Lancet Infectious Diseases* vol. 20 1371–1372 (2020).

80. Abubakar, I. *et al.* Prognostic value of interferon-gamma release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *Lancet. Infect. Dis.* **18**, 1077–1087 (2018).

81. WHO consolidated guidelines on tuberculosis 2020: module 1: prevention: tuberculosis preventive treatment. https://www.who.int/publications/i/item/who-consolidated-guidelines-on-tuberculosis-module-1-prevention-tuberculosis-preventive-treatment.

82. Programmatic management of latent tuberculosis infection in the European Union 2018. https://www.ecdc.europa.eu/en/publications-data/programmatic-management-latent-tuberculosis-infection-european-union.

83. Dunlap, N. E. *et al.* Diagnostic standards and classification of tuberculosis in adults and children. *American Journal of Respiratory and Critical Care Medicine* vol. 161 1376–1395 (2000).

84. Andrews, J. R. *et al.* Serial QuantiFERON testing and tuberculosis disease risk among young children: an observational cohort study. *Lancet Respir. Med.* **5**, 282–290 (2017).

85. Winje, B. A. *et al.* Stratification by interferon-γ release assay level predicts risk of incident TB. *Thorax* **73**, 652–661 (2018).

86. Diel, R., Bottger, E. C. & Diel, R. Interferon-gamma release assays and the risk of developing active tuberculosis: Reply. *Am. J. Respir. Crit. Care Med.* **185**, 787 (2012).

87. Geis, S. *et al.* How can we achieve better prevention of progression to tuberculosis among contacts? *Eur. Respir. J.* **42**, 1743–1746 (2013).

88. Zellweger, J.-P. *et al.* Risk Assessment of Tuberculosis in Contacts by IFN-γ Release Assays. A Tuberculosis Network European Trials Group Study. *Am. J. Respir. Crit. Care Med.* **191**, 1176–1184 (2015).

89. Ruhwald, M. *et al.* Safety and efficacy of the C-Tb skin test to diagnose Mycobacterium tuberculosis infection, compared with an interferon γ release assay and the tuberculin skin test: a phase 3, double-blind, randomised, controlled trial. *Lancet Respir. Med.* **5**, 259–268 (2017).

90. Day, C. L. *et al.* Functional capacity of Mycobacterium tuberculosis-specific T cell responses in humans is associated with mycobacterial load. *J. Immunol.* **187**, 2222–32 (2011).

91. Nikolova, M. *et al.* Antigen-specific CD4- and CD8-positive signatures in different phases of Mycobacterium tuberculosis infection. *Diagn. Microbiol. Infect. Dis.* **75**, 277–281 (2013).

92. Barcellini, L. *et al.* First independent evaluation of QuantiFERON-TB Plus performance. *Eur. Respir. J.* **47**, 1587–1590 (2016).

93. Barcellini, L. *et al.* First evaluation of QuantiFERON-TB Gold Plus performance in contact screening. *Eur. Respir. J.* **48**, 1411–1419 (2016).

94. Clifford, V., He, Y., Zufferey, C., Connell, T. & Curtis, N. Interferon gamma release assays for monitoring the response to treatment for tuberculosis: A systematic review. *Tuberculosis* **95**, 639–650 (2015).

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

96. Atuk, N. O. & Hunt, E. H. Serial tuberculin testing and isoniazid therapy in general hospital employees. *JAMA* **218**, 1795–8 (1971).

97. Houk, V. N., Kent, D. C., Sorensen, K. & Baker, J. H. The eradication of tuberculosis infection by isoniazid chemoprophylaxis. *Arch. Environ. Health* **16**, 46–50 (1968).

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

99. Machingaidze, S. *et al.* Predictive value of recent quantiFERON conversion for tuberculosis disease in adolescents. *Am. J. Respir. Crit. Care Med.* **186**, 1051–1056 (2012).

100. Trauer, J. M. *et al.* Risk of Active Tuberculosis in the Five Years Following Infection . . . 15%?. *Chest* **149**, 516–525 (2016).

101. WHO. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. https://www.who.int/tb/publications/tpp_report/en/ (2014).

102. Bm2Dx: The Biomarker Database. https://www.bm2dx.org/.

103. MacLean, E. *et al.* A systematic review of biomarkers to detect active tuberculosis. *Nature Microbiology* vol. 4 748–758 (2019).

104. Berry, M. P. R. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973–977 (2010).

105. Sweeney, T. E., Braviak, L., Tato, C. M. & Khatri, P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir. Med.* **4**, 213–224 (2016).

106. Gliddon, H. D. *et al.* Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification. *bioRxiv* 583674 (2019) doi:10.1101/583674.

107. Kaforou, M. *et al.* Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med.* **10**, e1001538 (2013).

108. Anderson, S. T. *et al.* Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N. Engl. J. Med.* **370**, 1712–1723 (2014).

109. Roe, J. K. *et al.* Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI insight* **1**, e87238 (2016).

110. Gjoen, J. E. *et al.* Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children. *Sci. Rep.* **7**, 5839 (2017).

111. Walter, N. D. *et al.* Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: a Case-Control Study with Systematic Cross-Classifier Evaluation. *J. Clin. Microbiol.* **54**, 274–282 (2016).

112. de Araujo, L. S. *et al.* Transcriptomic Biomarkers for Tuberculosis: Evaluation of DOCK9. EPHA4, and NPC2 mRNA Expression in Peripheral Blood. *Front. Microbiol.* **7**, 1586 (2016).

113. Warsinske, H. C. *et al.* Assessment of Validity of a Blood-Based 3-Gene Signature Score for Progression and Diagnosis of Tuberculosis, Disease Severity, and Treatment Response. *JAMA Netw. Open* **1**, e183779 (2018).

114. Thompson, E. G. *et al.* Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis (Edinb).* **107**, 48–58 (2017).

115. Penn-Nicholson, A. *et al.* RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. *Sci. Rep.* **10**, 1–21 (2020).

116. Bloom, C. I. *et al.* Detectable Changes in The Blood Transcriptome Are Present after Two Weeks of Antituberculosis Therapy. *PLoS One* **7**, e46191 (2012).

117. Mahomed, H. *et al.* TB Incidence in an Adolescent Cohort in South Africa. *PLoS One* **8**, e59652 (2013).

118. Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **387**, 2312–2322 (2016).

119. Suliman, S. *et al.* Four-gene pan-African blood signature predicts progression to tuberculosis. *Am. J. Respir. Crit. Care Med.* **197**, 1198–1208 (2018).

120. Roe, J. *et al.* Blood transcriptomic stratification of short-term risk in contacts of tuberculosis. *Clin. Infect. Dis.* (2019) doi:10.1093/cid/ciz252.

121. Singhania, A. *et al.* A modular transcriptional signature identifies phenotypic

heterogeneity of human tuberculosis infection. *Nat. Commun.* **9**, 2308 (2018).

122. Usher-Smith, J. A., Sharp, S. J. & Griffin, S. J. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ* **353**, (2016).

123. Scriba, T. J. *et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *Lancet Infect. Dis.* **0**, (2021).

124. Warsinske, H., Vashisht, R. & Khatri, P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLOS Med.* **16**, e1002786 (2019).

125. Scriba, T. Comment on PLOS Medicine: Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. https://journals.plos.org/plosmedicine/article/comment?id=10.1371/annotation/0e5eb688-93f9-4e18-b0ec-8692580ea90e.

126. Penn-Nicholson, A. *et al.* Discovery and validation of a prognostic proteomic signature for tuberculosis progression: A prospective cohort study. *PLOS Med.* **16**, e1002781 (2019).

127. Weiner, J. *et al.* Metabolite changes in blood predict the onset of tuberculosis. *Nat. Commun.* **9**, 5208 (2018).

128. Verma, R. *et al.* A novel, high-sensitivity, bacteriophage-based assay identifies low-level mycobacterium tuberculosis bacteremia in immunocompetent patients with active and incipient tuberculosis. *Clin. Infect. Dis.* **70**, 933–936 (2020).

129. Martineau, A. *et al.* S89 Detection of M. Tuberculosis DNA in CD34-positive peripheral blood mononuclear cells of asymptomatic TB contacts. in *TB: still playing the long game* A54.2-A55 (BMJ Publishing Group Ltd and British Thoracic Society, 2021). doi:10.1136/thorax-2020-BTSabstracts.94.

130. Vaara, J. & Kokkola, K. Inactive pulmonary lesions: a potent risk factor of tuberculosis. *Ann. Clin. Res.* **7**, 331–3 (1975).

131. Esmail, H. *et al.* Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18 F]fluoro-D-glucose positron emission and computed tomography. *Nat. Med.* **22**, 1090–1093 (2016).

132. Molton, J. S. *et al.* Sub-clinical abnormalities detected by PET/MRI in household tuberculosis contacts. *BMC Infect. Dis.* **19**, 83 (2019).

133. Iseman, M. D. Tuberculosis therapy: Past, present and future. in *European Respiratory Journal, Supplement* vol. 20 87S-94s (European Respiratory Society, 2002).

134. Fox, W. Clinical trial of six-month and four-month regimens of chemotherapy in the treatment of pulmonary tuberculosis. Singapore Tuberculosis Service/British

Medical Research Council. *Am. Rev. Respir. Dis.* **119**, 579–585 (1979).

135. WHO | Guidelines for treatment of drug-susceptible tuberculosis and patient care (2017 update). *WHO* http://www.who.int/tb/publications/2017/dstb_guidance_2017/en/ (2018).

136. Jindani, A. *et al.* High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis. *N. Engl. J. Med.* **371**, 1599–1608 (2014).

137. Gillespie, S. H. *et al.* Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N. Engl. J. Med.* **371**, 1577–1587 (2014).

138. Merle, C. S. *et al.* A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N. Engl. J. Med.* **371**, 1588–1598 (2014).

139. Imperial, M. Z. *et al.* A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nat. Med.* **24**, 1708–1715 (2018).

140. Girling, D. J. A controlled trial of 3-month, 4-month, and 6-month regimens of chemotherapy for sputum-smear-negative pulmonary tuberculosis. Results at 5 years. *Am. Rev. Respir. Dis.* **139**, 871–876 (1989).

141. TBTC Study 31: Rifapentine-containing Tuberculosis Treatment Shortening Regimens - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT02410772.

142. Media Statement: Landmark TB Trial Identifies Shorter-Course Treatment Regimen | CDC. https://www.cdc.gov/nchhstp/newsroom/2020/landmark-tb-trial-media-statement.html.

143. Two-month Regimens Using Novel Combinations to Augment Treatment Effectiveness for Drug-sensitive Tuberculosis - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03474198.

144. Conradie, F. *et al.* Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *N. Engl. J. Med.* **382**, 893–902 (2020).

145. Ferebee, S. H. & Mount, F. W. Tuberculosis morbidity in a controlled trial of the prophylactic use of isoniazid among household contacts. *Am. Rev. Respir. Dis.* **85**, 490–510 (1962).

146. Moulding, T. & Comstock, G. W. How much isoniazid is needed for prevention of tuberculosis among immunocompetent adults? [1]. *International Journal of Tuberculosis and Lung Disease* vol. 4 485–487 (2000).

147. Salazar-Austin, N., Dowdy, D. W., Chaisson, R. E. & Golub, J. E. Seventy Years of Tuberculosis Prevention: Efficacy, Effectiveness, Toxicity, Durability, and Duration. *Am. J. Epidemiol.* **188**, 2078–2085 (2019).

148. Menzies, D. *et al.* Adverse Events with 4 Months of Rifampin Therapy or 9 Months

of Isoniazid Therapy for Latent Tuberculosis Infection. *Ann. Intern. Med.* **149**, 689 (2008).

149. Sterling, T. R. *et al.* Three Months of Rifapentine and Isoniazid for Latent Tuberculosis Infection. *N. Engl. J. Med.* **365**, 2155–2166 (2011).

150. Swindells, S. *et al.* One Month of Rifapentine plus Isoniazid to Prevent HIV-Related Tuberculosis. *N. Engl. J. Med.* **380**, 1001–1011 (2019).

151. Menzies, D. *et al.* Four Months of Rifampin or Nine Months of Isoniazid for Latent Tuberculosis in Adults. *N. Engl. J. Med.* **379**, 440–453 (2018).

152. Zenner, D. *et al.* Treatment of Latent Tuberculosis Infection. *Ann. Intern. Med.* **167**, 248 (2017).

153. Sterling, T. R. *et al.* Three months of weekly rifapentine and isoniazid for treatment of Mycobacterium tuberculosis infection in HIV-coinfected persons. *AIDS* **30**, 1607–1615 (2016).

154. Ross, J. M. *et al.* Isoniazid preventive therapy plus antiretroviral therapy for the prevention of tuberculosis: a systematic review and meta-analysis of individual participant data. *Lancet HIV* **8**, e8–e15 (2021).

155. Samandari, T. *et al.* Tuberculosis incidence after 36 months' isoniazid prophylaxis in HIV-infected adults in Botswana: a posttrial observational analysis. *AIDS* **29**, 351–359 (2015).

156. Churchyard, G. J. *et al.* A Trial of Mass Isoniazid Preventive Therapy for Tuberculosis Control. *N. Engl. J. Med.* **370**, 301–310 (2014).

157. Den Boon, S., Matteelli, A., Ford, N. & Getahun, H. Continuous isoniazid for the treatment of latent tuberculosis infection in people living with HIV. *AIDS* **30**, 797–801 (2016).

158. Seddon, J. A. *et al.* Levofloxacin versus placebo for the prevention of tuberculosis disease in child contacts of multidrug-resistant tuberculosis: Study protocol for a phase III cluster randomised controlled trial (TB-CHAMP). *Trials* **19**, 693 (2018).

159. Fox, G. J. *et al.* Levofloxacin versus placebo for the treatment of latent tuberculosis among contacts of patients with multidrug-resistant tuberculosis (the VQUIN MDR trial): A protocol for a randomised controlled trial. *BMJ Open* **10**, (2020).

160. Protecting Households On Exposure to Newly Diagnosed Index Multidrug-Resistant Tuberculosis Patients - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03568383.

161. Kopanoff, D. E., Snider, D. E. & Caras, G. J. Isoniazid-related hepatitis. A U.S. Public Health Service cooperative surveillance study. *Am. Rev. Respir. Dis.* **117**,

991–1001 (1978).

162. Campbell, J. R. *et al.* Adverse events in adults with latent tuberculosis infection receiving daily rifampicin or isoniazid: post-hoc safety analysis of two randomised controlled trials. *Lancet Infect. Dis.* **20**, 318–329 (2020).

163. Public Health England. Collaborative Tuberculosis Strategy for England. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/403231/Collaborative_TB_Strategy_for_England_2015_2020_.pdf (2015).

164. Chapter 6: Canadian Tuberculosis Standards 7th Edition: 2014. https://www.canada.ca/en/public-health/services/infectious-diseases/canadian-tuberculosis-standards-7th-edition/edition-18.html#s3.

165. Martinez, L. *et al.* The risk of tuberculosis in children after close exposure: a systematic review and individual-participant meta-analysis. *Lancet* **395**, 973–984 (2020).

166. Menzies, D., Gardiner, G., Farhat, M., Greenaway, C. & Pai, M. Thinking in three dimensions: a web-based algorithm to aid the interpretation of tuberculin skin test results. *Int. J. Tuberc. Lung Dis.* **12**, 498–505 (2008).

167. TSTin3D: About. https://www.tstin3d.com/en/about.html.

168. Puyat, J. H. *et al.* How Well Does TSTin3D Predict Risk of Active Tuberculosis in the Canadian Immigrant Population? An External Validation Study. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa780.

169. Alsdurf, H., Hill, P. C., Matteelli, A., Getahun, H. & Menzies, D. The cascade of care in diagnosis and treatment of latent tuberculosis infection: a systematic review and meta-analysis. *Lancet Infect. Dis.* **16**, 1269–1278 (2016).

170. Gao, J. *et al.* Knowledge and Perceptions of Latent Tuberculosis Infection among Chinese Immigrants in a Canadian Urban Centre. *Int. J. Family Med.* **2015**, 546042 (2015).

171. Steyerberg, E. W. *Clinical Prediction Models*. (Springer International Publishing, 2019). doi:10.1007/978-3-030-16399-0.

172. National Institute for Clinical Excellence. Tuberculosis Clinical diagnosis and management of tuberculosis, and measures for its prevention and control. http://www.nice.org.uk/guidance/cg117/resources/guidance-tuberculosis-pdf (2011).

173. Leung, C. C. *et al.* A strong tuberculin reaction in primary school children predicts tuberculosis in adolescence. *Pediatr. Infect. Dis. J.* **31**, 150–153 (2012).

174. Moran-Mendoza, O., FitzGerald, J. M., Marion, S., Elwood, K. & Patrick, D. Tuberculin skin test size and risk of TB among non-treated contacts of TB cases.

Can. Respir. J. **17**, 5B (2010).

175. Frank E Harrell. Biostatistical Modeling. http://biostat.mc.vanderbilt.edu/wiki/pub/Main/BioMod/notes.pdf (2004).

176. Gupta, R. K. *et al.* Quantitative Interferon Gamma Release Assay and Tuberculin Skin Test Results to Predict Incident Tuberculosis: A Prospective Cohort Study. *Am. J. Respir. Crit. Care Med.* rccm.201905-0969OC (2019) doi:10.1164/rccm.201905-0969OC.

177. Steyerberg, E. W. & Vergouwe, Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur. Heart J.* **35**, 1925–31 (2014).

178. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–45 (1988).

179. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–5 (1950).

180. BTS Guidelines Control and prevention of tuberculosis in the United Kingdom: Code of Practice 2000. doi:10.1136/thorax.55.11.887.

181. Farhat, M., Greenaway, C., Pai, M. & Menzies, D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria?. *Int. J. Tuberc. Lung Dis.* **10**, 1192–1204 (2006).

182. Katelaris, A. L. *et al.* Effectiveness of BCG Vaccination Against Mycobacterium tuberculosis Infection in Adults: A Cross-sectional Analysis of a UK-Based Cohort. *J. Infect. Dis.* **221**, 146–155 (2020).

183. Al Zahrani, K., Al Jahdali, H. & Menzies, D. Does size matter? Utility of size of tuberculin reactions for the diagnosis of mycobacterial disease. *Am. J. Respir. Crit. Care Med.* **162**, 1419–1422 (2000).

184. Aldridge, R. W., Shaji, K., Hayward, A. C. & Abubakar, I. Accuracy of Probabilistic Linkage Using the Enhanced Matching System for Public Health and Epidemiological Studies. *PLoS One* **10**, e0136179 (2015).

185. Tagmouti, S. *et al.* Reproducibility of Interferon Gamma (IFN-γ) Release Assays. A Systematic Review. *Ann. Am. Thorac. Soc.* **11**, 1267–1276 (2014).

186. Altman, D. G. & Bland, J. M. Measurement in Medicine: The Analysis of Method Comparison Studies. *Stat.* **32**, 307 (1983).

187. Gupta, R. K. *et al.* Evaluation of QuantiFERON-TB Gold Plus for Predicting Incident Tuberculosis among Recent Contacts: A Prospective Cohort Study. *Ann. Am. Thorac. Soc.* AnnalsATS.201905-407RL (2020) doi:10.1513/AnnalsATS.201905-407RL.

188. Shafeque, A., Bigio, J., Hogan, C. A., Pai, M. & Banaei, N. Fourth-generation quanti FERON-TB gold plus: What is the evidence? *Journal of Clinical Microbiology* vol. 58 (2020).

189. Sloot, R., Van Der Loeff, M. F. S., Kouw, P. M. & Borgdorff, M. W. Risk of tuberculosis after recent exposure: A 10-year follow-up study of contacts in Amsterdam. *Am. J. Respir. Crit. Care Med.* **190**, 1044–1052 (2014).

190. Haldar, P. *et al.* Single-step QuantiFERON screening of adult contacts: a prospective cohort study of tuberculosis risk. *Thorax* **68**, 240–246 (2013).

191. Erkens, C. G. M. *et al.* Risk of developing tuberculosis disease among persons diagnosed with latent tuberculosis infection in the Netherlands. *Eur. Respir. J.* **48**, 1420–1428 (2016).

192. Diel, R., Loddenkemper, R., Niemann, S., Meywald-Walter, K. & Nienhaus, A. Negative and positive predictive value of a whole-blood interferon-gamma release assay for developing active tuberculosis: an update. *Am. J. Respir. Crit. Care Med.* **183**, 88–95 (2011).

193. Kruijshaar, M. E. & Abubakar, I. Increase in extrapulmonary tuberculosis in England and Wales 1999-2006. *Thorax* **64**, 1090–5 (2009).

194. Mack, U. *et al.* LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. *Eur. Respir. J.* **33**, 956–973 (2009).

195. Stewart, L. A. *et al.* Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data. *JAMA* **313**, 1657 (2015).

196. Wells, G. *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

197. Riley, R. D., Windt, D. van der, Croft, P. & Moons, K. G. M. *Prognosis research in healthcare : concepts, methods, and impact*.

198. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Heal.* **22**, 153–160 (2019).

199. Spittal, M. J., Pirkis, J. & Gurrin, L. C. Meta-analysis of incidence rate data in the presence of zero events. *BMC Med. Res. Methodol.* **15**, 42 (2015).

200. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **350**, (2015).

201. Seddon, J. A. *et al.* The impact of BCG vaccination on tuberculin skin test responses in children is age dependent: evidence to be considered when

screening children for tuberculosis infection. *Thorax* **71**, 932–939 (2016).

202. Cobelens, F. G. *et al.* Tuberculin Skin Testing in Patients with HIV Infection: Limited Benefit of Reduced Cutoff Values. *Clin. Infect. Dis.* **43**, 634–639 (2006).

203. Gupta, R. K. *et al.* Discovery and validation of a personalized risk predictor for incident tuberculosis in low transmission settings. *Nat. Med.* **26**, 1941–1949 (2020).

204. Audigier, V. *et al.* Multiple imputation for multilevel data with continuous and binary variables. *Stat. Sci.* **33**, 160–183 (2018).

205. Audigier, V. & Resche-Rigon, M. micemd: Multiple Imputation by Chained Equations with Multilevel Data. (2019).

206. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* **30**, 377–399 (2011).

207. Rubin, D. B. *Multiple imputation for nonresponse in surveys*. (Wiley-Interscience, 2004).

208. Royston, P. & Parmar, M. K. B. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat. Med.* **21**, 2175–2197 (2002).

209. Clements, M. & Liu, X.-R. rstpm2: Smooth Survival Models, Including Generalized Survival Models. (2019).

210. Bower, H. *et al.* Capturing simple and complex time-dependent effects using flexible parametric survival models: A simulation study. *Commun. Stat. - Simul. Comput.* 1–17 (2019) doi:10.1080/03610918.2019.1634201.

211. Debray, T. P. A. *et al.* Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLOS Med.* **12**, e1001886 (2015).

212. Steyerberg, E. W. & Harrell, F. E. Prediction models need appropriate internal, internal–external, and external validation. *J. Clin. Epidemiol.* **69**, 245–247 (2016).

213. Westeneng, H.-J. *et al.* Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol.* **17**, 423–433 (2018).

214. Snell, K. I., Ensor, J., Debray, T. P., Moons, K. G. & Riley, R. D. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the *C* -statistic and calibration measures? *Stat. Methods Med. Res.* **27**, 3505–3522 (2018).

215. Vickers, A. J., van Calster, B. & Steyerberg, E. W. A simple, step-by-step guide to

interpreting decision curve analysis. *Diagnostic Progn. Res.* **3**, 18 (2019).

216. Zhang, Z. *et al.* Decision curve analysis: a technical note. *Ann. Transl. Med.* **6**, 19 (2018).

217. Vickers, A. J., Van Calster, B. & Steyerberg, E. W. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* **352**, (2016).

218. Sjoberg, D. D. dca: Decision Curve Analysis R package version 0.1.0.9000. (2020).

219. Altet, N. *et al.* Predicting the Development of Tuberculosis with the Tuberculin Skin Test and QuantiFERON Testing. *Ann. Am. Thorac. Soc.* **12**, 680–688 (2015).

220. Dobler, C. C. & Marks, G. B. Risk of tuberculosis among contacts in a low-incidence setting. *Eur. Respir. J.* **41**, 1459–1461 (2013).

221. Yoshiyama, T., Harada, N., Higuchi, K., Saitou, M. & Kato, S. Use of the QuantiFERON-TB Gold in Tube test for screening TB contacts and predictive value for active TB. *Infect. Dis. (London, England)* **47**, 542–549 (2015).

222. Aichelburg, M. C. *et al.* Detection and prediction of active tuberculosis disease by a whole-blood interferon-gamma release assay in HIV-1-infected individuals. *Clin. Infect. Dis.* **48**, 954–962 (2009).

223. Doyle, J. S. *et al.* Latent Tuberculosis screening using interferon-gamma release assays in an Australian HIV-infected cohort: is routine testing worthwhile?. *J. Acquir. Immune Defic. Syndr.* **66**, 48–54 (2014).

224. Lange, B., Vavra, M., Kern, W. V & Wagner, D. Development of tuberculosis in immunocompromised patients with a positive tuberculosis-specific IGRA. *Int. J. Tuberc. Lung Dis.* **16**, 492–495 (2012).

225. Sester, M. *et al.* Risk assessment of tuberculosis in immunocompromised patients. A TBNET study. *Am. J. Respir. Crit. Care Med.* **190**, 1168–1176 (2014).

226. Munoz, L. *et al.* Immunodiagnostic Tests' Predictive Values for Progression to Tuberculosis in Transplant Recipients: A Prospective Cohort Study. *Transplant. direct* **1**, e12 (2015).

227. Roth, D. Z. *et al.* Impact of interferon-gamma release assay on the latent tuberculosis cascade of care: A population-based study. *Eur. Respir. J.* **49**, 1601546 (2017).

228. Zenner, D., Loutet, M. G., Harris, R., Wilson, S. & Ormerod, L. P. Evaluating 17 years of latent tuberculosis infection screening in north-west England: a retrospective cohort study of reactivation. *Eur. Respir. J.* **50**, (2017).

229. Bergot, E. *et al.* Observational study of quantiferon-TB gold in-tube assay in

tuberculosis contacts in a low incidence area. *PLoS One* **7**, e43520 (2012).

230. Blount, R. J. *et al.* Tuberculosis progression rates in U.S. Immigrants following screening with interferon-gamma release assays. *BMC Public Health* **16**, 875 (2016).

231. Hand, J., Sigel, K., Huprikar, S., Hamula, C. & Rana, M. Tuberculosis after liver transplantation in a large center in New York City: QuantiFERON -TB Gold-based pre-transplant screening performance and active tuberculosis post-transplant. *Transpl. Infect. Dis.* **20**, e12845 (2018).

232. Harstad, I., Winje, B. A., Heldal, E., Oftung, F. & Jacobsen, G. W. Predictive values of QuantiFERON-TB Gold testing in screening for tuberculosis disease in asylum seekers. *Int. J. Tuberc. Lung Dis.* **14**, 1209–1211 (2010).

233. Kik, S. V *et al.* Predictive value for progression to tuberculosis by IGRA and TST in immigrant contacts. *Eur. Respir. J.* **35**, 1346–1353 (2010).

234. Pullar, N. D., Steinum, H., Bruun, J. N. & Dyrhol-Riise, A. M. HIV patients with latent tuberculosis living in a low-endemic country do not develop active disease during a 2 year follow-up; a Norwegian prospective multicenter study. *BMC Infect. Dis.* **14**, 667 (2014).

235. Reichler, M. R. *et al.* Risk and Timing of Tuberculosis Among Close Contacts of Persons with Infectious Tuberculosis. *J. Infect. Dis.* **218**, 1000–1008 (2018).

236. Saunders, M. J. *et al.* A score to predict and stratify risk of tuberculosis in adult contacts of tuberculosis index cases: a prospective derivation and external validation cohort study. *Lancet. Infect. Dis.* **17**, 1190–1199 (2017).

237. Saunders, M. J. *et al.* A household-level score to predict the risk of tuberculosis among contacts of patients with tuberculosis: a derivation and external validation prospective cohort study. *Lancet Infect. Dis.* **20**, 110–122 (2020).

238. Li, R. *et al.* Two Clinical Prediction Tools to Improve Tuberculosis Contact Investigation. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciz1221.

239. Farmer, G. D., Pearson, G. M., Skylark, W. J., Freeman, A. L. J. & Spiegelhalter, D. J. Making prognostic algorithms useful in shared decision-making: Patients and clinicians' requirements for the Predict:Breast Cancer interface. *medRxiv* 2020.11.16.20232348 (2020) doi:10.1101/2020.11.16.20232348.

240. Nemes, E. *et al.* Prevention of M. Tuberculosis infection with H4:IC31 vaccine or BCG revaccination. *N. Engl. J. Med.* **379**, 138–149 (2018).

241. Riley, R. D., Lambert, P. C. & Abo-Zaid, G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* **340**, c221 (2010).

242. Groenwold, R. H. H. *et al.* Explicit inclusion of treatment in prognostic modeling

was recommended in observational and randomized settings. *J. Clin. Epidemiol.* **78**, 90–100 (2016).

243. Bloom, C. I. *et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS One* **8**, e70630 (2013).

244. Maertzdorf, J. *et al.* Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol. Med.* **8**, 86–95 (2016).

245. Rajan, J. V *et al.* A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus. *Clin. Infect. Dis.* (2018) doi:10.1093/cid/ciy835.

246. Suliman, S. *et al.* Four-gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *Am. J. Respir. Crit. Care Med.* **197**, 1198–1208 (2018).

247. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

248. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).

249. Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589–W598 (2015).

250. Jollife, I. T. & Cadima, J. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* vol. 374 (2016).

251. Sweeney, T. E., Wong, H. R. & Khatri, P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci. Transl. Med.* **8**, 346ra91-346ra91 (2016).

252. Leek, J. *et al.* sva: Surrogate Variable Analysis. (2019).

253. Qian, Z. *et al.* Expression of nuclear factor, erythroid 2-like 2-mediated genes differentiates tuberculosis. *Tuberculosis (Edinb).* **99**, 56–62 (2016).

254. Morpheus. https://software.broadinstitute.org/morpheus/.

255. Reitsma, J. B. *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidemiol.* **58**, 982–990 (2005).

256. Gupta, R. K. *et al.* Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *Lancet. Respir. Med.* **0**, (2020).

257. Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in

high-throughput data. *Nat. Rev. Genet.* **11**, 733–739 (2010).

258. Huang, H.-H., Liu, X.-Y., Liang, Y., Chai, H. & Xia, L.-Y. Identification of 13 blood-based gene expression signatures to accurately distinguish tuberculosis from other pulmonary diseases and healthy controls. *Biomed. Mater. Eng.* **26 Suppl 1**, S1837-43 (2015).

259. Södersten, E. *et al.* Diagnostic accuracy study of a novel blood-based assay # for identification of TB in people living with HIV . *J. Clin. Microbiol.* (2020) doi:10.1128/jcm.01643-20.

260. Lin, P. L. *et al.* Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nat. Med.* **20**, 75–79 (2014).

261. Campbell, J. R., Dowdy, D. & Schwartzman, K. Treatment of latent infection to achieve tuberculosis elimination in low-incidence countries. *PLOS Med.* **16**, e1002824 (2019).

262. Duc, A. N. & Wolbers, M. Weighted analysis of composite endpoints with simultaneous inference for flexible weight constraints. *Stat. Med.* **36**, 442–454 (2017).

263. Turner, C. T. *et al.* Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. *Lancet Respir. Med.* **8**, 407–419 (2020).

264. Gupta, R. K. *et al.* Blood transcriptional biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection. *medRxiv* 2021.01.18.21250044 (2021) doi:10.1101/2021.01.18.21250044.

265. Gupta, R. K. *et al.* Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: An observational cohort study. *Eur. Respir. J.* (2020) doi:10.1183/13993003.03498-2020.

266. Wynants, L. *et al.* Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* **369**, (2020).

8 Appendix: Supplementary material

8.1 Supplementary material for Chapter 4

Table 8-1: Look-up table for transformation from raw latent TB test results to normalised percentile scale.

Percentile	QuantiFERON (IU / mL)	T-SPOT.TB (spots)	TST (mm)
1	0 to 0	0 to 0	0 to 1
39	0.001 to 0.009		
40	0.01 to 0.019		
48		1 to 1	
49	0.02 to 0.029		
52			2 to 2
53			3 to 3
54	0.03 to 0.039		
55			4 to 4
57	0.04 to 0.049		5 to 5
60	0.05 to 0.059		6 to 6
62	0.06 to 0.069		7 to 7
63	0.07 to 0.079		
64		2 to 2	8 to 8
65	0.08 to 0.089		
66	0.09 to 0.099		9 to 9
67	0.1 to 0.119		
68			10 to 10
69	0.12 to 0.129		
70	0.13 to 0.149	3 to 3	
71	0.15 to 0.169		
72	0.17 to 0.199		11 to 11
73	0.2 to 0.219		
74	0.22 to 0.259	4 to 4	12 to 12
75	0.26 to 0.299		
76	0.3 to 0.349	5 to 5	
77	0.35 to 0.399		13 to 13
78	0.4 to 0.449	6 to 6	
79	0.45 to 0.519		14 to 14
80	0.52 to 0.599	7 to 7	
81	0.6 to 0.689	8 to 8	15 to 15
82	0.69 to 0.809	9 to 9	
83	0.81 to 0.919	10 to 10	
84	0.92 to 1.069	11 to 11	16 to 16
85	1.07 to 1.269	12 to 13	
86	1.27 to 1.519	14 to 15	17 to 17
87	1.52 to 1.769	16 to 17	
88	1.77 to 2.059	18 to 20	18 to 18
89	2.06 to 2.479	21 to 24	19 to 19
90	2.48 to 2.939	25 to 28	
91	2.94 to 3.549	29 to 34	20 to 20
92	3.55 to 4.279	35 to 40	
93	4.28 to 5.099	41 to 49	
94	5.1 to 6.289	50 to 58	21 to 21
95	6.29 to 7.589	59 to 70	22 to 22
96	7.59 to 9.179	71 to 87	23 to 24
97	9.18 to 9.999	88 to 109	25 to 26
98	≥10	110 to 152	27 to 30
99		≥153	≥31

Table 8-2: Data missingness of candidate predictors considered for inclusion in prognostic model, stratified by source study.

Numbers and colours indicate % missing for each variable.

Study	Age	Sex	BCG	Exposure	Months since migration	Quantitative test result	LTBI treatment	HIV
Abubakar 2018	0.2	0.7	15.7	22.3	1.2	0.0	0.0	6.9
Aichelburg 2009	0.0	0.0	100.0	0.0	13.5	94.9	0.0	0.0
Altet 2015	0.1	0.1	0.1	0.0	0.0	0.0	0.2	0.0
Diel 2011	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Doyle 2014	0.0	0.1	100.0	5.0	20.6	0.0	0.0	0.0
Geis 2013	0.9	2.5	50.0	6.6	0.0	0.0	0.1	100.0
Gupta 2020	0.9	1.9	21.1	46.5	0.0	0.0	0.0	4.1
Haldar 2013	0.0	0.0	28.1	0.0	0.0	9.9	0.0	100.0
Lange 2012	0.0	0.0	100.0	29.4	30.6	0.0	0.0	0.3
Munoz 2015	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0
Roth 2017	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sester 2014	0.7	0.0	26.1	0.0	1.0	0.0	0.3	0.0
Sloot 2014	0.1	1.8	75.2	21.3	0.0	0.0	0.0	100.0
Yoshiyama 2015	0.8	100.0	100.0	0.8	0.0	21.3	1.3	100.0
Zellweger 2015	0.5	0.0	100.0	0.0	0.0	4.1	0.0	100.0
All	0.3	2.7	44.0	12.4	1.6	3.5	0.1	46.5

Table 8-3: Further characteristics of contributing studies included in IPD-MA.

Shown as n (%) or median (interquartile range (IQR)) as appropriate.

(a) Demographics, latent TB testing and follow-up.

Authors	Age (median (IQR))	Female (%)	LTBI positive (%)	LTBI tests used	LTBI treatment (%)	TB cases	Culture confirmed	Days to TB	Pulmonary TB	Prevalent TB definition (days)^
Abubakar et al.	32 (26-46)	5007 (50.2%)	4003 (40.7%)	TST, QFT-GIT, T-SPOT.TB	260 (2.6%)	147	61 (54%)	112 (21-366)	63 (48.1%)	21
Aichelburg et al.	39 (32-47)	249 (30%)	44 (5.6%)	QFT-GIT	0 (0%)	11	10 (90.9%)	0 (0-253)	10 (90.9%)	N/A
Altet et al.	23 (11-36)	665 (49.7%)	807 (60.3%)	TST, QFT-GIT, T-SPOT.TB	625 (46.7%)	95	78 (83%)	0 (0-0)	90 (94.7%)	48*
Diel et al.	30 (20-40)	684 (48.4%)	511 (36.1%)	QFT-GIT, T-SPOT.TB	67 (4.7%)	19	Not recorded	490 (112-560)	Not recorded	N/A
Dobler & Marks	29 (17-45)	6458 (54.8%)	4067 (33.3%)	TST	499 (4.1%)	94	Not recorded	42.5 (12-109)	24 (70.6%)	90
Doyle et al.	41 (33-48)	110 (12%)	29 (3.2%)	QFT-Gold	18 (2%)	2	2 (100%)	277.5 (10-545)	2 (100%)	N/A
Erkens et al.	33 (22-47)	7245 (50.9%)	14241 (100%)	TST, IGRA for subset	10038 (70.5%)	134	76 (56.7%)	280 (105-647)	113 (84.3%)	28
Geis et al.	35 (23-49)	589 (46.9%)	1283 (100%)	QFT-GIT, TST	277 (21.7%)	33	15 (100%)	133 (49-287)	Not recorded	N/A
Gupta et al.	33 (25-46)	308 (50.5%)	180 (33.8%)	QFT-Plus	39 (6.3%)	13	7 (53.8%)	222 (182-342)	3 (23.1%)	21
Haldar et al.	21 (9-31)	683 (48.4%)	287 (20.7%)	QFT-GIT	170 (12%)	37	16 (43.2%)	87 (12-141)	20 (54.1%)	N/A
Lange et al.	51 (40-62)	201 (44.1%)	42 (10.5%)	QFT-GIT	0 (0%)	1	1 (100%)	746 (746-746)	0 (0%)	N/A
Munoz et al.	56 (49-63)	30 (39.5%)	37 (48.7%)	QFT-GIT, TST	0 (0%)	2	2 (100%)	324 (116-532)	0 (0%)	N/A
Roth et al.	39 (29-49)	14432 (62.9%)	22949 (100%)	TST, IGRA for subset	5895 (25.7%)	58	58 (100%)	871.1 (478.5-1278.4)	45 (77.6%)	30
Sester et al.	48 (38-61)	549 (37.5%)	374 (25.6%)	TST, QFT-GIT, T-SPOT.TB	126 (8.6%)	11	8 (72.7%)	653 (61-1068)	9 (81.8%)	N/A
Sloot et al.	32 (18-45)	3015 (52.1%)	876 (14.9%)	TST, QFT-GIT for subset	380 (6.4%)	81	Not recorded	63 (4-123)	Not recorded	138*
Yoshiyama et al.	40 (30-50)	Not recorded	144 (23%)	QFT-GIT	115 (18.6%)	12	7 (100%)	375.5 (110.5-843)	11 (91.7%)	N/A
Zellweger et al.	34 (24-46)	2829 (54%)	1457 (28%)	QFT-GIT, T-SPOT.TB	1054 (20.3%)	55	Not recorded	25 (4-97)	Not recorded	39*
Zenner et al.	25 (22-29)	620 (46.3%)	366 (27.3%)	QFT-GIT	243 (18.1%)	21	13 (100%)	86 (50-716)	10 (47.6%)	60
Total	**34 (24-47)**	**43674 (53.9%)**	**51697 (63.1%)**		**19806 (24.1%)**	**826**	**354 (69.3%)**	**98 (12-392)**	**400 (71.2%)**	**39 (28-60)**

*^Defined as days from LTBI testing to TB diagnosis. Some studies (indicated by *) reported definitions in days from index case diagnosis for recent TB contacts. 42 days has been subtracted from the prevalent TB definition for these studies, based on the assumption that contacts are screened at least 6 weeks from exposure. Median (IQR) shown in 'Total' column.*

(b) Indication for latent TB testing

Authors	Contacts	High risk contacts*	Migrants	TB incidence in country of birth (migrants)^	HIV	Viral Load (log10 copies/mL)	CD4 (cells/mm³)	ART status	Other immunocompromise
Abubakar et al.	5095 (51.3%)	494 (16.7%)	4845 (48.7%)	254 (223-289)	0 (0%)	N/A	N/A	N/A	0 (0%)
Aichelburg et al.	0 (0%)	N/A	0 (0%)	219 (215-366)	830 (100%)	1.7 (1.7-4.4)	392.5 (264-566)	495 (59.6%)	0 (0%)
Altet et al.	1339 (100%)	609 (45.5%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Diel et al.	1414 (100%)	514 (36.4%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Dobler & Marks	12212 (100%)	Not recorded	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Doyle et al.	0 (0%)	N/A	0 (0%)	366 (240-559)	919 (100%)	1.7 (1.7-4.3)	490.6 (348-676.8)	N/A	0 (0%)
Erkens et al.	8499 (91.9%)	Not recorded	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	745 (8.1%)
Geis et al.	1283 (100%)	173 (14.4%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Gupta et al.	623 (100%)	76 (23.9%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Haldar et al.	1411 (100%)	183 (13%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Lange et al.	0 (0%)	N/A	0 (0%)	137.5 (132-175)	45 (9.9%)	N/A	N/A	N/A	411 (90.1%)
Munoz et al.	0 (0%)	N/A	0 (0%)	120 (120-120)	0 (0%)	N/A	N/A	N/A	76 (100%)
Roth et al.	Not recorded	0 (0%)	Not recorded	N/A	Not recorded	N/A	N/A	N/A	Not recorded
Sester et al.	0 (0%)	N/A	0 (0%)	219 (176-327)	713 (48.7%)	2.5 (1.7-4.3)	399.5 (227-598)	486 (68.2%)	751 (51.3%)
Sloot et al.	5895 (100%)	1272 (27.4%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Yoshiyama et al.	625 (100%)	116 (18.7%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Zellweger et al.	5237 (100%)	839 (16%)	0 (0%)	N/A	0 (0%)	N/A	N/A	N/A	0 (0%)
Zenner et al.	0 (0%)	N/A	1341 (100%)	276 (254-276)	0 (0%)	N/A	N/A	N/A	0 (0%)
Total	**43633 (80.3%)**	**4276 (21.5%)**	**6186 (11.4%)**	**261 (228-282)**	**2507 (4.6%)**	**1.7 (1.7-4.3)**	**431 (284.9-616)**	**981 (63.6%)[$]**	**1983 (3.7%)**

**Defined as household contacts of sputum smear positive index cases. Percentage reflects proportion of contacts who are high risk, where known.*

^Annual incidence per 100,000 persons.

[$]Percentage reflects percentage of those living with HIV

Table 8-4: Baseline characteristics of participants lost to follow-up in studies with >5% lost.

		Lost to follow-up	
		No	*Yes*
Diel 2011	*n*	1033	381
Age	*Median (IQR)*	28.00 [19.00, 38.00]	35.00 [24.00, 42.00]
Sex	*Female*	503 (48.7)	181 (47.5)
	Male	530 (51.3)	200 (52.5)
Exposure	*Household, smear+*	365 (35.3)	149 (39.1)
	Other contacts	668 (64.7)	232 (60.9)
LTBI	*Negative*	670 (64.9)	233 (61.2)
	Positive	363 (35.1)	148 (38.8)
LTBI treatment	*No*	982 (95.1)	365 (95.8)
	Yes	51 (4.9)	16 (4.2)
Doyle 2014	*n*	872	47
Age	*Median (IQR)*	40.95 [33.48, 48.40]	38.10 [29.20, 45.20]
Sex	*Female*	101 (11.6)	9 (19.1)
	Male	770 (88.3)	38 (80.9)
	Missing	1 (0.1)	0 (0.0)
Exposure	*No contact, migrant*	143 (16.4)	11 (23.4)
	No contact, non-migrant	686 (78.7)	30 (63.8)
	Missing	43 (4.9)	6 (12.8)
LTBI	*Negative*	842 (96.6)	42 (89.4)
	Positive	26 (3.0)	3 (6.4)
	Missing	4 (0.5)	2 (4.3)
LTBI treatment	*No*	850 (97.5)	43 (91.5)
	Yes	17 (1.9)	1 (2.1)
	Missing	5 (0.6)	3 (6.4)
HIV	*Yes*	872 (100.0)	47 (100.0)
Lange 2012	*n*	414	42
Age	*Median (IQR)*	51.00 [41.00, 63.00]	47.00 [37.00, 59.00]
Sex	*Female*	177 (42.8)	24 (57.1)
	Male	237 (57.2)	18 (42.9)
Exposure	*No contact, migrant*	6 (1.4)	2 (4.8)
	No contact, non-migrant	281 (67.9)	17 (40.5)
	Missing	127 (30.7)	23 (54.8)
LTBI	*Negative*	332 (80.2)	25 (59.5)
	Positive	35 (8.5)	7 (16.7)
	Missing	47 (11.4)	10 (23.8)
LTBI treatment	*No*	414 (100.0)	42 (100.0)
HIV	*No*	368 (88.9)	42 (100.0)
	Yes	45 (10.9)	0 (0.0)
	Missing	1 (0.2)	0 (0.0)
Transplant	*No*	189 (45.7)	34 (81.0)
	Yes	221 (53.4)	8 (19.0)
	Missing	4 (1.0)	0 (0.0)
Zellweger 2015	*n*	3898	1339
Age	*Median (IQR)*	34.00 [23.00, 46.00]	34.00 [24.00, 48.00]
Sex	*Female*	2171 (55.7)	658 (49.1)
	Male	1727 (44.3)	681 (50.9)

Exposure	*Household, smear+*	653 (16.8)	186 (13.9)
	Other contacts	3244 (83.2)	1153 (86.1)
	Missing	1 (0.0)	0 (0.0)
LTBI	*Negative*	2706 (69.4)	1045 (78.0)
	Positive	1170 (30.0)	287 (21.4)
	Missing	22 (0.6)	7 (0.5)
LTBI treatment	*No*	3087 (79.2)	1059 (79.1)
	Yes	811 (20.8)	243 (18.1)
	Missing	0 (0.0)	37 (2.8)

Table 8-5: Quality assessment of studies contributing individual participant data.

Selection	Abubakar et al	Aichelburg et al	Altet et al	Diel et al	Dobler & Marks	Doyle et al	Erkens et al	Geis et al	Gupta et al	Haldar et al	Lange et al	Munoz et al	Roth et al	Sester et al	Sloot et al	Yoshiyama et al	Zellweger et al	Zenner et al
1) Representativeness of the exposed cohort																		
a) truly representative of the target population in the community Ø	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
b) somewhat representative of the target population in the community Ø																		
c) selected group of users eg nurses, volunteers																		
d) no description of the derivation of the cohort																		
2) Selection of the non exposed cohort																		
a) drawn from the same community as the exposed cohort Ø	✓	✓	✓	✓	✓	✓	N/A	N/A	✓	✓	✓	✓	N/A	✓	✓	✓	✓	✓
b) drawn from a different source																		
c) no description of the derivation of the non exposed cohort																		
3) Ascertainment of LTBI test result																		
a) secure record Ø	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
b) structured interview Ø																		
c) written self report																		
d) no description																		
4) Assessment for prevalent TB at start of study																		
a) yes Ø	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
b) no																		
Comparability																		
1) Comparability of cohorts on the basis of the design or analysis	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Outcome																		
1) Assessment of outcome																		
a) independent assessment blinded to LTBI test resultØ																		
b) data available on microbiological confirmationØ		✓	✓					✓		✓	✓	✓		✓		✓	✓	
b) record linkage Ø	✓			✓	✓	✓	✓		✓				✓		✓			✓
c) self report																		
d) no description																		
2) Was follow-up long enough for outcomes to occur																		
a) yes (median >1 year) Ø	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
b) no																		
3) Adequacy of follow up of cohorts																		
a) complete follow up - all subjects accounted for Ø	✓		✓						✓			✓	✓		✓			✓
b) subjects lost to follow up unlikely to introduce biasØ: - small number lost (<5%); - or characteristics of those lost similar to those followed-up)		✓		✓	✓	✓	✓	✓		✓	✓			✓			✓	
c) follow up rate < 95% and no description of those lost																✓		
d) no statement																		
Total	7	7	7	7	7	7	6	6	7	7	7	7	6	7	7	6	7	7
Maximum score	7	7	7	7	7	7	6	6	7	7	7	7	6	7	7	7	7	7

Using modified version of Newcastle-Ottawa Scale for cohort studies.

8.2 Supplementary material for Chapter 5

Table 8-6: Quality assessment of studies included in RNA biomarkers IPD-MA.

Quality assessment of four studies representing datasets included in IPD-MA, using Newcastle-Ottawa scale for (a) case-control studies; or (b) cohort studies, as appropriate. ACS = adolescent cohort study; GC6-74 = Bill and Melinda Gates Foundation Grand Challenges 6-74 TB contacts study. Material from [256].

(a)

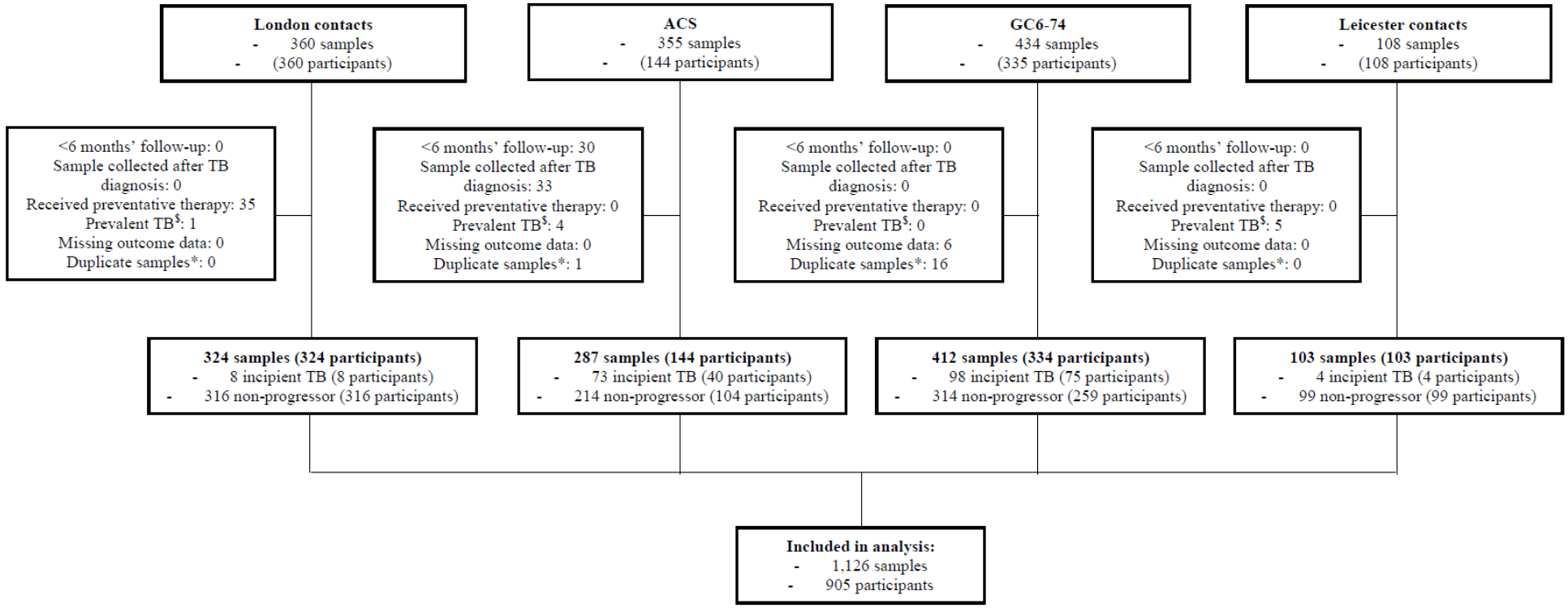
Case-control studies	**GC6-74**	**Comments**	**ACS**	**Comments**
Selection				
1) Is the case definition adequate?				
a) yes, with independent validation Ø	✓	Detailed data available	✓	Microbiologically confirmed only
b) yes, eg record linkage or based on self reports				
c) no description				
2) Representativeness of the cases				
a) consecutive or obviously representative series of cases Ø	✓		✓	
b) potential for selection biases or not stated				
3) Selection of Controls				
a) community controls Ø	✓		✓	
b) hospital controls				
c) no description				
4) Definition of Controls				
a) no history of disease (endpoint) Ø	✓		✓	
b) no description of source				
Comparability				
1) Comparability of cases and controls on the basis of the design or analysis				
a) study controls for age Ø	✓		✓	
b) study controls for any additional factor Ø	✓	Recruitment region, sex, enrolment year	✓	Gender, ethnicity, school, previous TB
Exposure				
1) Ascertainment of exposure	✓	Raw RNAseq data available	✓	Raw RNAseq data available
a) secure record (eg surgical records) Ø				
b) structured interview where blind to case/control status Ø				
c) interview not blinded to case/control status				
d) written self report or medical record only				
e) no description				
2) Same method of ascertainment for cases and controls				
a) yes Ø	✓	Calculated using RNAseq data	✓	Calculated using RNAseq data
b) no				
3) Non-Response rate				
a) same rate for both groups Ø	✓		✓	

(b)

Selection	London Contacts	Comments	Leicester Contacts	Comments
1) Representativeness of the exposed cohort				
a) truly representative of the average TB contact in the community Ø	✓		✓	
b) somewhat representative of the average TB contact in the community Ø				
c) selected group of users eg nurses, volunteers				
d) no description of the derivation of the cohort				
2) Selection of the non exposed cohort				
a) drawn from the same community as the exposed cohort Ø	✓	Based on RNAseq data	✓	Based on RNAseq data
b) drawn from a different source				
c) no description of the derivation of the non exposed cohort				
3) Ascertainment of exposure				
a) secure record (eg surgical records) Ø	✓	Based on RNAseq data	✓	Based on RNAseq data
b) structured interview Ø				
c) written self report				
d) no description				
4) Demonstration that outcome of interest was not present at start of study				
a) yes Ø	✓		✓	
b) no				
Comparability				
1) Comparability of cohorts on the basis of the design or analysis	N/A		N/A	
Outcome				
1) Assessment of outcome				
a) independent blind assessment Ø			✓	
b) record linkage Ø	✓			
c) self report				
d) no description				
2) Was follow-up long enough for outcomes to occur				
a) yes (>1 year) Ø	✓	Median 1.9 years	✓	2 years (clarified with authors)
b) no				
3) Adequacy of follow up of cohorts				

Figure 8-1: Inclusion of samples from contributing datasets in RNA biomarkers IPD-MA.

ACS = adolescent cohort study; GC6-74 = Bill and Melinda Gates Foundation Grand Challenges 6-74 TB contacts study. Material from [256].

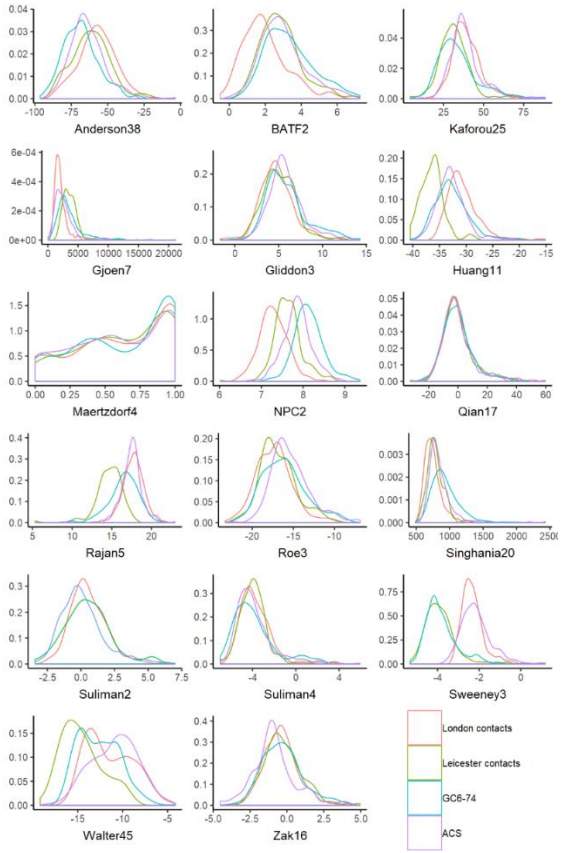


*$Prevalent TB defined as TB diagnosed within 21 days of sample collection. *Indicates >1 sample collected from the same participant within a 6-month interval.*

Figure 8-2: Density plots of RNA signature expression (a) before and (b) after batch correction, stratified by source study.

ACS = adolescent cohort study; GC6-74 = Bill and Melinda Gates Foundation Grand Challenges 6-74 TB contacts study. Material from [256].

A



B

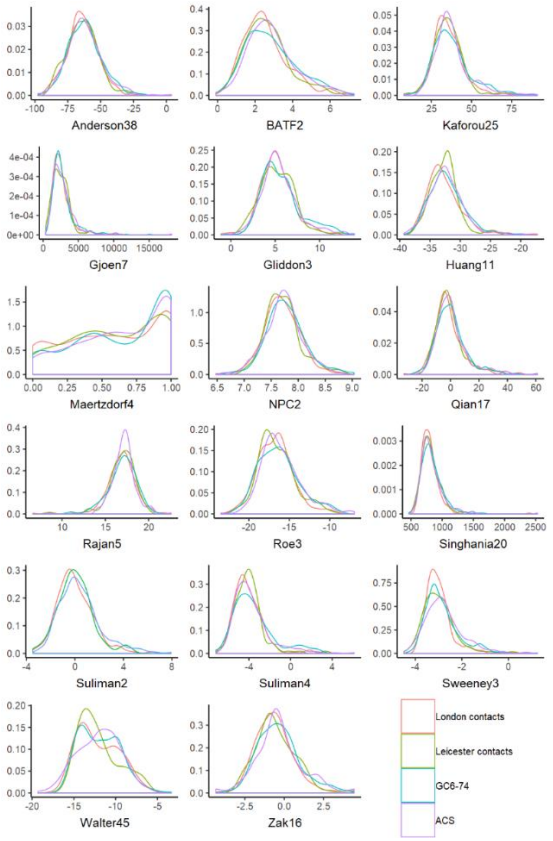


Table 8-7: AUROCs (95% CIs) for 17 RNA signatures for identification of incipient TB over a two-year period, stratified by (a) study, and (b) study and time interval to disease.

ACS = adolescent cohort study; GC6-74 = Bill and Melinda Gates Foundation Grand Challenges 6-74 study.

(a)

Signature	ACS	GC6-74	London contacts	Leicester contacts
Anderson38	0.71 (0.63 - 0.79)	0.63 (0.56 - 0.7)	0.72 (0.56 - 0.88)	0.66 (0.41 - 0.9)
BATF2	0.81 (0.74 - 0.88)	0.68 (0.61 - 0.76)	0.81 (0.61 - 1)	0.71 (0.33 - 1)
Gjoen7	0.72 (0.64 - 0.8)	0.64 (0.57 - 0.71)	0.83 (0.71 - 0.94)	0.66 (0.28 - 1)
Gliddon3	0.72 (0.64 - 0.8)	0.74 (0.67 - 0.8)	0.84 (0.62 - 1)	0.66 (0.21 - 1)
Huang11	0.69 (0.61 - 0.77)	0.67 (0.6 - 0.74)	0.66 (0.46 - 0.85)	0.62 (0.25 - 0.98)
Kaforou25	0.79 (0.72 - 0.86)	0.7 (0.63 - 0.77)	0.84 (0.64 - 1)	0.66 (0.22 - 1)
Maertzdorf4	0.75 (0.68 - 0.83)	0.63 (0.56 - 0.7)	0.81 (0.67 - 0.95)	0.66 (0.23 - 1)
NPC2	0.64 (0.55 - 0.72)	0.69 (0.62 - 0.75)	0.84 (0.71 - 0.97)	0.8 (0.64 - 0.96)
Qian17	0.7 (0.62 - 0.78)	0.61 (0.54 - 0.68)	0.77 (0.61 - 0.94)	0.71 (0.44 - 0.97)
Rajan5	0.7 (0.62 - 0.78)	0.54 (0.47 - 0.61)	0.45 (0.17 - 0.72)	0.53 (0.32 - 0.74)
Roe3	0.83 (0.76 - 0.89)	0.64 (0.56 - 0.71)	0.79 (0.6 - 0.99)	0.74 (0.33 - 1)
Singhania20	0.69 (0.61 - 0.77)	0.66 (0.6 - 0.73)	0.73 (0.61 - 0.85)	0.45 (0.12 - 0.78)
Suliman2	0.8 (0.74 - 0.87)	NA	0.8 (0.63 - 0.98)	0.62 (0.21 - 1)
Suliman4	0.75 (0.67 - 0.83)	0.63 (0.52 - 0.74)	0.85 (0.72 - 0.98)	0.7 (0.28 - 1)
Sweeney3	0.78 (0.71 - 0.85)	0.69 (0.62 - 0.76)	0.75 (0.55 - 0.95)	0.65 (0.22 - 1)
Walter45	0.59 (0.5 - 0.68)	0.54 (0.47 - 0.62)	0.49 (0.27 - 0.71)	0.53 (0.35 - 0.71)
Zak16	0.69 (0.5 - 0.88)	0.67 (0.6 - 0.74)	0.79 (0.59 - 0.99)	0.76 (0.43 - 1)

(b)

Months to TB	12 to 24		6 to 12		3 to 6		0 to 3	
Study	**GC6-74**	**ACS**	**GC6-74**	**ACS**	**GC6-74**	**ACS**	**GC6-74**	**ACS**
Anderson38	0.58 (0.47 - 0.69)	0.71 (0.59 - 0.82)	0.72 (0.6 - 0.85)	0.65 (0.54 - 0.77)	0.64 (0.54 - 0.75)	0.59 (0.26 - 0.92)	0.61 (0.35 - 0.88)	0.8 (0.68 - 0.93)
BATF2	0.62 (0.5 - 0.74)	0.72 (0.61 - 0.83)	0.73 (0.6 - 0.86)	0.8 (0.7 - 0.9)	0.72 (0.62 - 0.83)	0.82 (0.57 - 1)	0.79 (0.55 - 1)	0.9 (0.82 - 0.98)
Gjoen7	0.58 (0.47 - 0.7)	0.69 (0.58 - 0.8)	0.7 (0.57 - 0.83)	0.68 (0.56 - 0.81)	0.68 (0.58 - 0.77)	0.76 (0.6 - 0.93)	0.68 (0.37 - 1)	0.74 (0.55 - 0.94)
Gliddon3	0.64 (0.53 - 0.76)	0.6 (0.49 - 0.7)	0.8 (0.68 - 0.92)	0.73 (0.64 - 0.82)	0.76 (0.67 - 0.85)	0.76 (0.45 - 1)	0.81 (0.57 - 1)	0.85 (0.7 - 1)
Huang11	0.56 (0.43 - 0.68)	0.59 (0.47 - 0.7)	0.78 (0.68 - 0.88)	0.72 (0.63 - 0.81)	0.6 (0.49 - 0.71)	0.69 (0.25 - 1)	0.92 (0.85 - 0.99)	0.7 (0.53 - 0.87)
Kaforou25	0.61 (0.49 - 0.73)	0.69 (0.58 - 0.79)	0.78 (0.65 - 0.91)	0.79 (0.71 - 0.87)	0.73 (0.64 - 0.83)	0.81 (0.57 - 1)	0.84 (0.63 - 1)	0.89 (0.78 - 1)
Maertzdorf4	0.55 (0.43 - 0.67)	0.66 (0.56 - 0.77)	0.76 (0.64 - 0.88)	0.74 (0.62 - 0.86)	0.65 (0.54 - 0.75)	0.86 (0.73 - 0.98)	0.71 (0.4 - 1)	0.85 (0.73 - 0.97)
NPC2	0.61 (0.5 - 0.71)	0.59 (0.47 - 0.72)	0.68 (0.53 - 0.83)	0.63 (0.52 - 0.75)	0.73 (0.64 - 0.81)	0.67 (0.32 - 1)	0.9 (0.8 - 1)	0.69 (0.5 - 0.89)
Qian17	0.55 (0.43 - 0.67)	0.57 (0.45 - 0.7)	0.7 (0.58 - 0.82)	0.74 (0.64 - 0.85)	0.62 (0.51 - 0.73)	0.76 (0.39 - 1)	0.78 (0.58 - 0.99)	0.78 (0.64 - 0.91)
Rajan5	0.55 (0.44 - 0.66)	0.6 (0.49 - 0.71)	0.48 (0.34 - 0.61)	0.72 (0.62 - 0.82)	0.52 (0.41 - 0.62)	0.68 (0.31 - 1)	0.64 (0.35 - 0.93)	0.77 (0.61 - 0.93)
Roe3	0.55 (0.43 - 0.67)	0.75 (0.65 - 0.84)	0.7 (0.56 - 0.84)	0.83 (0.75 - 0.91)	0.71 (0.6 - 0.81)	0.82 (0.53 - 1)	0.77 (0.51 - 1)	0.91 (0.85 - 0.98)
Singhania20	0.64 (0.53 - 0.75)	0.66 (0.53 - 0.78)	0.63 (0.51 - 0.76)	0.65 (0.53 - 0.77)	0.7 (0.62 - 0.78)	0.65 (0.44 - 0.86)	0.65 (0.37 - 0.93)	0.81 (0.65 - 0.98)
Suliman4	0.61 (0.42 - 0.81)	0.68 (0.56 - 0.8)	0.66 (0.47 - 0.84)	0.7 (0.58 - 0.82)	0.75 (0.56 - 0.93)	0.83 (0.58 - 1)	0.81 (0.64 - 0.97)	0.91 (0.84 - 0.97)
Sweeney3	0.59 (0.48 - 0.71)	0.7 (0.6 - 0.8)	0.78 (0.66 - 0.89)	0.73 (0.62 - 0.83)	0.69 (0.58 - 0.8)	0.81 (0.5 - 1)	0.91 (0.77 - 1)	0.9 (0.79 - 1)
Walter45	0.54 (0.43 - 0.65)	0.45 (0.33 - 0.57)	0.53 (0.4 - 0.65)	0.57 (0.44 - 0.69)	0.69 (0.58 - 0.79)	0.55 (0.31 - 0.79)	0.53 (0.19 - 0.87)	0.47 (0.3 - 0.64)
Zak16	0.58 (0.46 - 0.69)	0.42 (0.19 - 0.65)	0.71 (0.57 - 0.85)	0.67 (0.36 - 0.99)	NA	NA	0.76 (0.47 - 1)	0.94 (0.86 - 1)

Figure 8-3: Density plots of signature expression of eight best performing RNA signatures for incipient TB, among control population.

Plots include participants with negative interferon-gamma release assay tests only, with approximately Normal distributions. Material from [256].

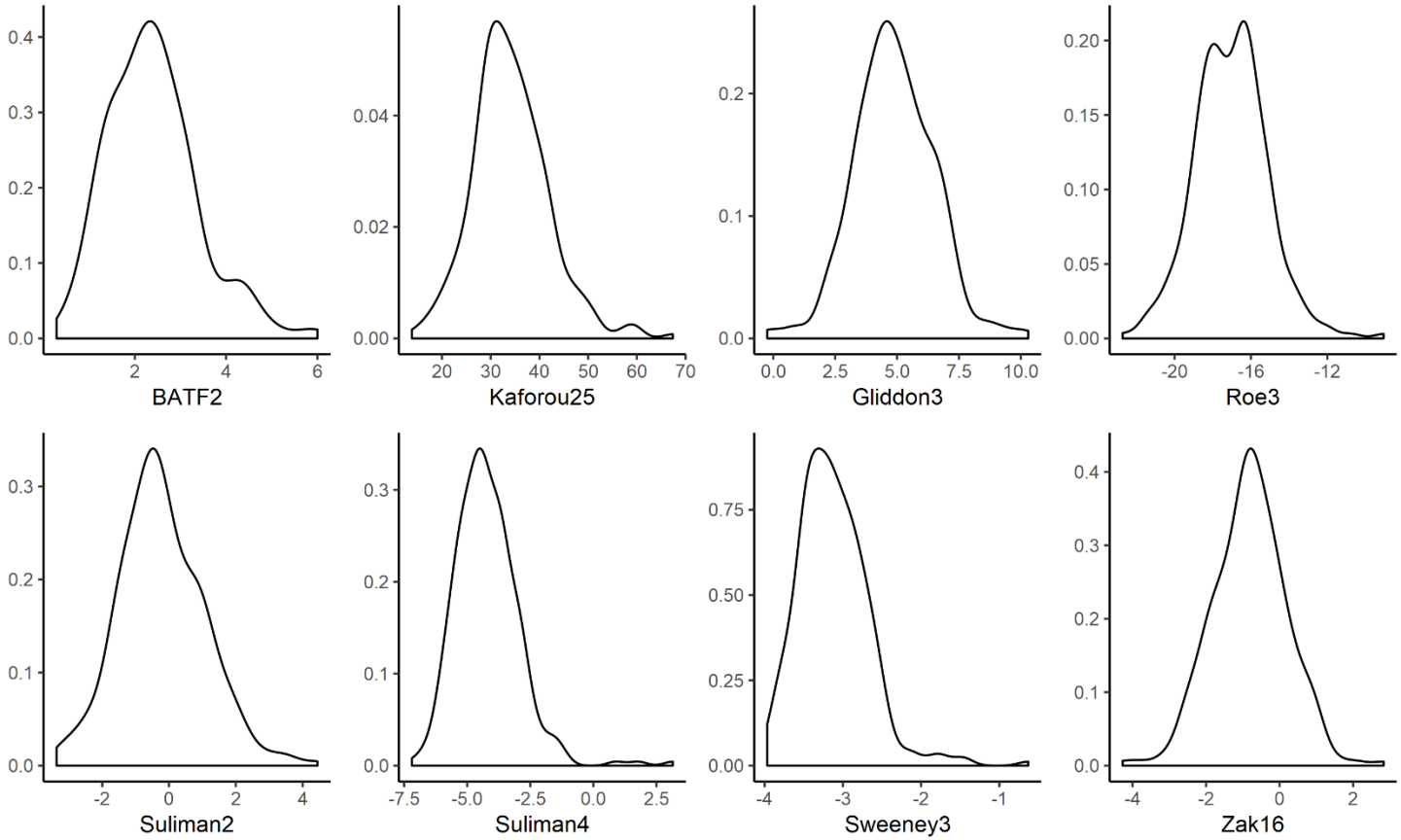


Table 8-8: Number of incipient tuberculosis and non-progressor samples included in analysis.

Stratified by (a) primary analysis; (b) sensitivity analysis including only TB cases with microbiological confirmation; (c) sensitivity analysis including only one sample per participant; (d) sensitivity analysis with mutually exclusive time intervals to TB. Data presented as n (% of all available samples per time interval). NP = non-progressor.

Timeframe	**Signature**	**0 to 24 months**		**0 to 12 months**		**0 to 6 months**		**0 to 3 months**	
		NP	*Incipient TB*	*NP*	*Incipient TB*	*NP*	*Incipient TB*	*NP*	*Incipient TB*
(a) Primary analysis	*Zak16*	489 (87.5)	119 (66.9)	723 (85.1)	80 (68.4)	777 (82.4)	49 (80.3)	777 (82.4)	12 (57.1)
	Suliman4	444 (79.4)	122 (68.5)	706 (83.1)	81 (69.2)	771 (81.8)	34 (55.7)	771 (81.8)	17 (81.0)
	Suliman2	351 (62.8)	81 (45.5)	584 (68.7)	53 (45.3)	629 (66.7)	20 (32.8)	629 (66.7)	15 (71.4)
	All other signatures	559 (100.0)	178 (100.0)	850 (100.0)	117 (100.0)	943 (100.0)	61 (100.0)	943 (100.0)	21 (100.0)
(b) Microbiological confirmation	*Zak16*	489 (87.5)	106 (64.2)	723 (85.1)	70 (65.4)	777 (82.4)	44 (78.6)	777 (82.4)	11 (55.0)
	Suliman4	444 (79.4)	111 (67.3)	706 (83.1)	73 (68.2)	771 (81.8)	30 (53.6)	771 (81.8)	16 (80.0)
	Suliman2	351 (62.8)	76 (46.1)	584 (68.7)	49 (45.8)	629 (66.7)	18 (32.1)	629 (66.7)	14 (70.0)
	All other signatures	559 (100.0)	165 (100.0)	850 (100.0)	107 (100.0)	943 (100.0)	56 (100.0)	943 (100.0)	20 (100.0)
(c) One sample per participant	*Zak16*	463 (91.3)	94 (74.6)	676 (91.8)	64 (73.6)	704 (90.5)	41 (85.4)	704 (90.5)	10 (66.7)
	Suliman4	406 (80.1)	81 (64.3)	609 (82.7)	58 (66.7)	638 (82.0)	24 (50.0)	638 (82.0)	11 (73.3)
	Suliman2	319 (62.9)	51 (40.5)	504 (68.5)	37 (42.5)	519 (66.7)	14 (29.2)	519 (66.7)	10 (66.7)
	All other signatures	507 (100.0)	126 (100.0)	736 (100.0)	87 (100.0)	778 (100.0)	48 (100.0)	778 (100.0)	15 (100.0)
(d) Mutually exclusive time periods		**12 to 24 months**		**6 to 12 months**		**3 to 6 months**		**0 to 3 months**	
	Zak16	489 (87.5)	39 (63.9)	723 (85.1)	31 (55.4)	777 (82.4)	37 (92.5)	777 (82.4)	12 (57.1)
	Suliman4	444 (79.4)	41 (67.2)	706 (83.1)	47 (83.9)	771 (81.8)	17 (42.5)	771 (81.8)	17 (81.0)
	Suliman2	351 (62.8)	28 (45.9)	584 (68.7)	33 (58.9)	629 (66.7)	5 (12.5)	629 (66.7)	15 (71.4)
	All other signatures	559 (100.0)	61 (100.0)	850 (100.0)	56 (100.0)	943 (100.0)	40 (100.0)	943 (100.0)	21 (100.0)