Diagnostic accuracy of low and moderate complexity assays for pulmonary and extrapulmonary tuberculosis detection: systematic review and meta-analyses

Mikashmi Kohli

Department of Epidemiology, Biostatistics and Occupational Health
Faculty of Medicine and Health Sciences
McGill University
January 2023

A thesis to be submitted to McGill University partial fulfilment of the requirements of the degree of
Master of Science in Epidemiology
Montreal, Quebec, Canada

TABLE OF CONTENTS

Contents

FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

FRONT MATTER

Abstract (English)

Background

The WHO estimates that globally in 2021, 10.6 million people fell ill with tuberculosis (TB). Nearly 4.2 million people were either not diagnosed or not reported during 2021, and this makes TB diagnosis the biggest challenge. With the emergence of molecular rapid tests for TB, there is some hope for improving case detection. Meta-analyses offer an efficient way of summarizing test accuracy for guideline development. We performed two systematic reviews on two technology classes - low and moderate complexity assays - for diagnosing pulmonary and extrapulmonary TB, respectively.

Methodology

A comprehensive search of six databases for relevant citations was performed. Cross-sectional, case-control, cohort studies, and randomised controlled trials of any of the index tests were included. Data on Xpert MTB/RIF and Ultra (low complexity assay) were extracted for extrapulmonary TB. Additionally, for moderate complexity assays for diagnosing pulmonary TB, five technologies were included. Study quality was assessed using Quadas instrument, and pooled test accuracy estimates were generated via bivariate random effects regression. Latent class analysis was used to estimate the accuracy of Xpert MTB/RIF and Ultra as culture is not a perfect reference standard for diagnosing extrapulmonary TB.

Results

Low complexity assays for extrapulmonary TB (based on site of the disease)

For this meta-analysis, we included 69 studies: 58 evaluated Xpert MTB/RIF, 2 evaluated Xpert Ultra, and 9 evaluated both tests. Most studies were conducted in China, India, South Africa, and Uganda. Overall, risk of bias was low for patient selection, index test, and flow and timing domains. For reference standard, 49% had low and 43% had unclear risk of bias. Applicability for the patient selection domain was unclear for most studies because clinical settings were uncertain. For cerebrospinal fluid, Ultra pooled sensitivity and specificity (95% CrI) against culture were 89.4% (79.1 to 95.6) (89 participants) and 91.2% (83.2 to 95.7) (386 participants). Xpert pooled sensitivity and specificity against culture were 71.1% (62.8 to 79.1) (571 participants) and 96.9% (95.4 to 98.0) (2824 participants). For pleural fluid, Ultra pooled sensitivity and specificity against culture were 75.0% (58.0 to 86.4) (158 participants) and 87.0% (63.1 to 97.9) (240 participants).

Xpert pooled sensitivity and specificity against culture were 49.5% (39.8 to 59.9) (644 participants) and 98.9% (97.6 to 99.7) (2421 participants). For Lymph node aspirate, Ultra sensitivity and specificity (95% CI) against composite reference standard were 70% (51 to 85) (30 participants) and 100% (92 to 100) (43 participants). Xpert pooled sensitivity and specificity against composite reference standard were 81.6% (61.9 to 93.3) (377 participants) and 96.4% (91.3 to 98.6) (302 participants). For lymph node aspirates, Xpert MTB/RIF pooled specificity against culture was 86.2% (78.0 to 92.3. Using the latent class model, Xpert pooled specificity was 99.5% (99.1 to 99.7), similar to that observed with a composite reference standard.

Moderate complexity assays for pulmonary TB

A total of 21 studies were included. We could only meta-analyse data for three of the five assays identified, as data were limited for the remaining two. For TB detection, the included assays had a sensitivity of 91% or more and the specificity ranged from 97% to 100%. For rifampicin resistance detection, all the included assays had a sensitivity of more than 92%, with a specificity of 99% to 100%. Sensitivity for isoniazid resistance detection varied from 70% to 91%, with higher specificity of 99% to 100% across all index tests. Studies that included head-to-head comparisons of these assays with Xpert MTB/RIF for detection of TB and rifampicin resistance suggested comparable diagnostic accuracy.

Conclusions

Xpert Ultra and Xpert MTB/RIF may be helpful in diagnosing extrapulmonary tuberculosis. While sensitivity varies across different extrapulmonary specimens, for most specimen types, specificity is high, which is helpful in ruling in people for confirmed disease. Xpert Ultra and Xpert MTB/RIF had similar sensitivity and specificity for rifampicin resistance. In people with symptoms of pulmonary TB, the moderate complexity assays demonstrate comparable diagnostic accuracy for detection of TB, rifampicin and isoniazid resistance to Xpert MTB/RIF assay, a WHO recommended molecular test.

Résumé (Français)

Contexte

L'OMS estime qu'en 2021, dans le monde, 10,6 millions de personnes ont contracté la tuberculose (TB). Près de 4,2 millions de personnes n'ont pas été diagnostiquées ou n'ont pas été signalées en 2021, ce qui fait du diagnostic de la tuberculose le plus grand défi. Avec l'émergence des tests moléculaires rapides pour la TB, on peut espérer améliorer la détection des cas. Les méta-analyses offrent un moyen efficace de résumer l'exactitude des tests pour l'élaboration de lignes directrices. Nous avons effectué deux revues systématiques sur deux classes de technologies - les tests de complexité faible et modérée - pour le diagnostic de la TB pulmonaire et extrapulmonaire, respectivement.

Méthodologie

Une recherche exhaustive de six bases de données pour les citations pertinentes a été effectuée. Des études transversales, des cas-témoins, des études de cohorte et des essais contrôlés randomisés de l'un ou l'autre des tests index ont été inclus. Les données sur Xpert MTB/RIF et Ultra (test de faible complexité) ont été extraites pour la TB extrapulmonaire. De plus, pour les tests de complexité modérée pour le diagnostic de la TB pulmonaire, cinq technologies ont été incluses. La qualité de l'étude a été évaluée à l'aide de l'instrument Quadas, et les estimations de précision des tests regroupés ont été générés via une régression à effets aléatoires bivariés. L'analyse de classe latente a été utilisée pour estimer la précision de Xpert MTB/RIF et Ultra car la culture n'est pas une norme de référence parfaite pour le diagnostic de la TB extrapulmonaire.

Résultats

Tests de faible complexité pour la TB extrapulmonaire (basés sur le site de la maladie)

Pour cette méta-analyse, nous avons inclus 69 études: 58 évaluaient Xpert MTB/RIF, 2 évaluaient Xpert Ultra, et 9 évaluaient les deux tests. La plupart des études ont été menées en Chine, en Inde, en Afrique du Sud et en Ouganda. Dans l'ensemble, le risque de biais était faible pour la sélection des patients, le test d'index et les domaines de flux et de synchronisation. Pour le standard de référence, 49% avaient un risque faible et 43% un risque incertain. L'applicabilité au domaine de

sélection des patients n'était pas claire pour la plupart des études car les paramètres cliniques étaient incertains.

Pour le liquide céphalo-rachidien, la sensibilité et la spécificité Ultra regroupées (ICr à 95 %) par rapport à la culture étaient de 89,4 % (79,1 à 95,6) (89 participants) et de 91,2 % (83,2 à 95,7) (386 participants). La sensibilité et la spécificité regroupées de Xpert par rapport à la culture étaient de 71,1 % (62,8 à 79,1) (571 participants) et de 96,9 % (95,4 à 98,0) (2 824 participants). Pour le liquide pleural, la sensibilité et la spécificité regroupées Ultra contre la culture étaient de 75,0% (58,0 à 86,4) (158 participants) et de 87,0% (63,1 à 97,9) (240 participants). La sensibilité et la spécificité regroupées de Xpert par rapport à la culture étaient de 49,5% (39,8 à 59,9) (644 participants) et de 98,9% (97,6 à 99,7) (2 421 participants). Pour l'aspiration des ganglions lymphatiques, la sensibilité et la spécificité Ultra (IC à 95%) par rapport à la norme de référence composite étaient de 70% (51 à 85) (30 participants) et de 100% (92 à 100) (43 participants). La sensibilité et la spécificité regroupées de Xpert par rapport à la norme de référence composite étaient de 81,6% (61,9 à 93,3) (377 participants) et de 96,4% (91,3 à 98,6) (302 participants). Pour l'aspiration des ganglions lymphatiques, la spécificité regroupée de Xpert MTB/RIF par rapport à la culture était de 86,2% (78,0 à 92,3). En utilisant le modèle de classe latente, la spécificité regroupée de Xpert était de 99,5% (99,1 à 99,7), similaire à celle observée avec une norme de référence composite.

Tests de complexité modérée pour la TB pulmonaire

Au total, 21 études ont été incluses. Nous n'avons pu effectuer une méta-analyse des données que pour trois des cinq tests identifiés, car les données étaient limitées pour les deux autres. Pour la détection de la TB, les tests inclus avaient une sensibilité de 91% ou plus et la spécificité variait de 97% à 100%. Pour la détection de la résistance à la rifampicine, tous les tests inclus avaient une sensibilité de plus de 92%, avec une spécificité de 99% à 100%. La sensibilité pour la détection de la résistance à l'isoniazide variait de 70% à 91%, avec une spécificité plus élevée de 99% à 100% pour tous les tests d'indice. Les études qui comprenaient des comparaisons directes de ces tests avec Xpert MTB/RIF pour la détection de la TB et de la résistance à la rifampicine ont suggéré une précision diagnostique comparable.

Conclusions

Xpert Ultra et Xpert MTB/RIF peuvent être utiles pour diagnostiquer la TB extrapulmonaire. Bien que la sensibilité varie selon les différents échantillons extrapulmonaires, pour la plupart des types d'échantillons, la spécificité est élevée, ce qui est utile pour statuer chez les personnes atteintes d'une maladie confirmée. Xpert Ultra et Xpert MTB/RIF avaient une sensibilité et une spécificité similaires pour la résistance à la rifampicine. Chez les personnes présentant des symptômes de TB pulmonaire, les tests de complexité modérée démontrent une précision diagnostique comparable pour la détection de la TB, la résistance à la rifampicine et la résistance à l'isoniazide à celle de Xpert MTB/RIF, un test moléculaire recommandé par l'OMS.

Acknowledgements

Parents are the most important and precious gift from God. I feel blessed to have such supporting parents who have rendered love, support and their blessings to me at every step of my life. I humbly express my obeisance to the Lord Almighty, who has bestowed me health and courage to fulfil this hard task to the best of my capabilities.

I would like to express my profound gratitude to Dr. Madhukar Pai for his supervision, advice, guidance and unwavering support throughout my master's experience. Deciding to do my post-doctoral fellowship and master's thesis under his supervision, was probably the best decision of my academic life as he not only supported and guided me throughout but also has always pushed all his students to be better and aim for higher academic goals while keeping a healthy work life balance. I have learnt from him that we can do rigorous research and work hard while being compassionate. He has always inspired me to do better and achieve higher towards my next milestone in my professional career. A strong believer in making sure that all his students reach their highest potential. He constantly asks all his students to go back to the reason why we chose this career path and work towards a bigger cause and to never give up. I could not have imagined a better mentor for my MSc thesis. Without his constant support and supervision, this work could not have been possible. Thanking him for all his efforts would be an understatement.

I gratefully acknowledge my co-supervisors, Nandini Dendukuri. I learnt a lot about analyses and developing concrete plans through our discussions. She was always available to help me with any question I had about my thesis analysis, and she helped me think critically about developing innovative methods of analyses for systematic reviews. I aspire to be able to replicate her critical thinking, enthusiasm for research and maintain her poise and calm mind.

I must also thank my committee member, Dr. Samuel Schumacher. I am very grateful to him for always making the time to mentor and support me. He is an excellent role model and taught me a lot about thinking through the problem critically and seeing the bigger picture and relevance of my work in public health. I value his insightful feedback and the time he took out to answer my questions throughout my thesis.

I would like to thank Dr. Karen Steingart for being an exceptional mentor and taking the time to guide me through the rigorous process of performing systematic review using Cochrane methods and resources. I am so thankful for her incredible support, kindness, and patience.

I thank almighty to usher all his love on me by giving me two loving brothers, Sahil Kohli and Karan Kohli who have stood by my side no matter how hard the situations were and how impatient I got during the tenure of my thesis. I thank them for bearing all my tantrums and mood swings when it got tough. They have been my strongest support system and the driving force to keep going and never let me give up.

I might fall short of words to express my gratitude towards my grandmother, Lt. Smt. Urmila Kohli, who never gave in to what the world said and encouraged me to work harder and aim higher. She never let anyone or any old- school belief get into the way of my work and ambition.

Many thanks also to past and present members of Team Pai! Working with them has been a true privilege and they impress and amuse me endlessly. Gratitude is particularly owed to Caroline Vadnais for keeping everything running and making all things doable. Thank you to my fellow students and cohort colleagues. Their generosity, care, and reality checks have helped me through all the most challenging parts of the program.

No words would suffice to express my thanks to all those patients and volunteers for participating in the included studies, without them this work would not have been possible.

Financial support

I am extremely grateful for the financial support I received. I held the Becklake Fellowship from RI MUHC for the first year of my MSc.

Preface and contribution of authors

As first author on all manuscripts included in this thesis, I developed the research questions and protocols for all chapters, with feedback and support from my co-supervisors, Dr Madhukar Pai, Dr Nandini Dendukuri, committee member Dr Samuel Schumacher, and manuscript co-authors. All analyses were conducted by me. I was responsible for interpretation of the results, drafting of the manuscripts, and subsequent submissions and revisions for all published and submitted papers. All manuscripts included in this thesis were written by me. Detailed author contributions for each manuscript are described below.

Manuscript 1: Xpert MTB/RIF Ultra and Xpert MTB/RIF assays for extrapulmonary tuberculosis and rifampicin resistance in adults. This is a Cochrane systematic revie and meta-analyses. I wrote the protocol with inputs and support from Dr. Karen Steingart who is a senior editor of the Cochrane infectious Diseases Diagnostic. I designed the data extraction form, QUADAS-2 rules and statistical analyses plan. Dr. Claudia Denkinger, Samuel Schumacher and Dr. Karen Steingart, provided inputs on the methodological assessment of the included studies. All authors provided critical feedback.

Manuscript 2: Diagnostic accuracy of centralised assays for TB detection and detection of resistance to rifampicin and isoniazid: a systematic review and meta-analysis. I wrote the protocol and analyses plan for this review. Dr. Madhukar Pai, Dr. Claudia Denkinger and Dr. Samuel Schumacher provided inputs on methodological and statistical analyses for the study. Dr. Emily MacLean was involved in reading and assessing the studies included in this review. She helped with the data extraction of the included studies and supported the data analyses plan. All authors provided critical feedback.

Both of these manuscripts provided evidence for policy recommendations on diagnostic accuracy of molecular tests for pulmonary and extrapulmonary TB detection.

Abbreviations

ATT – anti-tuberculosis therapy

BD – Becton Dickinson

CI – confidence interval

CrI – credible interval

COVID-19 – Coronavirus disease 2019

CRS – composite reference standard

CSF – cerebrospinal fluid

CXR – chest x-ray

DR-TB – drug resistant TB

DST – drug susceptibility testing

EPTB – extrapulmonary tuberculosis

HIV- Human immunodeficiency virus

INH – isoniazid

LAM – lipoarabinomannan

LCA – latent class analysis

LPA – line probe assay

LTBI – latent TB infection

MDR-TB – multi drug resistant tuberculosis

MRS- microbiological reference standard

NA – not applicable

NAAT – nucleic acid amplification test

NE – not estimable

NGS – next generation sequencing

No. – number of

NTP – National TB Program

POC – point-of-care

PTB – pulmonary TB

PLHIV- People living with HIV

QUADAS-2 – The Quality Assessment of Diagnostic Accuracy Studies-2

RIF – rifampicin

SLID – second line injectable drug

Smear – smear microscopy

TB – tuberculosis

TPP – target product profile

Ultra – Xpert MTB/RIF Ultra

UNHLM – UN High level meeting

WGS – whole genome sequencing

WHO – World Health Organization

XDR-TB – extensively drug resistant tuberculosis

Xpert – Xpert MTB/RIF

CHAPTER 1: INTRODUCTION

Tuberculosis, an infection caused by *Mycobacterium tuberculosis* (Mtb) is world's second leading cause (after Covid-19) of infectious disease related deaths and is one of the top 10 causes of death worldwide. According to the WHO, in 2021, an estimated 10.6 million people fell ill with TB. The incidence rate rose by 3.6% between 2020 and 2021, reversing declines of about 2% per year for most of the past 2 decades[1]. Nearly 1.6 million people died from TB, making TB the second leading infectious killer after Covid-19.

Due to pandemic disruptions of essential health services in both 2020 and 2021, there have been huge and sustained drops in the number of people newly detected with TB. From a peak of 7.1 million in 2019, the number of people with newly diagnosed disease fell to 5.8 million in 2020, back to the level last seen in 2012. In 2021, there was a partial recovery, to 6.4 million, still lower than the pre-pandemic levels. This means, nearly 4.2 million people with TB were either not diagnosed, or not reported to National TB Programs during 2021[1]. This, in turn, meant that people with undiagnosed TB were transmitting TB infection to people in their families and communities. Early, rapid, and accurate detection of TB is essential to achieve global control of the disease. If the patient gets timely and appropriate treatment, TB disease is largely curable.

Over the last decade, the field of TB diagnostics has seen advances in the form of new molecular tests. Often referred to as nucleic acid amplification tests (NAATs), these assays rely on amplification of a targeted genetic region of the *Mycobacterium tuberculosis* complex, typically by PCR. NAATs can detect TB and perform drug susceptibility testing (DST) for key drugs, such as rifampin (RIF) and isoniazid (INH), more quickly than conventional mycobacterial culture and are also available at different levels of health care systems. Ideally, rapid and point-of-care (POC) testing would then help administering appropriate treatment to the patient without being lost to follow up in the care cascade. As such, NAATs are disrupting the field of TB diagnostics and are helping to improve the quality of TB care[2,3].

TB predominantly affects the lungs (pulmonary TB). Extrapulmonary TB refers to TB in parts of the body other than the lungs, and is known to affect virtually every part, with lymph nodes and pleura being the most common sites[4]. While active pulmonary TB is transmissible by aerosolized

droplets spread by coughing individuals, extrapulmonary TB is thought to result from hematogenous spread from an initial lung infection, and is typically not infectious. Extrapulmonary TB can occur alone or together with pulmonary TB.

Diagnosis of extrapulmonary TB is challenging for several reasons. Many forms of extrapulmonary TB require invasive diagnostic sampling and collecting adequate specimens can pose a risk of harm to the patient and be costly. Most forms of extrapulmonary TB are paucibacillary (TB disease caused by a smaller number of bacteria), making diagnosis by the conventional method of smear microscopy less sensitive. Molecular tests could help address this gap.

Xpert MTB/RIF(Xpert) and Xpert Ultra (Ultra) is one such molecular TB diagnostic test that helps identifies Mtb and rifampicin resistance in the specimen. It is a semiquantitative test that provides results in less than 90 minutes which was a major game changer in the field of TB when it was first introduced in 2010. The WHO published updated guidance on the use of Xpert® MTB/RIF in 2013[5]. Currently the manufacturer, Cepheid Incorporated (Sunnyvale, CA, USA), has made no specific recommendations for the use of Xpert® MTB/RIF in non-sputum specimens, and accordingly, Xpert® MTB/RIF is approved by the United States Food and Drug Administration (FDA) for use in raw sputum specimens and concentrated sputum sediment only[6]. I performed a systematic review on accuracy estimates of Xpert and Ultra for diagnosing extrapulmonary form of TB.

Recently, several companies have developed molecular tests for tuberculosis and RIF/INH resistance detection on centralised, high throughput platforms, many of which have already been established as multi-disease platforms, primarily for detection of HIV, human papillomavirus, SARS-CoV2 and hepatitis C virus. I performed another systematic review intended to evaluate the diagnostic accuracy of five of these tests for M. tuberculosis and RIF/INH resistance detection to assess their diagnostic accuracy for pulmonary TB. The tests included were Abbott RealTime MTB, Abbott RealTime RIF/INH, FluoroType MTB, FluoroType MTDBR , BD Max MDR-TB, COBAS MTB assay.

Both these systematic reviews have been published and were used by WHO for guideline development on use of molecular nucleic acid amplification tests (NAATs) for TB detection.

CHAPTER 2: LITERATURE REVIEW

2.1 Tuberculosis etiology and epidemiology

Tuberculosis is an infectious airborne disease that is caused by *Mycobacterium tuberculosis (Mtb)* and its genetically related species known as the mycobacterium tuberculosis complex (MTC). Approximately 86% of *Mtb* infections manifest as a pulmonary disease, however, the bacterial pathogen can cause disease anywhere in the body (except nails and hair in the humans), referred to as extrapulmonary TB [1] Only active PTB is considered contagious. Once in the lung, *Mtb* can spread through the air via actions such as coughing, sneezing and talking. *M. tuberculosis* can exist in the body in different forms – an asymptomatic and non-transmissible state known as latent TB infection (LTBI) or as active TB disease[7]. On average, 5-10% of individuals infected with *M. tuberculosis* will develop active TB disease. Newly infected individuals are at the highest risk for developing active TB. Two years after infection, development of active TB disease occurs infrequently.

Patients with active TB disease experience symptoms such as fever, loss of appetite, fatigue, cough, hemoptysis and in case of EPTB, the patients might also experience enlarged lymph nodes, headaches etc. based on the site of the body infected. Many studies have also shown that with incomplete treatment or active untreated disease, the pulmonary function of the lungs decrease and there is significant lung remodeling that takes place in these patients.

Currently, COVID-19 is the leading cause of death, globally. However, in 2019, 10 million people fell ill from TB and 1.6 million people died from the disease, and TB was the leading infectious cause of death before the pandemic. The highest burden of TB is consistently found among the most vulnerable and marginalized groups –prisoners, the homeless, injection drug users, minorities, migrants and refugees and people living in poverty[8]. Risk of exposure to *Mtb* is related to the underlying disease burden and the environmental conditions where people work and live, including ventilation, crowding, and air pollution and quality. Among those infected, risk of developing active disease is increased by other variables such as HIV infection, malnutrition, smoking, alcohol abuse and diabetes[8,9]

In 2019, approximately 44% of all the TB cases were from South-East Asia, followed by 25% and 18% in the Western Pacific region. There were 8 major countries that accounted for the two-third of the global TB burden with India accounting for 26% of these cases. Recently, WHO published an updated list of 30 high TB burden countries that accounted for 21% of the global burden[1]. TB incidence is very varied across the globe, ranging from 5 cases per 100,000 to over 500 cases per 100,000.

During the Covid-19 pandemic, essential health services were seriously disrupted, with devastating consequences for TB care. TB notifications dropped in most countries. According to the WHO, in 2021, an estimated 10.6 million people fell ill with TB. The incidence rate rose by 3.6% between 2020 and 2021, reversing declines of about 2% per year for most of the past 2 decades. Nearly 1.6 million people died from TB, making TB the second leading infectious killer after Covid-19. TB is the leading cause of death among people living with HIV. On the flip side, it is also worth mentioning that possibly due to the global policy and stringent rules on use of masks, it is possible that there was a decline in the TB disease transmission as well.

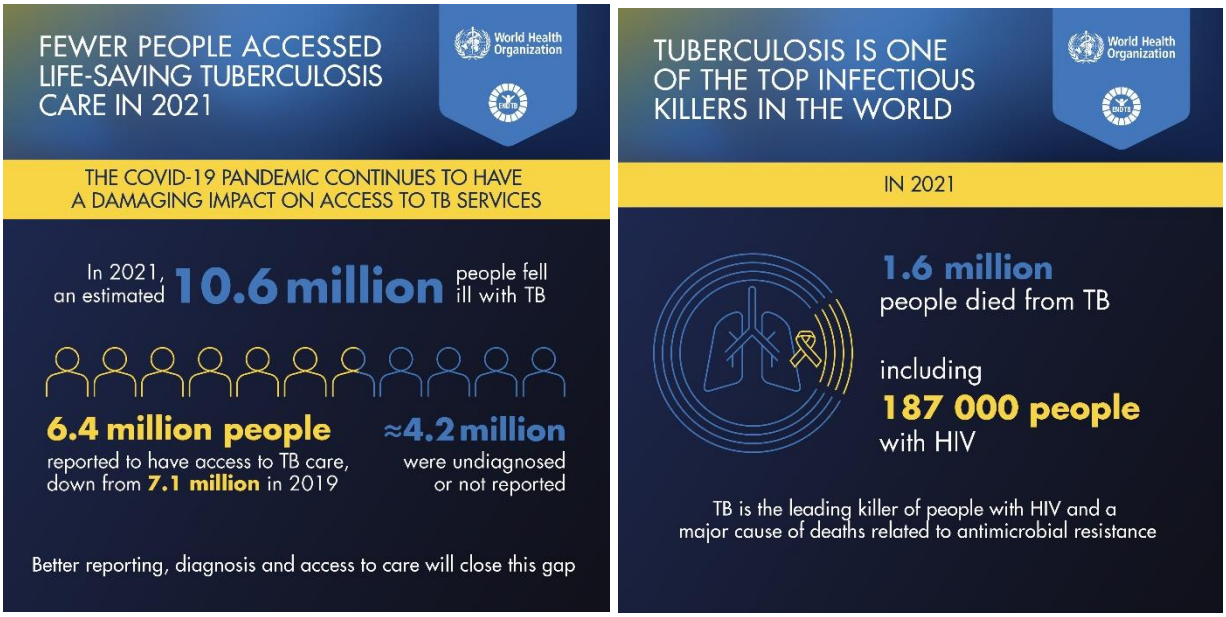


Figure 2.1: Global TB burden

Source: WHO, Global TB Report 2022

As shown in Figure 2.1 above, of the 10.6 million people who fell ill with TB during 2021, only 6.4 million were diagnosed and reported to national TB programs. This left a gap of nearly 4.2 million people who were either not diagnosed, or not notified to TB programs. Therefore, improving TB case detection is an urgent priority for TB control. Box 1 captures the key facts relevant to TB epidemiology, from the most recent WHO Global TB Report 2022.

Box 1. Key facts about global TB epidemiology (WHO, Global TB Report 2022)

TB BURDEN

- In 2021, an estimated 10.6 million (95% confidence interval 9.9-11 million) people fell ill with TB worldwide, of which 6.0 million were men, 3.4 million were women and 1.2 million were children. People living with HIV accounted for 6.7% of the total.
- The TB incidence rate (new cases per 100 000 population per year) rose by 3.6% between 2020 and 2021, reversing declines of about 2% per year for most of the past 2 decades.
- Globally, the estimated number of deaths from TB increased between 2019 and 2021, reversing years of decline between 2005 and 2019. In 2021, 1.6 million people died from TB, including 187 000 people with HIV.
- Eight countries accounted for more than two thirds of the global total: India, Indonesia, China, the Philippines, Pakistan, Nigeria, Bangladesh and the Democratic Republic of the Congo.

TB CARE AND PREVENTION

- TB treatment saved 74 million lives globally between 2000 and 2021.
- Globally, the number of people newly diagnosed with TB and those reported to national governments fell from 7.1 million in 2019 to 5.8 million in 2020. There was a partial recovery to 6.4 million in 2021.
- The cumulative number of people treated between 2018 and 2021 was 26.3 million, equivalent to 66% of the 5-year (2018–2022) UN High Level Meeting TB target of 40 million. This included 1.9 million children, 54% of the 5-year target of 3.5 million.
- There is still a large global gap between the estimated number of people who fell ill with TB and the number of people newly diagnosed, with 4.2 million people not diagnosed with the disease, or not officially reported to national authorities in 2021, up from 3.2 million in 2019.

DRUG-RESISTANT TB

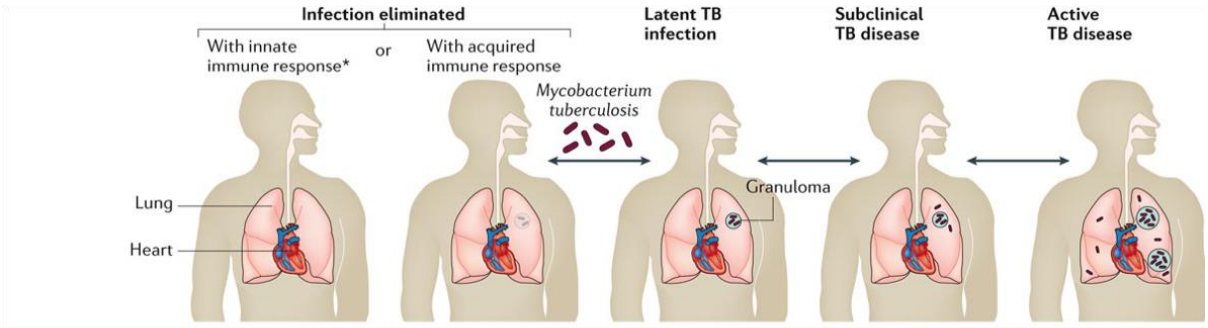
- The burden of drug-resistant TB (DR-TB) is also estimated to have increased between 2020 and 2021, with 450 000 (95%CI: 399 000–501 000) new cases of rifampicin-resistant TB (RR-TB) in 2021.
- The number of people provided with treatment for RR-TB and multidrug-resistant TB (MDR-TB) declined between 2019 and 2020. The reported number of people started on treatment for RR-TB and MDR-TB in 2021 was 161 746, covering only about one in three of those in need.
- The treatment success rate for drug-resistant TB, at 60% globally, remains low.

2.2 TB pathology

Exposure to *Mtb* can lead to natural elimination of the bacilli from the body where the innate and/or adaptive immune responses are at play. However, if the bacilli are not eliminated from the body, they could be present in quiescent stage in the body. This is known as latent TB infection (LTBI). At this stage, live bacilli are present in the human body, however, the patient is asymptomatic. This individual would probably benefit from LTBI treatment (also called preventive therapy).

Following the inhalation of aerosols containing *Mtb*, the bacteria move to the lower respiratory tract and alveolar macrophages phagocytose these bacilli. *Mtb* blocks the formation of phagolysosome, hence ensuring its survival in the body. Once, the bacilli have infected the alveolar macrophages in the respiratory tract, it then can invade the lung interstitial tissue where the subsequent immune response leads to the formation of a granuloma. If the host immune response is unable to contain the infection within granulomas, as is commonly the case in immunocompromised individuals, bacteria disseminate to the blood or re-enter the respiratory tract to be released. This state, where the individual starts to manifest symptoms (and could become contagious), is known as active TB disease[7,10,11].

Multiple stages of PTB can exist simultaneously in one lung and across patients there is no single, consistent presentation of PTB in lungs. Various factors such as host and bacterial genetics, host immune status, nutritional status and the presence of co-infections (HIV, diabetes, other auto-immune diseases) all contribute to the diversity of pathologies, resulting in the 'spectrum of TB' (Figure 2.2)[10].



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

Nature Reviews | Disease Primers

Figure 2.2. The spectrum of tuberculosis

(Source: Pai et al, 2016).
TB: tuberculosis; TST: tuberculin skin test; IGRA: interferon gamma release assay.

2.3 Extrapulmonary TB

The various forms of extrapulmonary TB cause signs and symptoms related to the structures affected. Table 2.1 describes the forms of extrapulmonary TB, as well as the different specimens that may be acquired for diagnosis. The presentation of extrapulmonary TB varies depending on the body site affected and may imitate other diseases, such as cancer and bacterial and fungal infections. The signs and symptoms of extrapulmonary TB are often non-specific and may include fever, night sweats, fatigue, loss of appetite, and weight loss (as seen in pulmonary TB) or specific complaints related to the involved site (for example, headache for TB meningitis and back pain for TB of the spine). The clinical presentation of extrapulmonary disease may be acute, but is more often subacute (falling between acute and chronic) or chronic, meaning that patients may have symptoms for days to months before they seek care. Signs and symptoms for the forms of extrapulmonary TB are described in Table 2.

Table 2.1. Forms of extrapulmonary TB

Form of extrapulmonary TB	Characteristics	Diagnostic specimens and means of collection
Lymph node TB, also called TB lymphadenitis	MTB infection of lymph nodes. May affect one node or a group of nodes, or multiple groups within a chain. Lymph node TB is relatively more common among children than adults. The most common presentation is of a single, firm, non-tender enlarged node in the neck, although any lymph node group can be affected. This may be accompanied by fever, weight loss, and night sweats, particularly in people with HIV. Patients with TB in deep lymph nodes, such as the mediastinal or mesenteric lymph nodes, may present with fever, night sweats, and weight loss, or more rarely with symptoms related to compression of adjacent structures. Over time lymph nodes become fluctuant and may discharge via a sinus to the skin or an adjacent viscus. It should be noted that lymphadenopathy may also be seen in other forms of TB as part of the immune response, but this is not usually caused by direct infection of the lymph nodes.	Fine needle aspirate from affected lymph node, with or without radiological guidance; excisional biopsy of superficial lymph nodes; endoscopic biopsy of deep lymph nodes with ultrasound guidance
Pleural TB, also called TB pleurisy	MTB infection of the pleura presents with the gradual onset of pleuritic chest pain, shortness of breath, fever, night sweats, and weight loss. Chest X-ray may demonstrate unilateral or occasionally bilateral pleural effusion. The severity of symptoms is highly variable, with many patients experiencing spontaneous resolution of symptoms, while others may develop severe pleural effusions requiring drainage. Pleuro-pulmonary TB, where	Pleural aspirate; pleural biopsy, which may be performed via thoracoscopy or percutaneously with an Abram's needle, with or without ultrasound guidance

	there is parenchymal lung involvement visible on a chest X-ray, is associated with higher mortality than isolated pleural infection, which appears to be rarely fatal.	
TB meningitis, also called tuberculous meningitis	MTB infection of the meninges affects people of all ages, but is most common in children and people with untreated HIV infection. In adults, TB meningitis presents with the gradual onset of headache, neck stiffness, malaise, fever, and, if untreated, can progress to altered sensorium, focal neurological deficits, coma, and death. Young children may present with poor weight gain, low-grade fever, and listlessness. Infants, may present with fever, cough (related to the primary pulmonary infection which occurs before TB meningitis develops), change of consciousness at presentation, bulging anterior fontanel, and seizures. TB meningitis is sometimes associated with a concurrent cerebral tuberculoma, or more rarely a tuberculous abscess.	CSF, acquired by lumbar puncture with or without radiological guidance; biopsy of tuberculoma, acquired surgically
Bone and joint TB	MTB infection of the bones or joints or both causes chronic pain, deformity and disability, and TB of the cervical spine can be life-threatening. The usual presenting symptom is pain. Fever and weight loss, with or without signs of spinal cord compression, may be present. Patients with advanced disease may have severe pain, spinal deformity, paraspinal muscle wasting, and neurological deficit. Children may have failure to thrive and difficulty walking.	Aspiration of joint fluid or periarticular abscesses; percutaneous computed tomography guided biopsy of lesions is preferred, but some patients may require open biopsy
Genitourinary TB	MTB infection of the genitourinary tract. Renal TB presents with flank pain, haematuria, and dysuria. Female genital TB presents with	Urine; biopsy of affected organs, acquired under radiological guidance or surgically

	infertility (and may be otherwise asymptomatic), pelvic pain, and vaginal bleeding. Testicular TB presents with a scrotal mass and infertility.	
Pericardial TB, also called TB pericarditis	MTB infection of the pericardium presents with fever, malaise, night sweats, and weight loss. Chest pain and shortness of breath are also commonly experienced symptoms. Pericardial TB may be associated with pericardial effusion, which can be severe and lead to tamponade, which is life-threatening. Some patients go on to develop pericardial constriction, which can lead to heart failure and death, and may require surgical intervention even after mycobacterial cure.	Pericardial fluid acquired by pericardiocentesis; pericardial biopsy, acquired under radiological guidance or surgically
Peritoneal TB	MTB infection of the peritoneum. Peritoneal TB usually presents with pain and abdominal swelling, which may be accompanied by fever, weight loss, and anorexia.	Ascitic fluid acquired by paracentesis; peritoneal biopsy
Disseminated TB, also called miliary TB	Disseminated TB refers to TB that involves two or more distinctly separate sties. Manifestations may be quite varied, ranging from acute fulminant disease to non-specific symptoms of fever, weight loss, and weakness. HIV-positive people are more likely to have disseminated TB than HIV-negative people. In a systematic review of the prevalence of TB in post-mortem evaluations of HIV-positive people, in adults disseminated TB was found in 87.9% (82.2% to 93.7%) of TB cases and considered the cause of death in 91.4% (95% CI 85.8% to 97.0%) of TB cases (.	Blood; specimens acquired from affected extrapulmonary sites

Abbreviations: CSF: cerebrospinal fluid; HIV: human immunodeficiency virus; MTB: *Mycobacterium tuberculosis*; TB: tuberculosis.
Adapted the table from Index-TB 2016[12].

The clinician should take a careful history noting history of TB exposure, prior TB disease, and medical conditions that increase the risk for TB disease (such as HIV, diabetes mellitus, and low body weight). In comparison with HIV-negative people, HIV-positive people have higher rates of extrapulmonary TB or mycobacteraemia (TB bloodstream infection). HIV-positive patients with signs or symptoms of extrapulmonary TB should have specimens taken from the suspected site(s) of involvement to increase the likelihood of TB diagnosis. In general, children with extrapulmonary TB present in a similar way to that of adults. However, infants and young children are at the highest risk of developing disseminated TB disease and TB meningitis, the most severe forms of TB. In TB meningitis, diagnosis is often delayed with negative consequences for patients. For all forms of extrapulmonary TB, patients may be evaluated in a primary or secondary care setting. However, if more complex or invasive tests are needed, patients may be referred to a tertiary medical centre[4,13,14].

2.4 Diagnosis of tuberculosis

It is estimated that 10.6 million individuals developed tuberculosis (TB) in 2021, although only 6.4 million cases were notified to National TB Programs. This gap of 4.2 million people could be due to delayed patient care, disruptions in health services during the pandemic, suboptimal diagnostics, inaccessibility of accurate and rapid diagnostics or inefficient follow up in the healthcare system. It is a multi-faceted problem, but we will start here with the issues around TB diagnostics.

Early diagnosis and treatment of patients with TB are critical to prevent mortality and control the spread of the disease. However, as exemplified by the global gap in TB diagnosis, achieving early and accurate diagnosis remains a challenge, particularly for low- and middle-income countries which contribute to the majority of the global disease burden[15].

Diagnosis of extrapulmonary TB is challenging for several reasons. Many forms of extrapulmonary TB require invasive diagnostic sampling and gathering adequate specimens can pose a risk of harm to the patient and be costly. Most forms of extrapulmonary TB are paucibacillary (TB disease caused by a smaller number of bacteria), making diagnosis by the conventional method of smear microscopy less sensitive. This problem particularly affects resource-limited settings, where the more sensitive methods of mycobacterial culture and histological examination are not widely available. There are also limitations associated with culture and histology: culture takes several weeks, requires a highly-

equipped laboratory, and has reduced sensitivity in paucibacillary disease; histology relies on highly trained operators and characteristic morphology is shared with other diseases. Additionally, access to histology in low resource settings is a challenge for most health programmes in low-and middle-income countries. As a result of these difficulties, diagnosis of extrapulmonary TB is often made on the grounds of clinical suspicion alone, and many people receive the wrong diagnosis leading to unnecessary TB treatment or poor outcomes from untreated extrapulmonary TB. The demand for faster, reliable diagnostics that are suitable for resource-limited settings is clear, and has been defined by the research community[16]. In 2014, the World Health Assembly unanimously approved the End TB Strategy, a 20-year strategy to end the global TB epidemic. The END TB strategy calls for early diagnosis of TB including universal drug susceptibility testing (DST)[17].

Broadly, diagnosis of TB falls under three major groups of methods: microbiological techniques, imaging techniques and clinical assessment.

2.4.1 Diagnostic tests

Microbiological tests for diagnosing tuberculosis are those which can help detect the presence of *Mtb* bacilli in the specimens collected from the patients. For some tests, they assess the presence of bacilli, its genetic material, or host related biomarkers, which could be in the pathophysiological pathway of TB infection. Currently, some of the widely used microbiological tests for detecting TB are smear microscopy, culture, antigen detection, biomarker detection and nuclei acid amplification tests.

Since the 19th century, sputum smear microscopy is the oldest existing technology for the diagnosis of TB. This method allows for the identification of acid-fast bacilli (AFB) in a sputum sample with the use of a microscope[18]. Even though this technique has major limitations such as low and varied sensitivity, requiring highly trained laboratory technician, it still remains the most widely used tool for TB diagnosis in low- and middle-income countries[19]. The major reason for its high uptake is that it is very inexpensive and can be performed with limited infrastructure. Florescence microscopy was introduced in the 1930s as an alternative to conventional light microscopy. This method uses an acid-fast fluorochrome dye and an intense light source enabling the use of a lower power objective lens compared to the conventional method. This technique has been shown to be more efficient than conventional sputum smear microscopy and offers moderate gains in sensitivity[18,19]. However, the

diagnostic performance varies based on the population as well. For example, for people living with HIV (PLHIV) or pediatric patients with TB, sputum smear microscopy is not optimal, due to lower bacillar load in the specimens or even due to difficulty in getting sputum specimens from these patients. For PLHIV, the sensitivity is even lower, ranging from 22-43%[20,21].

Culture is the current best reference standard for diagnosing TB. There are two different media for culturing *Mtb* from the patients' specimens- liquid and solid media. Irrespective of the smear microscopy or any molecular test results, the current recommendation for TB diagnosis requires performing culture on the collected specimens. Once the culture is positive, it can then be further used for assessing anti-TB drug susceptibility testing to evaluate if the patient is susceptible or resistant to the anti-TB drugs. One major limitation of this method is that it takes weeks (21 days for liquid culture to 8 weeks for solid culture) to get a result for both liquid and solid culture and once it is positive, it takes even longer for drug susceptibility testing. Another major limitation of this technique is that access to culture requires optimal resources, trained human resources and infrastructure. Implementing this in high burden low resource settings is challenging.

WHO has endorsed a large number of new diagnostic technologies during the past 10 years. The amplification and detection of *Mtb* complex (MTBC) nucleic acids has proven to be a highly sensitive and accurate technique for diagnosing TB. Recently published guidelines on rapid TB diagnostics included recommendations on Xpert, moderate complexity automated NAATs for TB detection and resistance to rifampicin and isoniazid, antigen detection in a lateral flow format, low complexity automated NAATs for resistance to isoniazid and second line anti-TB drugs and line probe assays (LPA)[22].

Table 2.2 provides an overview of all currently available NAATs that are endorsed by WHO, along with information on diagnostic accuracy.

Table 2.2: WHO endorsed molecular tests for pulmonary TB detection and drug susceptibility testing[22]

Technology	Xpert® MTB/RIF	Xpert MTB/RIF Ultra	First-line Line probe assays (E.g. GenoType MTBDRplus and NIPRO)	Second-line Line probe assays (E.g. GenoType MTBDRsl)	Loopamp™ MTBC assay	TrueNAT MTB	**Xpert MTB/XDR assay**	**Centralized MTB assay**
Year endorsed	2010	2017	2008	2016	2016	2020	2020	2020
Method principle	qPCR	qPCR / melting temperature analysis (RIF resistance)	PCR, hybridization	PCR, hybridization	loop-mediated isothermal amplification	micro RT-PCR	qPCR / melting temperature analysis (RIF resistance)	qPCR
Intended use	MTB diagnosis & RIF resistance detection	MTB diagnosis & RIF resistance detection	Diagnosis of RIF & INH resistance	Diagnosis of FLQ & SLID resistance	MTB diagnosis	MTB diagnosis	MTB diagnosis of INH, FQs, ETH, SLIDs	MTB diagnosis of RIF & INH resistance
Sensitivity*	98% (SSM+/C+) 67% (SSM-/C+) 95% (RIF resistance)	95% (pooled) 95% (RIF resistance)	98% (RIF resistance) 84% (INH resistance)	86% (FLQ resistance) 87% (SLID resistance)	78% (pooled)	89% (pooled) 93% (RIF resistance)	94% (INH resistance), 93% (FQ resistance), 98 % (ETH resistance) 86 % (AMK resistance)	MTB: 93% INH: 86% RIF: 97%
Specificity*	99% (SSM-/C-) 98% (RIF resistance)	96% (SSM-/C-) 98% (RIF resistance)	99% (RIF resistance) >99% (INH resistance)	99% (FLQ resistance) 99% (SLID resistance)	98% (pooled)	99% (pooled) 95% (RIF resistance)	98% (INH resistance) 98% (FQ resistance) 99,7% (ETH resistance)	MTB: 98% INH: 99% RIF: 99%

							99% (AMK resistance)	
Target setting of use	District or sub-district laboratory	District or sub-district laboratory	Reference laboratory	Reference laboratory	Peripheral laboratory	Peripheral laboratory	Reference laboratory	Reference laboratory
Turnaround time	<2 hours	<2 hours	5 hours	5 hours	<2 hours	<2 hours	<2 hours	5 hours
Amenable to rapid test-and-treat?	Yes	Yes	No	No	Yes	Yes	Yes	Yes

FLQ – fluoroquinolone. INH – isoniazid. LAMP – loop-mediated isothermal amplification. MTB – Mycobacterium tuberculosis. NAAT – nucleic acid amplification tests. RIF – rifampicin. SLID – second line injectable drugs. SSM+/C- – sputum smear microscopy positive / culture positive. 23 SSM-/C+ - sputum smear microscopy negative / culture positive. WHO – World Health Organization.

*n.b.: performance estimates in Table 3.1 have been retrieved from different studies, and are not the result of head to head comparisons. Therefore, comparing performances between tests must be made with caution. All reported values are from the policy guidance document cited.

Line probe assays

Line probe assays (LPA) for first-line TB drugs (INH and RIF) have been endorsed by WHO for over a decade for the detection of multiple-drug-resistant TB (MDR-TB)[23]. These assays include GenoType MTBDRplus (Hain Lifesciences-Bruker, Nehren, Germany) and Nipro NTM+MDRTB II (Osaka, Japan). New-generation LPAs have emerged with higher sensitivity, and some (e.g., GenoType MTBDRsl version 2.0; Hain Lifesciences-Bruker) can detect mutations associated with fluoroquinolones (FLQs) and second-line injectables, kanamycin, amikacin, and capreomycin, and are recommended to guide MDR-TB treatment initiation[24].

Loop-mediated isothermal amplification

Loop-mediated isothermal amplification (LAMP) is an isothermal PCR amplification technique that can be performed in peripheral health care settings. The LAMP-based TB-LAMP assay (Eiken Chemical Company, Tokyo, Japan) has been recommend by WHO as a potential replacement for smear microscopy since 2016, owing to its superior diagnostic performance. It also does not require much sophisticated laboratory equipment[25]. Despite this, TB-LAMP is underutilized[26], but some countries are creating their own LAMP assays for in-country use. Hopefully country-specific versions of LAMP will increase uptake.

Next-generation Xpert testing

In 2010, WHO endorsed Xpert MTB/RIF use with the GeneXpert platform (Cepheid, Sunnyvale, USA) [27], and an updated policy was released in 2013[28]. In 2017, WHO recommended Xpert Ultra (Cepheid) (Ultra), the next generation of Xpert MTB/RIF, as the initial TB diagnostic test for adults and children, regardless of HIV status, over smear microscopy and culture[29]. As in previous generations, Ultra detects RIF resistance by employing four probes with targets in the *rpoB* gene and melting temperature analysis. Compared with previous generations, Ultra test cartridges have a larger chamber for DNA amplification than Xpert MTB/RIF and two multicopy amplification targets for TB, namely, *IS6110* and *IS1081*, for a lower limit of detection of 16 CFU/ml. These modifications have increased Ultra's overall sensitivity from 85% (95% confidence interval [CI], 82% to 88%) to 88% (95% CI, 85% to 91%); however, compared with the previous generation, Ultra's specificity is lower at 96% (95% CI, 90% to 98%) versus 98% (95% CI, 97% to 98%), seemingly because it detects nonviable bacteria, particularly in people with recent TB [22,30]. This lower specificity is proving to be

an important issue in certain settings, such as areas with high numbers of HIV-TB coinfections or recurrent TB cases, like South Africa. In a recent study by Mishra and colleagues, it was shown that the Xpert Ultra assay had significantly lower specificity and positive predictive value than the Xpert MTB/RIF assay and high numbers of Ultra positive/culture negative people with previous treatment [31]. The clinical consequences of treating such patients are unclear, and ongoing studies are attempting to shed light on this information.

The Xpert Ultra test also has a semiquantitative "trace" category, indicating bacilli at the lowest limits of detection. In instances of trace positives (termed "trace calls"), one of the two multicopy amplification targets, but not the *rpoB* sequences, are detected. In instances of suspected extrapulmonary TB, children, and people living with HIV (PLHIV), trace positives should be treated as positives, as these cases tend to be paucibacillary. For other cases, a fresh specimen should be retested to rule out false positives[29]. Trace calls may be difficult to interpret, as in the aforementioned study by Mishra et al., where it was observed that among people who were previously treated for TB, trace positives were a substantial portion of all positives, and these individuals by definition had indeterminate results for RIF resistance and were culture negative, precluding further DST [31]. Trace calls may be improving Ultra's sensitivity for extrapulmonary TB, particularly in the context of definite or probable TB meningitis.

As an automated PCR-based test, Ultra can be used by minimally trained technicians, but as it runs on the GeneXpert platform, it requires a continuous power supply and computer which limits its use as a true point-of-care (POC) test. Alternatively, the recently launched GeneXpert Edge system is battery powered and utilizes a tablet, making it more portable.

Truelab by Molbio

Truenat MTB, Truenat MTB Plus, and Truenat MTB-Rif Dx (Molbio Diagnostics, Goa, India) are chip-based, micro real-time PCR-based assays for TB detection that produce results in 1 hour on the portable Truelab platform (Molbio Diagnostics). Already being rolled out in India, Truenat is characterized as a more affordable alternative to Xpert that is made in India. Products that are developed and manufactured in a country with a high TB burden might be quicker and more straightforward to scale up in that country than products developed in another country, as governments

often already have a degree of buy-in, data from locally run studies will have accumulated, and supply chain and regulatory issues are simpler to solve[32,33].

Truenat MTB and Truenat MTB Plus assays detect *M. tuberculosis* bacilli in sputum after extraction using the separate TruePrep instrument and kits, with Truenat MTB-Rif Dx available as an optional add-on chip for sequential RIF resistance detection [33]. Truelab, which comes in Uno-, Duo-, and Quattro-throughput formats, was designed to be "rugged" and POC friendly, as it has a dust filter and runs in temperatures up to 30°C, but multiple micropipetting steps necessitate a trained technician for its operation.

In December 2019, WHO convened a guideline development group meeting to determine recommended use cases for Truenat assays and other rapid molecular tests. The subsequent rapid communication reported that Truenat MTB, MTB Plus, and MTB-Rif Dx assays displayed comparable sensitivities and specificities to Xpert MTB/RIF and Ultra for the detection of TB and RIF resistance, although this report was based on an interim analysis of a multicenter study that is still ongoing. The 2020 WHO Consolidated Guidelines on Molecular Diagnostics recommend using Truenat MTB or MTB Plus rather than smear microscopy as an initial diagnostic test for TB in adults and children with signs and symptoms of pulmonary TB. This is a conditional recommendation, as test accuracy certainty is moderate. Regarding DST, with a Truenat MTB- or MTB Plus-positive result, Truenat MTB-RIF Dx may be used as an initial test for rifampicin resistance rather than phenotypic DST. This is also a conditional recommendation, as there is very low certainty of evidence for test accuracy [22].

Moderate complexity assays

Recently, centralized, high-throughput NAATs for TB diagnosis and drug resistance detection have been developed and are currently undergoing WHO evidence evaluation. RealTime MTB (Abbott Molecular, Abbott Park, USA), RealTime RIF/INH (Abbott Molecular), FluoroType MTB (Hain Lifescience, Nehren, Germany), FluoroType MTDBR (Hain Lifescience), Cobas MTB (Roche, Rotkreuz, Switzerland), and Max MDR-TB (BD, Franklin Lakes, USA) assays run on established multidisease platforms that are already employed for such diseases as human immunodeficiency virus (HIV), human papillomavirus, and hepatitis C virus[34]. These almost entirely automated tests are all intended for tertiary laboratory use. In 2020, WHO convened a guideline development group and reported that the centralized assays' performance for detecting resistance to INH and RIF was similar

to LPA and that RealTime MTB, Cobas MTB, and Max MDR-TB performed similarly to Xpert MTB/RIF for TB detection[22].

The RealTime MTB is a multiplex NAAT that targets the MTB *IS6100* and *PAB* genes with a limit of detection (LOD) of 17 CFU/ml. Up to 96 respiratory specimens can be inactivated and processed by the Abbott m2000 platform per run[34].

Another centralized test is the semiautomated FluoroType MTB, a beacon-based PCR assay performed on the Hain Fluorocycler platform. Specimen decontamination, sample preparation, and DNA isolation must be performed manually, which requires 30 min of hands-on time, with the entire process taking 4 h to final results[35].

The Cobas 6800/8800 MTB assay runs on the high-throughput Cobas 8800 platform that can run 960 samples in 8 h. One internal manufacturer study of 744 samples reported a sensitivity and specificity of 95% (95% CI, 92% to 97%) and 98% (95% CI, 96% to 99%), respectively [22].

Finally, the Max MDR-TB test runs on the BD Max platform and targets the MTB 16S rRNA gene. Up to 24 specimens are manually decontaminated and prepared before extraction and amplification by the Max MDR-TB assay. Time to final results is 4 hours [36,37]. A manufacturer-sponsored validation study of 892 samples reported TB detection sensitivity of 93% (95% CI, 89% to 96%) and specificity of 97% (95% CI, 96% to 98%). Sensitivity for RIF resistance and INH resistance was 90% (95% CI, 55% to 100%) and 82% (95% CI, 63% to 92%), respectively, with 100% specificity in both cases [37].

Centralized TB assays are promising due to their high diagnostic accuracy and ability to run large numbers of samples simultaneously, and their automated nature reduces the hazard of contacting infectious respiratory specimens for health care workers and laboratory technicians. The developmental pipeline for centralized assays is quite robust, with platforms, such as MeltPro (Zeesan Biotech, Xiamen, China), Seegene (Seoul, South Korea), and MolecuTech (YD Diagnostics, Seoul, South Korea), currently under regulatory assessment[38]. All platforms are offering tests for MDR-TB and XDR-TB, which will provide more options in the future.

However, carry-over contamination is still possible with these assays, and quality assurance is critical. Additionally, the costs for each of these tests have not been made public, and no subsidized or concessional pricing schemes are yet in place. These tests do run on multi-disease platforms, which adds value, but it is unclear exactly who will be willing to pay to implement these tests if they can only perform DST for INH and RIF resistance, particularly when there are simpler NAATs available (Table 2.2). Furthermore, their centralized placement means they are unavailable where patients first present to care, and therefore, sample transportation is essential for success. Reliable systems for delivering test results to patients and health care providers must also be in place for these tests to have impact.

2. 5 Emerging technologies

Xpert XDR

Another PCR-based cartridge has been designed to run on the GeneXpert platform for the simultaneous detection of mutations associated with resistance to multiple first- and second-line TB drugs, or extensively drug resistant TB (XDR-TB). Against phenotypic drug-susceptibility testing, Xpert XDR displayed sensitivities (95%CI) of 83.3% (77.1-88.5) for isoniazid, 88.4% (80.2-94.1) for ofloxacin, 96.2% (87.0-99.5) for moxifloxacin at a critical concentration of 2.0 μg per milliliter, 71.4% (56.7-83.4) for kanamycin, and 70.7% (54.5-83.9) for amikacin[39]. However, as WHO updates treatment guidelines for MDR-TB and XDR-TB, it will be critical that molecular tools for DST can be updated to quickly reflect new recommendations. Already, this iteration of Xpert XDR may have less impact than it otherwise would have, as WHO has de-emphasized second-line injectable agents in treating drug resistant forms of TB. Future developments will need to focus on drugs that are now critical for MDR and XDR-TB management, including bedaquiline, pretomanid, and linezolid.

Indigenously developed Chinese diagnostics

Similar to Molbio in India, Chinese biotechnology firms have used their own expertise and developing TB NAATs for in-country use. Instead of waiting for WHO recommendation before attempting to commercialize their assays, these companies have undergone the CFDA regulatory processes, received approval, and rolled out the tests locally. The CFDA has approved several NAATs for commercial use in China. However, none of these technologies have been reviewed by WHO and international uptake, therefore, is limited. Table 2.3 summarizes the performance of some of these assays.

Table 2.3: CFDA-endorsed molecular test for TB diagnosis and drug susceptibility testing[40,41].

Technology	Easynat	SAT-TB	MeltPro TB	GeneChip MDR
Method principle	Cross priming amplification	Isothermal amplification of MTB 16s RNA	PCR, melt curve analysis	PCR, hybridization
Intended use	MTB diagnosis	MTB diagnosis	DST	MDR-TB diagnosis; INH, RIF resistance
Sensitivity	87% (pooled)	71% - 94% [range]	98% (RIF resistance) 85% (INH resistance) 64% (FLQ resistance) 83% (SLID resistance)	79% (MDR-TB) 89% (RIF resistance) 79% (INH resistance)
Specificity	97% (pooled)	54% – 83% [range]	97% (RIF resistance) 98% (INH resistance) 98% (FLQ resistance) 99% (SLID resistance)	98% (MDR-TB) 97% (RIF resistance) 97% (INH resistance)
Target setting of use	District or sub-district laboratory	District or reference laboratory	Reference laboratory	Reference laboratory

CFDA – China Food and Drug Administration. DST – drug susceptibility testing. INH – isoniazid. RIF – rifampicin. SLID – second line infectable drugs.

CFDA-approved since 2014, EasyNat (Ustar Biotechnologies, Hangzhou, China) replicates and detects Mycobacterial DNA from sputum via cross-priming amplification (CPA). As CPA is an isothermal technique, Easynat may be placed at low levels of healthcare systems as a thermal cycler is not required[40].

Simultaneous amplification and testing (SAT)-TB (Rendu Biotechnology, Shanghai) detects Mycobacterial 16S rRNA from sputum, which is isothermally amplified before the resultant cDNA is detected by fluorescent probes, requiring laboratory infrastructure[42].

For drug resistance testing, MeltPro TB (Zeesan Biotech, Xiamen) assays for RIF, INH, second-line injectables, and fluoroquinolones are available, allowing them to detect MDR-TB and XDR-TB. After manual DNA extraction, MeltPro TB detects drug resistance via melt curve analysis using a PCR

machine; the shift in melting temperature from wildtype to mutation in sequences covered by multiple probes can be qualitatively detected[43].

GeneChip MDR (CapitalBio Corporation) is a microarray assay that requires hands-on sample preparation before reverse hybridization and analysis on a fully automated system. As such, it requires sophisticated laboratory equipment. GeneChip MDR utilizes multiplexed asymmetric PCR to detect resistance to RIF and INH in one assay, and thus can detect MDR-TB[44].

Next-generation sequencing

Next-generation sequencing (NGS) is increasingly considered a promising option for rapid and comprehensive DST for TB[45]. Unlike probe-based assays where detection is limited to probes' specific targets, NGS-based assays can provide detailed and accurate sequence information for whole genomes, as with whole genome sequencing (WGS), or multiple gene regions of interest, as with targeted sequencing[46]. (Table 2.4).

Table 2.4: Strengths and limitations of WGS versus targeted sequencing.

Whole genome sequencing	Targeted sequencing
Strengths • Full genome sequenced • No pre-specified targets needed • Comprehensive solution • Detect rare mutations and hetero-resistance **Weaknesses** • Requires culture isolates • Slower than targeted NGS • Complicated bioinformatics • Expensive	**Strengths** • Sequence directly from sample • Large number of gene targets • Less expensive than WGS • Simpler bioinformatics and storage • Detect rare mutations and hetero-resistance **Weaknesses** • Knowledge of targets required • Less information than WGS • Expensive

Acknowledging the value of NGS, WHO has published guidance on the role of sequencing for detecting mutations associated with drug resistance in TB[46], along with a consensus-based Target Product Profile for sequencing. In 2019, a TB sequencing database called ReSeqTB was established at WHO to curate, standardize, and unify genotypic and phenotypic DST data, along with metadata on drug resistant-TB (DR-TB)[47].

There are ongoing efforts by multiple stakeholders to validate targeted sequencing as a full solution, from DNA extraction in respiratory samples, to library preparation and sequencing, to result reporting (Figure 2.3). One such assay is Deeplex® Myc-TB (Genoscreen, Lille, France). Deeplex Myc-TB uses ultra-deep sequencing of 24-plex amplicon mixes for Mycobacterial species identification, genotyping, and DST. In addition, it enables detection of hetero-resistance down to 3% of minority strains in case of multiple infections or emergent mutations[48]. Another newly-developed targeted sequencing assay for DR-TB is DeepChek®-TB (Translational Genomics Research Institute, Flagstaff, USA), which has recently been licensed by ABL (Luxembourg). Both tests are currently for research use only. This evidence is being generated and will be reviewed by the WHO in 2023.

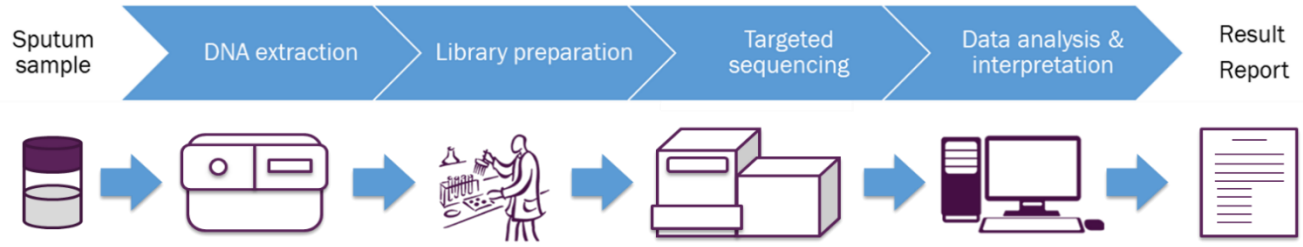


Figure 2.3: Targeted sequencing workflow schematic.

Sequencing is currently being successfully implemented for DR-TB surveillance purposes in at least seven countries - Azerbaijan, Bangladesh, Belarus, Pakistan, Philippines, South Africa, and Ukraine[49]. Select health systems in low burden settings, including the United Kingdom (Public Health England), the Netherlands, and New York state, have already transitioned to WGS from phenotypic culture for DST. More countries are expressing interest in switching to a sequencing-based approach for surveillance.

India has recently expressed interest in utilizing sequencing for surveillance and clinical care. In 2018, infrastructure and technical support for sequencing was introduced at five National TB Program laboratories across India with Global Fund funding. It is hoped that this will be the beginning of the foundations of a clinical diagnostic network in the future.

South Africa has implemented and integrated sequencing into their national drug resistance surveillance program as an alternative to phenotypic DST, and are considering its future potential for laboratory-based TB management and TB transmission investigations.

In Brazil, the interdisciplinary group Rede Brasileira de Pesquisas em Tuberculose (REDE-TB, Brazilian TB Research Network) identified NGS as a key technology for implementation. Through the Oswaldo Cruz Foundation (Fiocruz), Brazil has also signed memoranda of understanding with the Beijing Genomic Institute and the Chinese Centre for Disease Control and Prevention. One of the planned activities under this agreement is the establishment of a sequencing service at Fiocruz with applications in infectious disease, including TB.

Regarding sequencing for DST, centralized sequencing platforms have been the norm, but there is increasing interest in smaller and more portable sequencing devices, such as MinION (Oxford Nanopore, Oxford, UK) [50]and iSeq from Illumina (San Diego, USA)[51], validation for both of which is on-going.

In my thesis, I performed a systematic review, published in the Cochrane Library, of Xpert MTB/RIF and Ultra for different forms of extrapulmonary TB. Existing diagnostic tests for extrapulmonary TB are not sensitive enough or are invasive and costly. This Cochrane Systematic Review estimated sensitivity and specificity of Xpert® MTB/RIF for detection of extrapulmonary TB and rifampicin resistance. The purpose of Xpert® MTB/RIF is diagnosis of TB and detection of rifampicin resistance. The role of Xpert® MTB/RIF is a replacement for standard practice, which includes obtaining appropriate specimens from the suspected sites of involvement for microbiological (conventional microscopy and culture) and histological examination.

Additionally, I performed a systematic review and meta-analyses, published in the European Respiratory Journal, of accuracy of moderate complexity assays for TB, rifampicin and isoniazid resistance detection for pulmonary TB. The rationale for this review was that with the increasing use

of molecular NAATs for TB, especially in high TB burden countries and an increased notification of isoniazid mono-resistant cases, it is important to use high throughput molecular NAATs which can perform several samples at the same time and also provide information on both rifampicin and isoniazid resistance.

CHAPTER 3- Manuscript

Xpert MTB/RIF Ultra and Xpert MTB/RIF assays for extrapulmonary tuberculosis and rifampicin resistance in adults

Mikashmi Kohli[1], Ian Schiller[2], Nandini Dendukuri[2], Mandy Yao[2], Keertan Dheda[3,4,] Claudia M Denkinger[5,6,] Samuel G Schumacher[5], Karen R Steingart[7]

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
[2]Centre for Outcomes Research, McGill University Health Centre - Research Institute, Montreal, Canada
[3]Centre for Lung Infection and Immunity Unit, Department of Medicine and UCT Lung Institute, University of Cape Town, Cape Town, South Africa
[4]Faculty of Infectious and Tropical Diseases, Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, London, UK
[5]FIND, Geneva, Switzerland
[6]Division of Tropical Medicine, Centre for Infectious Diseases, University Hospital Heidelberg, Heidelberg, Germany
[7]Honorary Research Fellow, Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

Published in Cochrane Database Syst Rev. 2021 Jan 15;1(1):CD012768[52].

Background

Tuberculosis (TB) causes tremendous suffering worldwide and has surpassed HIV/AIDS as the world's leading infectious cause of death. The WHO estimates that from 2000 to 2019 more than 60 million lives were saved by diagnosing and treating tuberculosis. However, the COVID-19 pandemic threatens the gains made over recent years. A modelling study by the WHO suggests that there could be between 200,000 and 400,000 additional tuberculosis deaths in 2020 if, over a period of three months, 25% to 50% fewer people were detected and treated with tuberculosis [1].

Of the 7.1 million new cases of tuberculosis notified to the WHO in 2019, 16% were cases of extrapulmonary tuberculosis, (range, 8% in the WHO Western Pacific Region to 24% in the WHO Eastern Mediterranean Region) [1,53]. Among countries in the European Union, extrapulmonary tuberculosis was responsible for 19% of all notified cases (range, 6% to 44%)[54]. A large retrospective analysis from China found that of 19,279 hospitalised tuberculosis patients, around 33% had extrapulmonary tuberculosis[55]. The number of people affected by extrapulmonary tuberculosis is likely to be higher, given that, according to the WHO, extrapulmonary tuberculosis is notified as pulmonary tuberculosis when the two forms exist together, and diagnosing extrapulmonary tuberculosis is challenging, as described below. Additionally, extrapulmonary tuberculosis accounts for an increasing proportion of tuberculosis cases in some countries, in part because of host and genetic considerations, and the association of extrapulmonary tuberculosis and HIV [10,56–58]. Based on surveillance and epidemiological data, extrapulmonary tuberculosis affects a greater proportion of children than adults[59].

Drug-resistant tuberculosis is a serious threat to global health. For the purpose of surveillance and treatment, drug-resistant tuberculosis is classified as rifampicin-resistant tuberculosis, multidrug-resistant tuberculosis (MDR-TB), and extensively drug-resistant tuberculosis. MDR-TB is defined as resistance to at least isoniazid and rifampicin, the two most important first-line anti-tuberculosis drugs. Extensively drug-resistant tuberculosis is defined as MDR-TB plus resistance to at least one drug in the following two classes of medicines used in treatment of MDR-TB: fluoroquinolones and second-line injectable agents. In 2019, there were approximately half a million new cases of rifampicin-resistant tuberculosis (of which 78% had MDR-TB), with India (27%), China (14%) and the Russian

Federation (9%) accounting for the largest burden [1]. In 2019, 12,350 cases of extensively drug-resistant tuberculosis were reported [53].

In 2014, the World Health Assembly unanimously approved the WHO End TB Strategy, a 20-year strategy devised to end the global tuberculosis epidemic[17]. Early diagnosis of tuberculosis, including universal drug susceptibility testing (DST) and systematic screening of contacts and high-risk groups, is a part of pillar one of the strategies.

Objectives

To estimate the diagnostic accuracy of Xpert Ultra and Xpert MTB/RIF for a) extrapulmonary tuberculosis by site of disease and b) rifampicin resistance, in adults with presumptive extrapulmonary tuberculosis. Presumptive tuberculosis refers to a patient who presents with symptoms or signs suggestive of tuberculosis.

Secondary objectives

- To compare the diagnostic accuracy of Xpert Ultra and Xpert MTB/RIF for a) extrapulmonary tuberculosis by site of disease, and b) rifampicin resistance.
- To investigate the effects of potential sources of heterogeneity on test accuracy across the included studies.

For potential sources of heterogeneity, for extrapulmonary tuberculosis, we included smear status, HIV status, and prevalence of extrapulmonary tuberculosis. For cerebrospinal fluid (CSF), we considered the presence of a concentration step and specimen volume.

For rifampicin resistance, we planned to assess the impact of the prevalence of rifampicin resistance on accuracy estimates, but we had insufficient data for this analysis.

Methods

Criteria for considering studies for this review

Types of studies

We included cross-sectional and cohort studies. In addition, we had planned to include randomized controlled trials that evaluated the use of the index(s) test on patient health outcomes, but that also reported sensitivity and specificity. Although the study design was a randomized trial for the purpose of determining the impact of the test on participant outcomes, the study design was a cross-sectional study for the purpose of determining the diagnostic accuracy of the index tests in this review. However,

we did not identify any randomized controlled trials. We used abstracts to identify published studies and included these when they met the inclusion criteria. We only included studies that reported data comparing the index test(s) to an acceptable reference standard from which we could extract true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values. We excluded case-control studies and case reports.

Participants

We included studies where at least 85% of the participants enrolled were adults aged 15 years or older with presumptive extrapulmonary tuberculosis from all settings and countries. Restricting the age group to adults differs from the original review, where we also included children[60]. We did this because children are now included in a separate Cochrane Review[61]. We excluded studies where we could not disaggregate data on adults from those in children and studies where we could not tell the age of the participants enrolled.

We included non-respiratory specimens (such as CSF, pleural fluid, lymph node aspirate or tissue). We excluded sputum and other respiratory specimens, such as fluid obtained from bronchial alveolar lavage and tracheal aspiration. As we anticipated finding many studies, we set a bar to exclude smaller studies to reduce unnecessary work. We therefore required studies to provide data for at least five specimens for a given form of extrapulmonary tuberculosis included in the review. We excluded studies evaluating the use of Xpert Ultra and Xpert MTB/RIF to diagnose relapse of previously-treated extrapulmonary tuberculosis, so as to avoid the selection bias that may arise by limiting to a group that is already at elevated risk of extrapulmonary tuberculosis. We attempted to identify studies that included participants who were not taking anti-tuberculosis drugs or had taken anti-tuberculosis drugs for less than seven days.

Index tests

The index tests were Xpert Ultra and Xpert MTB/RIF.

Index test results are automatically generated (i.e. there is a single threshold), and the user is provided with a printable test result as follows.

Xpert Ultra

- MTB (M tuberculosis) DETECTED HIGH; RIF (rifampicin) Resistance DETECTED

- MTB DETECTED MEDIUM; RIF Resistance DETECTED
- MTB DETECTED LOW; RIF Resistance DETECTED
- MTB DETECTED VERY LOW; RIF Resistance DETECTED
- MTB DETECTED HIGH; RIF Resistance NOT DETECTED
- MTB DETECTED MEDIUM; RIF Resistance NOT DETECTED
- MTB DETECTED LOW; RIF Resistance NOT DETECTED
- MTB DETECTED VERY LOW; RIF Resistance NOT DETECTED
- MTB DETECTED HIGH; RIF Resistance INDETERMINATE
- MTB DETECTED MEDIUM; RIF Resistance INDETERMINATE
- MTB DETECTED LOW; RIF Resistance INDETERMINATE
- MTB DETECTED VERY LOW; RIF Resistance INDETERMINATE
- MTB Trace DETECTED; RIF Resistance INDETERMINATE
- INVALID (the presence or absence of MTB cannot be determined)
- ERROR (the presence or absence of MTB cannot be determined)
- NO RESULT (the presence or absence of MTB cannot be determined)

Xpert Ultra incorporates a semi-quantitative classification for results: trace, very low, low, moderate, and high. 'Trace' corresponds to the lowest bacterial burden for detection of *M tuberculosis*[62]. We considered a trace result to mean MTB (*M tuberculosis*) DETECTED. However, no rifampicin-resistance result was available for participants with trace results because the trace sample is always reported as 'INDETERMINATE' for rifampin resistance [63].

Xpert MTB/RIF

- MTB (*M tuberculosis*) DETECTED; Rif (rifampicin) resistance DETECTED
- MTB DETECTED; Rif resistance NOT DETECTED
- MTB detected; Rif resistance INDETERMINATE
- MTB NOT DETECTED
- INVALID (the presence or absence of MTB cannot be determined)
- ERROR (the presence or absence of MTB cannot be determined)
- NO RESULT (the presence or absence of MTB cannot be determined)

Target conditions

The target conditions were extrapulmonary tuberculosis and rifampicin resistance. We included eight common forms of extrapulmonary tuberculosis and considered subcategories of the target condition as separate diagnostic classifications [4,54,64].

- Tuberculous meningitis.
- Pleural tuberculosis.
- Lymph node tuberculosis.
- Genitourinary tuberculosis.
- Bone or joint tuberculosis.
- Peritoneal tuberculosis.
- Pericardial tuberculosis.
- Disseminated tuberculosis.

Table 2.1 lists the forms of extrapulmonary tuberculosis and specimens used for diagnosis in the review. We excluded less common forms, such as cutaneous tuberculosis, ocular tuberculosis, female genital tuberculosis, and tuberculosis of the breast, ear, and paranasal sinuses[4].

Reference standards

Detection of extrapulmonary tuberculosis

We included two reference standards.

- Solid or liquid mycobacterial culture.
    - ‘Tuberculosis' was defined as a positive *M tuberculosis* culture
    - ‘Not tuberculosis' was defined as a negative *M tuberculosis* culture
- Composite reference standard.
    - 'Tuberculosis' was defined as a positive *M tuberculosis* culture or positive composite reference test.
    - ‘Not tuberculosis' was defined as a negative *M tuberculosis* culture and a negative composite reference test.

The composite reference standard might be based on the results of microbiological tests, culture or NAAT other than Xpert Ultra and Xpert MTB/RIF; imaging studies; histology; and clinical

characteristics, and include at least one component test that is positive, according to the definition of the primary study authors.

For pleural tuberculosis, we defined the composite reference standard as the presence of granulomatous inflammation or a positive culture. We proposed this definition because we found evidence to support including histopathological examination in the definition. Around 60% of patients undergoing pleural biopsy will show granulomatous inflammation[65]. In a prospective cohort study of participants with clinical and radiological findings consistent with pleural tuberculosis, Conde 2003 found that histological examination of tissue obtained from pleural biopsy had a higher diagnostic yield (78%; 66/84) than that of culture (62%; 52/84).

Culture is considered the best reference standard for tuberculosis. However, culture may lead to misclassification of some cases of extrapulmonary tuberculosis as 'not tuberculosis', owing to the paucibacillary nature of the disease. This means that culture may have low sensitivity for extrapulmonary tuberculosis overall and further that culture sensitivity may differ for different forms of extrapulmonary tuberculosis[65]. This misclassification by culture may lead to biased estimates (overestimation or underestimation) of the diagnostic accuracy of Xpert Ultra and Xpert MTB/RIF. The extent of bias will depend on the frequency of errors by culture and the degree of correlation in errors by culture and the Xpert assays because culture and Xpert Ultra or Xpert MTB/RIF are likely to pick up cases with a higher bacterial load, and are likely to miss cases with a lower bacterial load. Ignoring this dependence could lead to an overestimation of the sensitivity of Xpert Ultra or Xpert MTB/RIF.

- Effect of low sensitivity of culture on Xpert sensitivity: the low sensitivity of culture means that index test FNs may be misclassified as TNs when culture is used as the reference standard. Therefore, when Xpert Ultra or Xpert MTB/RIF is evaluated against culture, the number of FNs (classified as negative by the index test and positive by the reference test) may be decreased and the sensitivity of the index test may be overestimated.
- Effect of low sensitivity of culture on Xpert specificity: the low sensitivity of culture means that index test TPs may be misclassified as FPs when culture is used as the reference standard. Therefore, when Xpert Ultra or Xpert MTB/RIF is evaluated against culture, the number of FPs (classified as positive by the index test and negative by the reference test) may be increased and specificity of the index test may be underestimated.

In contrast to culture, a composite reference standard that includes culture, other tests, and clinical characteristics may correctly classify index test results as TPs (instead of as FPs with respect to culture), especially in people with paucibacillary disease in whom culture may be negative. However, because of the uncertainties that surround a clinical diagnosis of tuberculosis and, in some instances, the conditional dependence of the index tests and other tests in the composite reference standard (for example, for most of these tests, detection of tuberculosis depends on bacillary load), a reference standard that uses additional tests and clinical characteristics (in culture-negative people) may incorrectly classify people without tuberculosis as having tuberculosis[66]. An additional challenge with including a composite reference standard is that the definition of the composite reference standard may vary across studies, making it difficult to interpret the accuracy estimates.

Thus both reference standards, culture and composite, are imperfect and may affect accuracy estimates. In an attempt to improve the estimation of diagnostic accuracy, we applied a latent class meta-analysis model to the three most commonly studied forms of extrapulmonary tuberculosis. This approach provides the sensitivity and specificity of culture in addition to the accuracy of the index tests, thus adjusting for imperfect culture accuracy.

Detection of rifampicin resistance

The reference standard was culture-based DST using solid or liquid media or line-probe assays, as recommended by the WHO [22].

Search methods for identification of studies

We attempted to identify all relevant studies, regardless of language or publication status (published, unpublished, in press, or ongoing). We monitored abstracts to see if these studies were published during the time we performed the review. We included only published studies in the review.

Electronic searches

For the original review, we searched the literature on 7 August 2017. For this review update, we searched the literature on 2 August 2019 and again on 28 January 2020, specifically for studies of

Xpert Ultra (studies could include Xpert Ultra alone or both Xpert Ultra and Xpert MTB/RIF), using the search terms and strategy described in Appendix. We searched the following databases:

- Cochrane Infectious Diseases Group Specialized Register;
- MEDLINE (OVID, from 1966);
- Embase (OVID, from 1974);
- Science Citation Index - Expanded (from 1900);
- Conference Proceedings Citation Index - Science (CPCI-S, from 1990);
- BIOSIS Previews (from 1926), all three from the Web of Science;
- Scopus (Elsevier, from 1970);
- Latin American Caribbean Health Sciences Literature (LILACS) (BIREME, from 1982).

We also searched ClinicalTrials.gov, the WHO International Clinical Trials Registry (ICTRP) Platform (www.who.int/trialsearch), and the International Standard Randomized Controlled Trials Number (ISRCTN) registry (www.isrctn.com/) for trials in progress, and ProQuest Dissertations & Theses A&I (www.proquest.com/pqdtglobal, from 1990) for dissertations.
To identify other systematic reviews and meta-analyses, we performed an additional search on 28 May 2020 in MEDLINE (PubMed), Embase (OVID), and the Cochrane Library, applying filters for systematic reviews (www.sign.ac.uk/search-filters.html) to search terms for Xpert and tuberculosis.

Searching other resources
We reviewed reference lists of included articles and any relevant review articles identified through the above methods. We also contacted researchers at FIND and other experts in the field of tuberculosis diagnostics for information on ongoing and unpublished studies.

Data collection and analysis
Selection of studies
We used Covidence to manage the selection of studies[67]. Two review authors independently scrutinized titles and abstracts identified by electronic literature searching to identify potentially eligible studies. We selected any citation identified by either review author as potentially eligible for full-text review. The same review authors independently assessed full-text papers for study eligibility using predefined inclusion and exclusion criteria, and resolved any discrepancies by discussion. We recorded all studies excluded after full-text assessment and their reasons for exclusion

in Characteristics of excluded studies. We illustrated the study selection process in a PRISMA diagram [68].

Data extraction and management

Using a previously-developed form (Appendix), two review authors worked independently to extract data on the following characteristics.

- Author; publication year; country; setting (outpatient, inpatient, or both outpatient and inpatient); study design; manner of participant selection; number of participants enrolled; number of participants for whom results are available.
- Characteristics of participants: gender; age; HIV status; history of prior tuberculosis; receipt of anti-tuberculosis treatment.
- Index test.
- Target condition and subcategories.
- Type of reference standard.
- Quality Assessment of Studies of Diagnostic Accuracy - Revised (QUADAS-2) items.
- Details of specimen: type (such as CSF, pleural fluid, or lymph node aspirate or tissue); condition (fresh or frozen); smear-positive or smear-negative.
- Specimen preparation; homogenization step (for tissue specimens); concentration step and specimen volume (for CSF); adherence to WHO standard operating procedures.
- Number of TP, FP, FN, TN (i.e. true-positives, false-positives, false-negatives, and true-negatives), and trace results; number of inconclusive results for detection of extrapulmonary tuberculosis; number of indeterminate results for detection of rifampicin resistance.
- Number of missing or unavailable test results.

We classified country income status as either low- and middle-income or high-income, according to the World Bank List of Economies[69].

We extracted TP, FP, FN, and TN values for the following specimens: CSF, pleural fluid and tissue, lymph node aspirate and tissue (the latter specimen acquired by surgical biopsy), bone or joint aspirate and tissue, urine, peritoneal fluid and tissue, pericardial fluid and tissue, and blood. We extracted these values for each of the specimen types separately. For example, we used one 2 × 2 table for lymph node

aspirate, and another 2 × 2 table for lymph node tissue. In situations in which a participant contributed more than one specimen but of different types, we extracted data for all specimens. When a study included data for both raw specimens and concentrated sediment involving the same participants, we preferentially extracted data for raw specimens, except in the case of CSF, for which we extracted data for concentrated sediment as recommended by the WHO[70]. We extracted accuracy data according to the defined reference standards. We did not encounter any situations in which a subset of participants in a study received the reference standard but others did not. Hence, there was no need to make corrections for verification bias in the statistical analysis [71].

In most studies, the number of specimens was the same as the number of participants. However, in some studies, the number of specimens exceeded the number of participants or study authors reported only the number of specimens. In the previous review [60], we added post hoc a sensitivity analysis limiting inclusion to studies that used one specimen per participant. In this review, we performed a similar sensitivity analysis for Xpert Ultra.

We contacted authors of primary studies for missing data or clarifications. We entered all data into Microsoft Excel 2014.

Assessment of methodological quality

We used the QUADAS-2 tool, tailored to this review, to assess the quality of the included studies (Appendix)[72]. QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing. We assessed all domains for risk of bias and the first three domains for concerns about applicability. Two review authors independently completed QUADAS-2 and resolved disagreements through discussion. We present the results of this quality assessment in text, tables, and graphs.

We followed Cochrane policy, which states that "authors of primary studies will not extract data from their own study or studies. Instead, another author will extract these data, and check the interpretation against the study report and any available study registration details or protocol".

Statistical analysis and data synthesis

We performed descriptive analyses of the characteristics of included studies using Stata 15 [73]. We used data reported in the TP, FP, FN, and TN format to calculate sensitivity and specificity estimates and 95% confidence intervals (CIs) for individual studies. We present individual study results graphically by plotting the estimates of sensitivity and specificity (and their 95% CIs) in forest plots and receiver operating characteristic (ROC) space using Review Manager 5 (RevMan 5) [74].

When data were sufficient, we performed meta-analyses to estimate pooled sensitivity and specificity and corresponding 95% credible interval (CrI, defined below) using an adaptation of the bivariate random-effects approach of Reitsma 2005, which uses the exact binomial likelihood for the observed proportions[75]. The bivariate random-effects approach allowed us to calculate the pooled estimates of sensitivity and specificity while dealing with potential sources of variation caused by (1) imprecision of sensitivity and specificity estimates within individual studies; (2) correlation between sensitivity and specificity across studies; and (3) variation in sensitivity and specificity between studies. The model has a hierarchical structure, with the logit sensitivity in individual studies assumed to come from a common probability distribution the mean of which is the pooled logit sensitivity, and the standard deviation is the between-study standard deviation, and likewise for the specificity. This structure allows for borrowing strength across studies. In the absence of sufficient studies, we simply present descriptive statistics. In addition, we determined predictive values at a pretest probability of 10%, a value suggested by the WHO.

We performed separate analyses grouped by type of extrapulmonary specimen (e.g. CSF, pleural fluid, peritoneal fluid) rather than determine summary accuracy estimates for all forms of extrapulmonary tuberculosis combined, because we considered the former approach to be most clinically meaningful. In addition, we performed separate analyses by reference standard.

Comparison of Xpert Ultra and Xpert MTB/RIF

We performed comparative meta-analyses by restricting the analyses to only those studies that made direct comparisons between Xpert Ultra and Xpert MTB/RIF within the same participants[76]. We extracted the median and the 95% CrI for the difference in the pooled sensitivities and the difference in the pooled specificities, respectively, of Xpert Ultra versus Xpert MTB/RIF. We also calculated the probability that the difference exceeds zero in each case.

For analysis of Xpert MTB/RIF or Xpert Ultra accuracy for detection of rifampicin resistance, we include participants who (1) were culture-positive; (2) had a valid culture-based DST or line-probe assay (LPA) result; (3) were Xpert MTB/RIF or Xpert Ultra tuberculosis-positive; and (4) had a valid Xpert MTB/RIF or Xpert Ultra result for rifampicin resistance, detected or not detected (susceptible).

- *Sensitivity = Xpert MTB/RIF (or Xpert Ultra) rifampicin resistance detected/phenotypic DST or LPA rifampicin-resistant.*
- *Specificity = Xpert MTB/RIF (or Xpert Ultra) rifampicin resistance not detected/phenotypic DST or LPA rifampicin-susceptible.*

For detection of rifampicin resistance, when a study included multiple types of specimens, we based our determination of Xpert Ultra and Xpert MTB/RIF and sensitivity and specificity on all available data in the study, including data for specimens that we did not include in the primary analyses for detection of extrapulmonary tuberculosis. For example, if a study provided data for several specimen types combined (e.g. all tissue specimens) and we could not disaggregate the data for a specific specimen type, we included all data (for all tissue specimens) in the analysis for rifampicin resistance detection. We did this because we did not expect the accuracy of Xpert Ultra or Xpert MTB/RIF for rifampicin resistance to vary by specimen type. We used the bivariate random-effects model to estimate pooled sensitivity and specificity.

We estimated all models using a Bayesian approach with low-information prior distributions using OpenBUGS software (Version 3.2.3)[77], along with R statistical software. Under the Bayesian approach, all unknown parameters must be provided a prior distribution that defines the range of possible values of the parameter and the weight of each of those values, based on information external to the data. To allow observed data to dominate the final results, we chose to use low-information prior distributions. We defined prior distributions on the log-odds scale over the pooled sensitivity and specificity parameters, their corresponding between-study standard deviations, and the correlation between the sensitivities and specificities across studies. For the pooled log odds of the sensitivity or the pooled log odds of the specificity, we used a normal prior distribution with mean 0 and a wide variance of 4 (or a precision of 0.25). This corresponds to a roughly uniform distribution over the pooled sensitivity and pooled specificity on the probability scale. For the between-study precision, we

used a gamma distribution with a shape parameter of 2 and a rate parameter of 0.5. This corresponds to a 95% prior CrI for the between-study standard deviation in the log odds of sensitivity or the log odds of specificity ranging from roughly 0.29 to 1.44, corresponding to moderate to high values of between-study heterogeneity. Covariance terms followed a uniform prior distribution whose upper and lower limits were determined by the sensitivity of the two tests. The OpenBUGS model used appears in Appendix . It is known that meta-analysis models can be sensitive to the choice of prior distributions over between-study standard deviation parameters. We therefore carried out sensitivity analyses and considered alternative prior distributions that are less informative, allowing a wider range of possible values. To study the sensitivity of all results to the choice of prior distributions given above, we considered alternative prior distributions that were less informative, allowing a wider range of possible values. We increased the variance of the normal distributions over the pooled log odds of sensitivity or specificity to 100. We used a uniform prior distribution ranging from 0 to 3 over the between-study standard deviation on the log odds scale (see programme in Appendix). We noted no appreciable change in pooled accuracy parameters but found that the posterior CrIs and prediction intervals were slightly wider, as expected.

We combined information from the prior distribution with the likelihood of the observed data, in accordance with Bayes' theorem, using the OpenBUGS programme, which provides a sample from the posterior distribution of each unknown parameter. We were particularly interested in the pooled sensitivity and specificity of Xpert and between-study variance in the sensitivity and specificity of Xpert on the log-odds scale. Using a sample from the posterior distribution, we calculated various descriptive statistics of interest. We estimated the median pooled sensitivity and specificity and their 95% CrI. The median or the 50% quantile is the value below which 50% of the posterior sample lies. We report the median because the posterior distributions of some parameters may be skewed and the median would be considered a better point estimate of the unknown parameter than the mean in such cases. The 95% CrI is the Bayesian equivalent of the classical (frequentist) 95% confidence interval (CI) (we will indicate 95% CI for individual study estimates and 95% CrI for pooled study estimates as appropriate). The 95% CrI may be interpreted as an interval that has a 95% probability of capturing the true value of the unknown parameter, given observed data and prior information. We prepared summary receiver operating characteristic (SROC) curves for each meta-analysis model, using the methods described in Harbord 2007.

We also determined the predicted sensitivity and specificity of Xpert MTB RIF and Xpert Ultra and their 95% CrIs. Predicted values represent our best guess for sensitivity and specificity in a future study and will be close to the pooled estimates. However, their CrIs may be different. If there is no heterogeneity at all between studies, the CrI around the predicted estimate will be the same as the CrI around the pooled estimate. On the other hand, if considerable heterogeneity is observed between studies, the CrI around the predicted estimate will be much wider than the CI around the pooled estimate.

In addition, we performed latent class analysis for three forms of extrapulmonary tuberculosis: tuberculous meningitis, pleural tuberculosis, and lymph node tuberculosis, using data from the two-by-two tables comparing the index test to culture as a reference standard. Latent class analysis is a statistical modelling technique that allows estimation of test accuracy in the absence of an adequate reference standard to define the presence or absence of disease [78]. The latent class meta-analysis model expands the traditional meta-analysis model in two ways: (1) we added parameters for the sensitivity and specificity of culture; and (2) we added covariance terms to adjust for the dependence between the index test and culture among disease-positive and disease-negative participants in each study. We used hierarchical prior distributions over the logit sensitivity and logit specificity of culture. In other words, we assumed that the logit sensitivities in the individual studies come from a common probability distribution whose mean is the pooled mean logit sensitivity of culture and whose standard deviation is the between-study standard deviation. Likewise for the specificities. We used the same low-information prior distributions over the pooled logit mean and between-study standard deviation parameters as we had for the corresponding parameters for the index test. We used uniform prior distributions for covariance terms over their ranges, which are determined by the sensitivities and the specificities of the two tests in each study (see Appendix for the OpenBUGS model). We found that we did not need to augment observed data with prior information from other sources for most models. However, in a post hoc analysis Xpert MTB/RIF in lymph node aspirate in which we suspected a systematic bias in the performance of culture, we used informative prior distributions over the specificity of culture (ranging from 99% to 100%) and the specificity of Xpert MTB/RIF (ranging from 98% to 100%) (see Appendix). We added the SROC plots of the latent class meta-analyses to the

SROC plots resulting from the models in which culture was treated as a perfect test, so they could be compared.

Based on work evaluating Xpert MTB/RIF for childhood tuberculosis (Schumacher 2016), we anticipated that latent class meta-analyses would lead to a decrease in the estimated pooled sensitivity of Xpert Ultra and Xpert MTB/RIF and an increase in the estimated pooled specificity of Xpert Ultra and Xpert MTB/RIF compared with the primary analyses. In other words, this method should help to correct the biases in Xpert Ultra and Xpert MTB/RIF sensitivity and specificity resulting from treating culture as a perfect reference standard, which we detailed earlier in the section on the reference standard.

Approach to inconclusive index test results

The proportion of inconclusive (non-determinate) rate for detection of pulmonary tuberculosis is the number of tests classified as ‘invalid', ‘error', or ‘no result' divided by the total number of index tests performed. The proportion of inconclusive (indeterminate) rate for detection of rifampicin resistance is the number of tests classified as ‘MTB DETECTED; Rif (rifampicin) resistance INDETERMINATE' divided by the total number of index test-positive results. For Xpert Ultra, we determined the proportion of inconclusive index test results = number of inconclusive test results divided by the total number of tests. In our previous review, we used a Bayesian hierarchical model for a single proportion to estimate the pooled proportion of inconclusive MTB/RIF test results[60]. We reported these findings again in this review update. As we found very few inconclusive results reported, we excluded these results from the quantitative analysis.

Investigations of heterogeneity

Initially, we investigated heterogeneity through visual examination of forest plots of sensitivities and specificities and through visual examination of the ROC space of the raw data. When data allowed, we evaluated potential sources of heterogeneity using subgroup analyses and bivariate meta-regression. We included the following covariates.

- HIV status.
- For tuberculous meningitis, concentration step used for preparing specimen (yes or no).
- CSF specimen volume used for Xpert MTB/RIF or Xpert Ultra testing.

We had planned to investigate smear status, history of tuberculosis, and whether WHO standard procedures for preparing tissue specimens were followed. However, we had insufficient data to do this. The impact of the prevalence of extrapulmonary tuberculosis on sensitivity and specificity is an important consideration. In a post hoc meta-regression analysis, for Xpert MTB/RIF we explored this question for CSF, pleural fluid, and lymph node aspirate. For Xpert Ultra we explored this question for CSF. We did not conduct other analyses, owing to an insufficient number of studies. For detection of rifampicin resistance, owing to a small number of studies, we could not assess the impact of prevalence of rifampicin resistance on accuracy estimates.

Nontuberculous mycobacteria

Nontuberculous mycobacteria (NTM), such as *M avium* complex and *M intracellulare*, constitute a multi-species group of human pathogens that are ubiquitous in water and soil. NTM can cause severe diseases that share clinical signs with tuberculosis but are treated differently. People living with HIV with severe immunosuppression are particularly vulnerable to infections caused by NTM[79]. Previous studies have shown that Xpert does not cross-react with other mycobacterial species [80,81]. In our original review, we summarized data for NTM separately by determining the percentage of false-positive Xpert MTB/RIF results in specimens that grew NTMs[60]. In this updated review, we therefore summarize data for NTM only for Xpert Ultra.

Sensitivity analyses

For Xpert Ultra testing in CSF, we performed sensitivity analyses to explore whether the overall findings were robust to potentially influential decisions. We did this by limiting inclusion in the meta-analysis to the following.

- Studies that used consecutive or random selection of participants.
- Studies in which the reference standard results were interpreted without knowledge of the index test results.
- Studies that included only one specimen per participant.

For Xpert Ultra, in CSF, we also planned to perform a sensitivity analysis by limiting studies to those that included only untreated participants. However, we were unable to confirm that studies met this criterion. We planned similar sensitivity analyses for pleural fluid and lymph node aspirate, but these

analyses were not carried out owing to an insufficient number of studies. For all other specimen types, we had an insufficient number of studies for sensitivity analyses.

For Xpert MTB/RIF, in the original review we performed sensitivity analyses by type of extrapulmonary specimen and found that for most analyses, the sensitivity analyses made little difference to any of these findings[60]. However, for Xpert MTB/RIF in CSF, in comparison with all studies, (sensitivity of 71.1% (60.9 to 80.4), and specificity of 98.0% (97.0 to 98.8)), studies that evaluated only one specimen per participant had lower pooled sensitivity at 63.5% (47.6 to 76.3) and lower pooled specificity at 96.1% (94.2 to 97.4).

Assessment of reporting bias

We did not perform a formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been helpful for diagnostic test accuracy studies [82].

Summary of findings and assessment of the certainty of the evidence

We assessed the certainty of evidence using the GRADE approach for diagnostic studies[83–86]. As recommended, we rated the certainty of evidence as either high (not downgraded), moderate (downgraded by one level), low (downgraded by two levels), or very low (downgraded by more than two levels) based on five domains: risk of bias, indirectness, inconsistency, imprecision, and publication bias. For each outcome, the certainty of evidence started as high when there high-quality studies (cross-sectional or cohort studies) that enrolled participants with diagnostic uncertainty. If we found a reason for downgrading, we used our judgement to classify the reason as either serious (downgraded by one level) or very serious (downgraded by two levels).Two review authors discussed judgments and applied GRADE in the following way[87,88].

- *Assessment of risk of bias*. We used QUADAS-2 to assess risk of bias.
- *Indirectness*. We assessed indirectness in relation to the population (including disease spectrum), setting, interventions, and outcomes (accuracy measures). We also used prevalence as a guide to whether there was indirectness in the population.
- *Inconsistency*. GRADE recommends downgrading for unexplained inconsistency in sensitivity and specificity estimates. We carried out prespecified analyses to investigate potential sources of heterogeneity and downgraded when we could not explain inconsistency in the accuracy estimates.

- *Imprecision*. We considered a precise estimate to be one that would allow a clinically meaningful decision. We considered the width of the CrI (or CI) and asked, "Would we make a different decision if the lower or upper boundary of the CrI (or CI) represented the truth?" In addition, we worked out projected ranges for TP, FN, TN, and FP for a given prevalence of tuberculosis and made judgements on imprecision from these calculations.
- *Publication bias*. We rated publication bias as undetected (not serious) for several reasons: the comprehensiveness of the literature search and extensive outreach to tuberculosis researchers to identify studies; the presence only of studies that produced precise estimates of high accuracy despite small sample size; and our knowledge about studies that were conducted but not published.

For the 'Summary of findings' tables for CSF and pleural fluid, we provide evidence using culture as the reference standard, which is considered the best reference standard for tuberculosis[65]. For lymph node aspirate, we provide evidence using a composite reference because, based on findings from the original review[60], we believe a composite reference standard is preferable for estimating accuracy.

Results

Results of the search

We identified and screened a total of 735 records for inclusion in this review update. Of these, we assessed 142 full-text papers against our inclusion criteria. We excluded 120 papers, mainly for the following reasons: study did not evaluate Xpert Ultra (n = 54); could not extract 2 x 2 values (n = 14); inappropriate reference standard (n = 13); could not extract data by specimen type (n = 12); did not include extrapulmonary specimen (n = 10): duplicate data (n = 4); case-control study (n = 4); index test other than Xpert MTB/RIF or Xpert Ultra (n = 3); study included children (n = 2); screening study (n = 2); case report (n = 1); and fewer than five specimens for a given specimen type (n = 1). From our previous review, we included 47 studies.

Thus, we included 69 unique studies that met the inclusion criteria in this review update.

Sixty-seven studies evaluated Xpert MTB/RIF[89–155]. Eleven studies evaluated Xpert Ultra. Of these 11 studies, nine evaluated both Xpert MTB/RIF and Xpert Ultra [91,99,118,136,139,142,149] and two studies evaluated Xpert Ultra alone [111,152]. All studies but two (one in Spanish: Peñata 2016[112], and one in Turkish: Ozkutuk 2014[115]), were written in English. Figure 1 shows the flow of studies in the review.

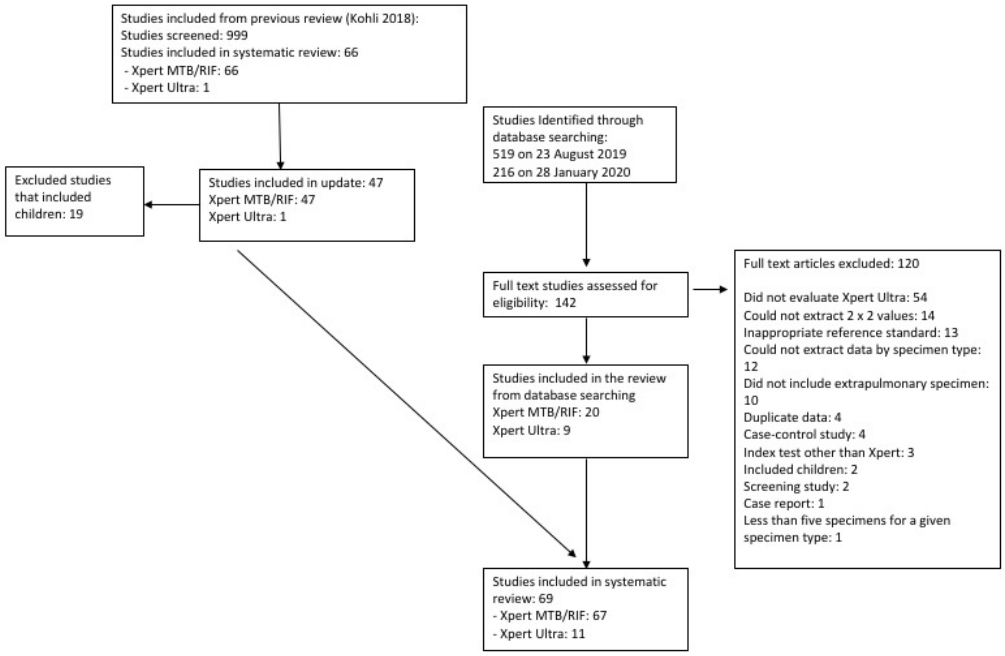


Figure 3.1: PRISMA diagram of included studies

Methodological quality of included studies

Studies evaluating Xpert MTB/RIF and Xpert Ultra for detection of extrapulmonary tuberculosis

Figure 3.2 show risk of bias and applicability concerns for the 69 studies included for tuberculosis detection. Risk of bias and applicability concerns are also presented specifically for studies evaluating Xpert Ultra and Xpert MTB/RIF for tuberculous meningitis, pleural tuberculosis, and lymph node tuberculosis.

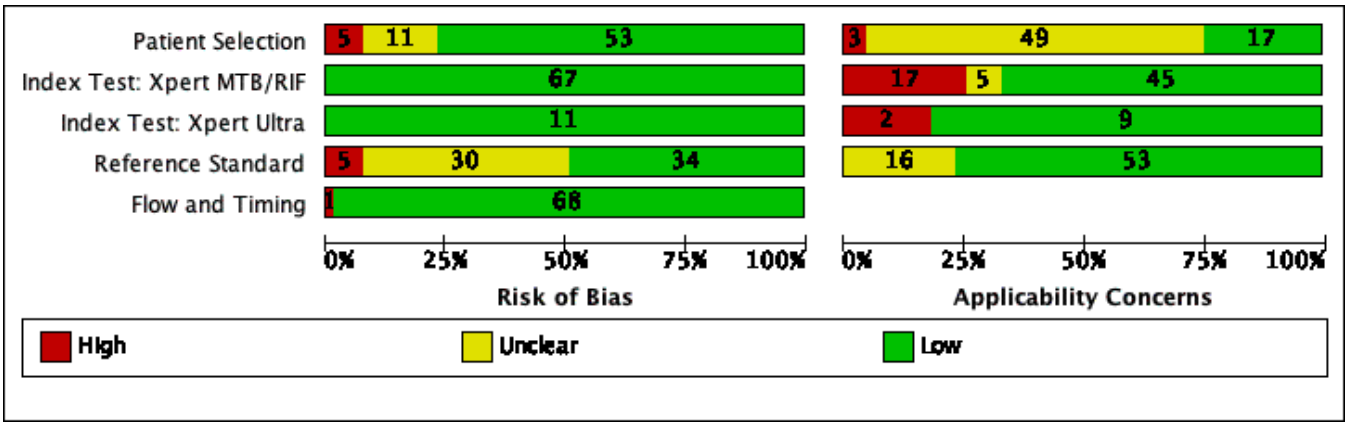


Figure 3.2: Risk of bias and applicability concerns graph for tuberculosis detection

In the patient selection domain, we thought that 53 studies (77%) had low risk of bias, and five studies (7%) had high risk of bias for the following reasons: one study selected participants by convenience[120] and four studies had inappropriate exclusions [95,100,111,135]. We thought that 11 studies (16%) had unclear risk of bias for the following reasons: the manner of patient selection was unclear in ten studies[89,92,105,107–109,123,124,130,151], and it was unclear whether the study avoided inappropriate exclusions: one study [148]. Regarding applicability (patient characteristics and setting), we thought that 17 studies (25%) had low concern because participants were evaluated in local hospitals or primary health settings: three studies [96,105,114], or in the case of tuberculous meningitis, tertiary centres: 14 studies [100,107,113,116,128,136,139,140,142,149–151,154]. Three studies (4%) had high concern because participants were evaluated exclusively as inpatients at a tertiary care centre [130,135,144]; and 52 (75%) studies had unclear or high concern because we could not tell the clinical setting or a high percentage of participants had received prior testing for tuberculosis [152].

In the index test domain, we thought that all studies had low risk of bias because the results of the index tests (Xpert Ultra and Xpert MTB/RIF) are automatically generated, the user is provided with printable test results, and the test threshold is prespecified. Regarding applicability, with respect to Xpert Ultra, we thought that 9/11 studies (82%) had low risk of bias[91,99,111,118,136,139,149,152] and 2/11 studies (18%) had high risk of bias because the index test was not performed according to WHO standard operating procedures [92,142]. With respect to Xpert MTB/RIF, we thought that 45/67 studies (67%) had low concern because at least 75% of the specimen types in these studies were processed according to WHO recommendations; 17/67 studies (25%) had high concern because fewer than 50% of the specimen types in these studies were processed according to WHO recommendations [90,92,95,98,107–109,116,120,121,128,132,133,138,142,144,145]. We thought that 5/67 studies (7%) had unclear concern because either the manner of specimen processing was not reported [140,148,151,154], or only 50% of the specimen types were processed according to WHO recommendations[141].

In the reference standard domain, 34 studies (49%) had low risk of bias because results of the reference standard were interpreted without knowledge of results of the index test and only non-sterile specimens were decontaminated [90,92,93,95,96,98,100,103,104,107,113–116,118–120,122,123,126,128,133,136,137,139–142,144,148–150,152,154]. Five studies (7%) had high risk of bias because results of the reference standard were interpreted with knowledge of results of the index test [89,106,112,129,146]Thirty studies (43%) had unclear risk of bias for

the following reasons: seven studies did not report whether there was blinding of the reference standard [91,111,121,134,151]); 21 studies decontaminated specimens generally considered to be sterile [94,97,99,102,103,108–110,117,124,125,127,130–132,135,138,143,145,147,153]; and two studies did not report blinding and decontaminated specimens generally considered to be sterile [105,155].

Regarding applicability of the reference standard, we thought that 53 studies (77%) had low concern because these studies performed a test to identify *M tuberculosis* species (speciation) and 16 studies (23%) had unclear concern because we could not tell whether the study performed speciation [91,92,95,96,105,112,121,126,134,137,140–142,148,151].

In the flow and timing domain, we considered almost all studies to have low risk of bias, noting that all participants were accounted for in the analysis. One study included fewer than 50% of eligible participants in the analysis (Trajman 2014).

Studies evaluating Xpert Ultra for detection of rifampicin resistance

Figure 3.3 and Figure 3.4 show risk-of-bias and applicability concerns for each of the five studies included for rifampicin resistance detection.

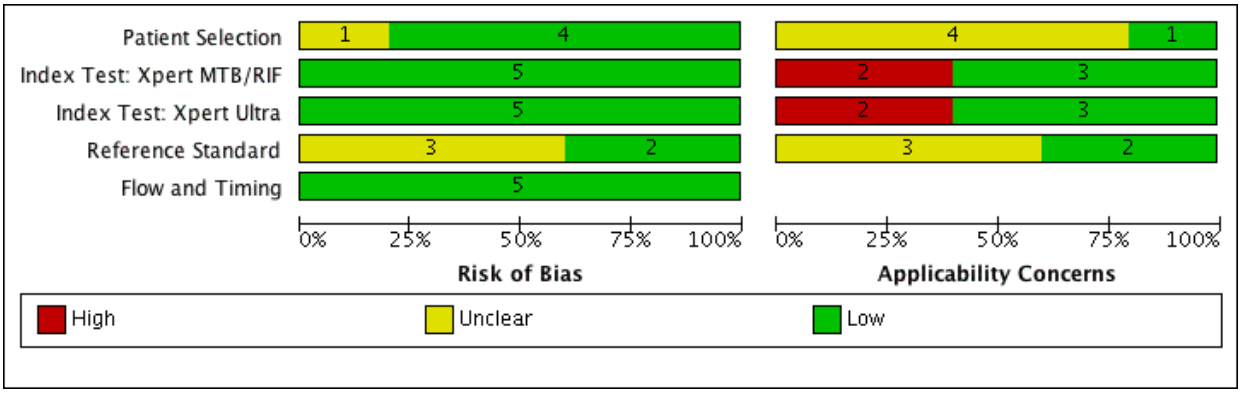


Figure 3.3: Risk of bias and applicability concerns graph for rifampicin resistance detection in comparative studies of Xpert Ultra and Xpert MTB/RIF

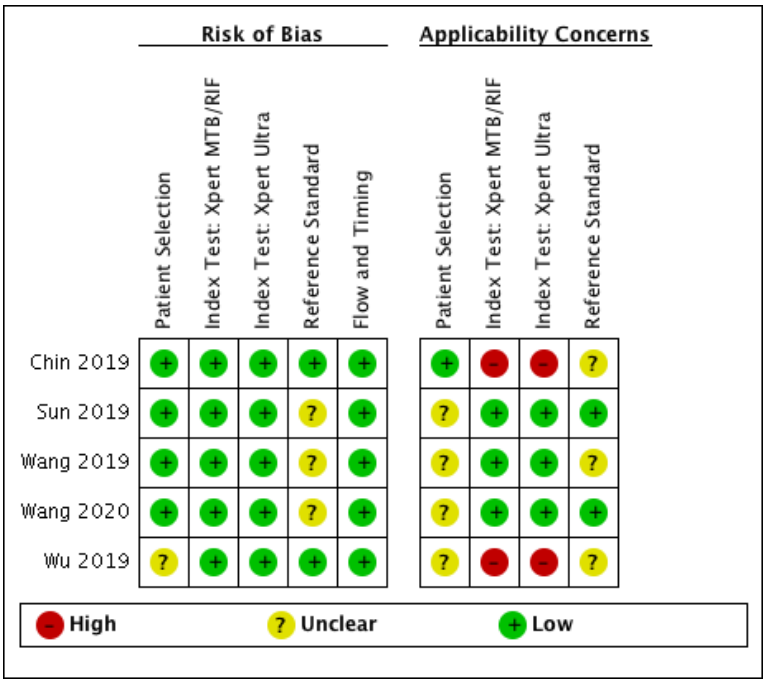


Figure 3.4: Risk of bias and applicability concerns summary for rifampicin resistance detection in each comparative studies of Xpert Ultra and Xpert MTB/RIF

In the patient selection domain, we thought that four studies (80%) had low risk of bias and one study (20%) had unclear risk of bias as the manner of patient selection was unclear (Figure 3.4). We thought that one study (20%) had low concern because participants were evaluated exclusively as inpatients at a tertiary care centre and four studies (80%) had unclear concern because we could not tell the details of the clinical setting (Figure 3.4).

In the index test domain, we thought that all studies had low risk of bias because the results of the index tests are automatically generated, the user is provided with printable test results, and the test threshold is prespecified. For applicability, with respect to Xpert Ultra, we thought that three studies (60%) had low concern because at least 75% of the specimen types in these studies were processed according to WHO recommendations; two studies (40%) had high concern because fewer than 50% of the specimen types in these studies were processed according to WHO recommendations (Figure 3.4).

In the reference standard domain, two studies (40%) had low risk of bias because results of the reference standard were interpreted without knowledge of results of the index test and only non-sterile

specimens were decontaminated[92,142]. Three studies (60%) had unclear risk of bias as it was unclear whether blinding of the reference standard was performe. For applicability of the reference standard, we thought that all studies had low concern because detection of rifampicin resistance occurs only when the *M tuberculosis* target is present within the specimen.

In the flow and timing domain, we considered all studies to have low risk of bias, noting that all participants were accounted for in the analysis.

Findings from included studies

The 69 studies were conducted in 28 different countries. Most of the studies were conducted in China (n = 10), India (n = 13), South Africa (n = 10), and Uganda (n = 6). Seven studies exclusively or mainly included HIV-positive participants [93,133,139,149–151,155]. Most studies performed the index tests and culture on the same specimen type, except for one study in which Xpert MTB/RIF was performed on blood and culture was performed on sputum [133]. Most studies did not report the exact number of cultures used to confirm a diagnosis of tuberculosis, but it is likely that many studies used a single culture.

I. Detection of extrapulmonary tuberculosis

Xpert Ultra: of the 11 studies, the number evaluating different specimens was as follows: tuberculous meningitis (CSF) six studies; pleural tuberculosis (pleural fluid) four studies; lymph node tuberculosis (lymph node aspirate) one study; genitourinary tuberculosis (urine) one study; bone or joint tuberculosis (bone or joint aspirate) two studies; and peritoneal tuberculosis (peritoneal fluid) one study.

Xpert MTB/RIF: of the 67 studies, the number of studies evaluating different specimens was as follows: tuberculous meningitis (CSF) 33 studies; pleural tuberculosis (fluid) 27 studies; lymph node tuberculosis (aspirate 15 studies, biopsy 11 studies); genitourinary tuberculosis (urine) 15 studies; bone or joint tuberculosis (aspirate 12 studies, tissue 3 studies); peritoneal tuberculosis (fluid 17 studies, tissue 1 study); pericardial tuberculosis (fluid 14 studies, tissue 2 studies); and disseminated tuberculosis (blood 2 studies). Several studies included more than one type of specimen.

Table 3.1 presents Xpert Ultra and Xpert MTB/RIF pooled sensitivity and specificity estimates and predictive values by reference standard for all forms of extrapulmonary tuberculosis and specimen types included in the review.

Table 3.1: Xpert Ultra and Xpert MTB/RIF accuracy estimates for different specimen types

Type of specimen	Test	Reference standard	Number studies (participants)	Number (%) with TB or rifampicin resistance	Pooled sensitivity (95% CrI)	Pooled specificity (95% CrI)	Positive predictive value (95% CrI)	Negative predictive value (95% CrI)
CSF	Xpert Ultra	culture	6 (475)	89 (18.7)	89.4% (79.1 to 95.6)	91.2% (83.2 to 95.7)	53.0% (36.6 to 69.6)	98.7% (97.5 to 99.5)
CSF	Xpert Ultra	composite	4 (496)	160 (32.2)	62.7% (45.7 to 77.0)	99.1% (96.6 to 99.9)	87.9% (65.5 to 99.0)	96.0% (94.2 to 97.5)
CSF	Xpert MTB/RIF	culture	30 (3395)	571 (16.8)	71.1% (62.8 to 79.1)	96.9% (95.4 to 98.0)	71.8% (62.3 to 80.7)	96.8% (95.9 to 97.7)
CSF	Xpert MTB/RIF	composite	14 (2203)	862 (39.1)	42.3% (32.1 to 52.8)	99.8% (99.3 to 100.0)	96.3% (87.2 to 100.0)	94.0% (93.0 to 95.0)
CSF	Ultra, direct comparison	culture	5 (471)	86 (18.3)	89.0% (77.9 to 95.2)	91.0% (82.7 to 95.6)	52.2% (35.6 to 69.0)	98.7% (97.3 to 99.4)
CSF	Xpert MTB/RIF, direct comparison	culture	5 (471)	87 (18.5)	62.2% (43.7 to 78.1)	96.8% (93.4 to 98.6)	68.4% (49.0 to 83.6)	95.8% (93.9 to 97.5)
Pleural fluid	Xpert Ultra	culture	4 (398)	158 (39.7)	75.0% (58.0 to 86.4)	87.0% (63.1 to 97.9)	38.8% (17.9 to 79.5)	96.9% (94.5 to 98.3)

Pleural fluid	Xpert MTB/RIF	culture	25 (3065)	644 (21.0)	49.5% (39.8 to 59.9)	98.9% (97.6 to 99.7)	83.2% (68.9 to 94.6)	94.6% (93.7 to 95.7)
Pleural fluid	Xpert MTB/RIF	composite	10 (1024)	616 (60.1)	18.9% (11.5 to 27.9)	99.3% (98.1 to 99.8)	73.6% (49.2 to 91.2)	91.7% (91.0 to 92.5)
Lymph node aspirate	Xpert MTB/RIF	culture	14 (1588)	627 (39.5)	88.9% (82.7 to 93.6)	86.2% (78.0 to 92.3)	41.7% (31.4 to 55.5)	98.6% (97.9 to 99.2)
Lymph node aspirate	Xpert MTB/RIF	composite	4 (679)	377 (55.5)	81.6% (61.9 to 93.3)	96.4% (91.3 to 98.6)	71.0% (51.1 to 86.1)	97.9% (95.8 to 99.2)
Lymph node biopsy	Xpert MTB/RIF	culture	11 (786)	220 (28.0)	82.4% (73.5 to 89.7)	80.3% (60.3 to 91.5)	31.6% (18.7 to 51.8)	97.6% (96.2 to 98.6)
Urine	Xpert MTB/RIF	culture	9 (943)	72 (7.6)	85.9% (71.4 to 94.3)	98.1% (93.1 to 99.7)	83.0% (58.3 to 96.7)	98.4% (96.9 to 99.4)
Bone or joint aspirate	Xpert MTB/RIF	culture	6 (471)	110 (23.4)	97.9% (93.1 to 99.6)	97.4% (80.2 to 100.0)	80.7% (35.4 to 99.5)	99.8% (99.2 to 100.0)
Peritoneal fluid	Xpert MTB/RIF	culture	13 (580)	94 (16.2)	59.1% (42.1 to 76.2)	97.6% (95.4 to 98.9)	73.0% (58.2 to 86.2)	95.5% (93.8 to 97.4)
Pericardial fluid	Xpert MTB/RIF	culture	5 (181)	57 (31.5)	61.4% (32.4 to 82.4)	89.7% (74.9 to 99.0)	39.4% (18.3 to 88.0)	95.4% (92.1 to 97.9)
Rifampicin resistance	Xpert Ultra	DSTor LPA	4 (129)	24 (18.6)	100.0% (95.1 to 100.0)	100.0% (99.0 to 100.0)	99.9% (91.7 to 100.0)	100.0% (99.5 to 100.0)
Rifampicin resistance	Xpert MTB/RIF	DSTor LPA	19 (970)	148 (15.3)	96.5% (91.9 to 98.8)	99.1% (98.0 to 99.7)	92.0% (84.3 to 97.3)	99.6% (99.1 to 99.9)

A: Tuberculous meningitis

Xpert Ultra

Culture reference standard

Six studies evaluated Xpert Ultra in cerebrospinal fluid (CSF) specimens against culture [91,111,136,139,142,149]. Xpert Ultra sensitivity ranged from 80% to 100% and specificity ranged from 50% to 100% (Figure 10). Chin 2019 reported the lowest specificity (50%). In this study, the investigators inoculated uncentrifuged CSF which could have led to lower culture positivity, thus resulting in a higher number of false positives. Perez-Risco 2018 (specificity 100%) contributed only one participant to this analysis. In CSF, Xpert Ultra pooled sensitivity and specificity (95% Crl) against culture were 89.4% (79.1 to 95.6) and 91.2% (83.2 to 95.7), (6 studies; 475 participants, 89 (18.7%) with tuberculosis); Table 3.1.

Composite reference standard

In CSF, Xpert Ultra pooled sensitivity and specificity against a composite reference standard were 62.7% (45.7 to 77.0) and 99.1% (96.6 to 99.9), (4 studies; 496 participants); Table 3.1, Figure 3.5.

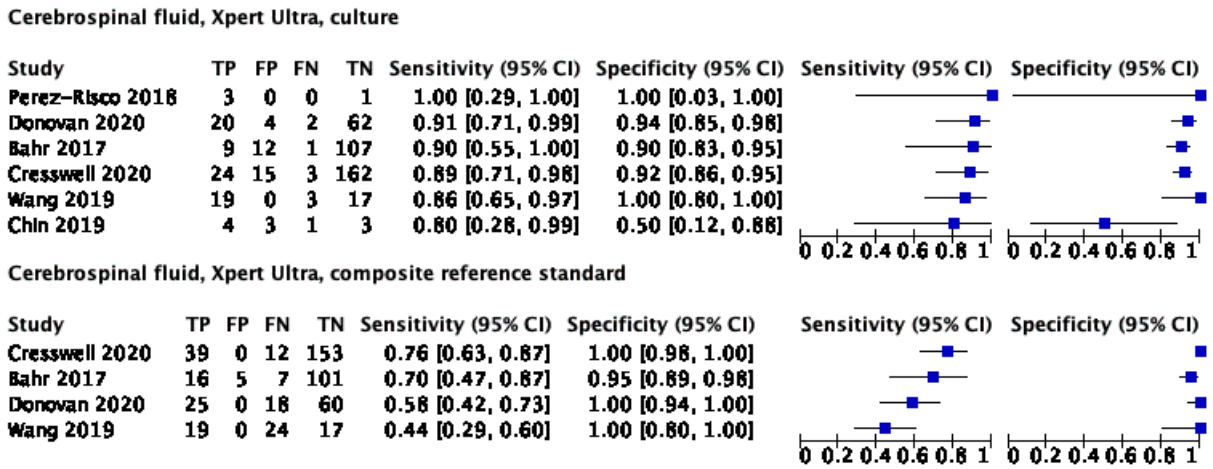


Figure 3.5: Forest plots of Xpert Ultra sensitivity and specificity in cerebrospinal fluid by reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

Latent class meta-analysis

We had insufficient data to obtain robust parameter estimates using the latent class model for Xpert Ultra in CSF.

Xpert MTB/RIF

Culture reference standard

Thirty-three studies evaluated Xpert MTB/RIF in CSF specimens against culture [89–91,94,95,98,100,103,106,107,110,112,113,115–117,120,125,127,129,136,139,140,142,145,146,149–151,153,154]. Xpert MTB/RIF sensitivity ranged from 0% to 100% and specificity ranged from 78% to 100% (Figure 6). For sensitivity, we thought that differences in CSF volume and processing could partly explain the heterogeneity. Three studies [106,127,153] did not contribute data to the meta-analysis because sensitivity was not estimable.

In CSF, Xpert MTB/RIF pooled sensitivity and specificity (95% Crl) against culture were 71.1% (62.8 to 79.1) and 96.9% (95.4 to 98.0), respectively (30 studies; 3395 participants, 571 (16.8%) with tuberculosis); Table 3.1, Figure 3.6.

Composite reference standard

In CSF, Xpert MTB/RIF pooled sensitivity and specificity against a composite reference standard were 42.3% (32.1 to 52.8) and 99.8% (99.3 to 100.0), (14 studies; 2203 participants); Table 3.1, Figure 3.6.

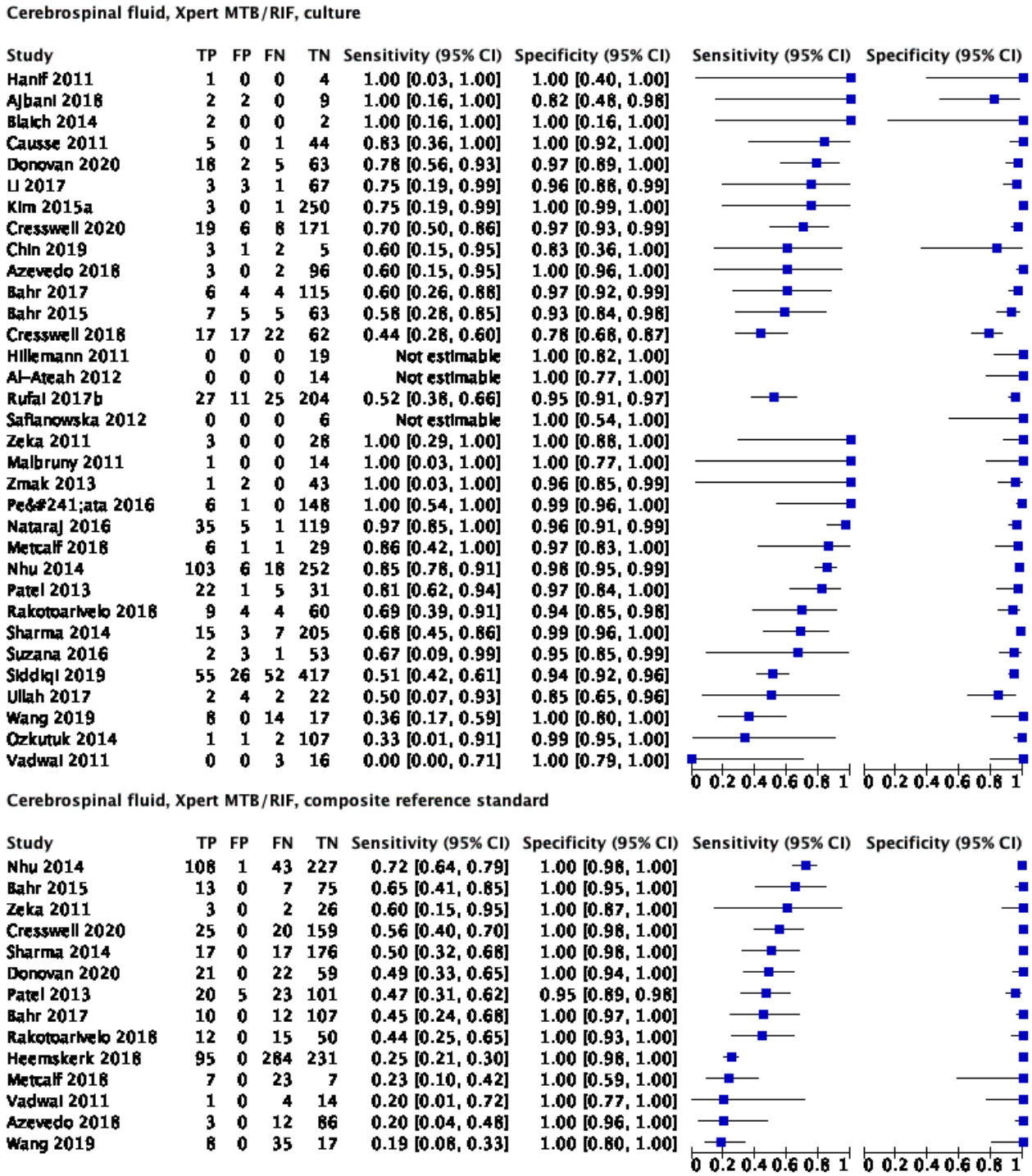


Figure 3.6: Forest plots of Xpert MTB/RIF sensitivity and specificity in cerebrospinal fluid by reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

Latent class meta-analysis

Based on the latent class meta-analysis model, Xpert MTB/RIF pooled sensitivity and specificity (95% Crl) for tuberculous meningitis were 74.7% (65.5 to 84.0) and 99.5% (99.1 to 99.7) (30 studies; 3395 participants); Table 2. The pooled sensitivity of culture at 80.8% (72.5 to 88.5) was estimated to be lower than 100%. The pooled specificity of culture was estimated to be 99.2% (98.7 to 99.5); Table 3.2.

Table 3.2: Latent class analyses

Form of extrapulmonary tuberculosis, type of specimen	Number of studies (participants)	Culture-confirmed tuberculosis (%)	Pooled sensitivity (95% CrI)	Pooled specificity (95% CrI)	Positive predictive value (95% CrI)	Negative predictive value (95% CrI)
Accuracy estimates of Xpert MTB/RIF						
Tuberculous meningitis, cerebrospinal fluid	30 (3395)	571 (16.8)	74.7% (65.5 to 84.0)	99.5% (99.1 to 99.7)	94.5% (89.7 to 96.9)	97.3% (96.3 to 98.3)
Pleural tuberculosis, fluid	25 (3065)	644 (21.0)	53.1% (42.8 to 64.1)	99.6% (99.3 to 99.8)	93.7% (89.5 to 96.5)	95.0% (94.0 to 96.2)
Lymph node tuberculosis, aspirate	14 (1588)	627 (39.5)	91.5% (85.2 to 95.9)	99.5% (99.1 to 99.7)	95.2% (91.4 to 97.5)	99.1% (98.4 to 99.5)
Accuracy estimates of culture						
Tuberculous meningitis, cerebrospinal fluid	30 (3395)	571 (16.8)	80.8% (72.5 to 88.5)	99.2% (98.7 to 99.5)	91.9% (86.9 to 95.1)	97.9% (97.0 to 98.7)
Pleural tuberculosis, fluid	25 (3065)	644 (21.0)	89.5% (80.5 to 96.3)	99.0% (98.2 to 99.5)	90.8% (84.2 to 94.7)	98.8% (97.9 to 99.6)
Lymph node tuberculosis, aspirate	14 (1588)	627 (39.5)	82.9% (69.9 to 91.8)	98.8% (97.8 to 99.4)	88.7% (80.1 to 94.0)	98.1% (96.7 to 99.1)

Xpert Ultra versus Xpert MTB/RIF

In comparative accuracy studies evaluating both index tests, Xpert Ultra pooled sensitivity and specificity (95% CrI) against culture were 89.0% (77.9 to 95.2) and 91.0% (82.7 to 95.6) and Xpert MTB/RIF pooled sensitivity and specificity were 62.2% (43.7 to 78.1) and 96.8% (93.4 to 98.6), (5 studies; 471 participants), direct comparison, Table 3.1; Figure 3.7; Figure 3.8. For CSF, the difference

between the sensitivities of Xpert Ultra and Xpert MTB/RIF was 26.2% (9.1 to 44.4). We estimated the probability that the pooled sensitivity of Xpert Ultra exceeds that of Xpert MTB/RIF was 0.997. The difference between the specificities of Xpert Ultra and Xpert MTB/RIF was −5.6% (−12.9 to −0.1). We estimated the probability that the pooled specificity of Xpert Ultra was less than that of Xpert MTB/RIF was 0.978; Table 3.3.

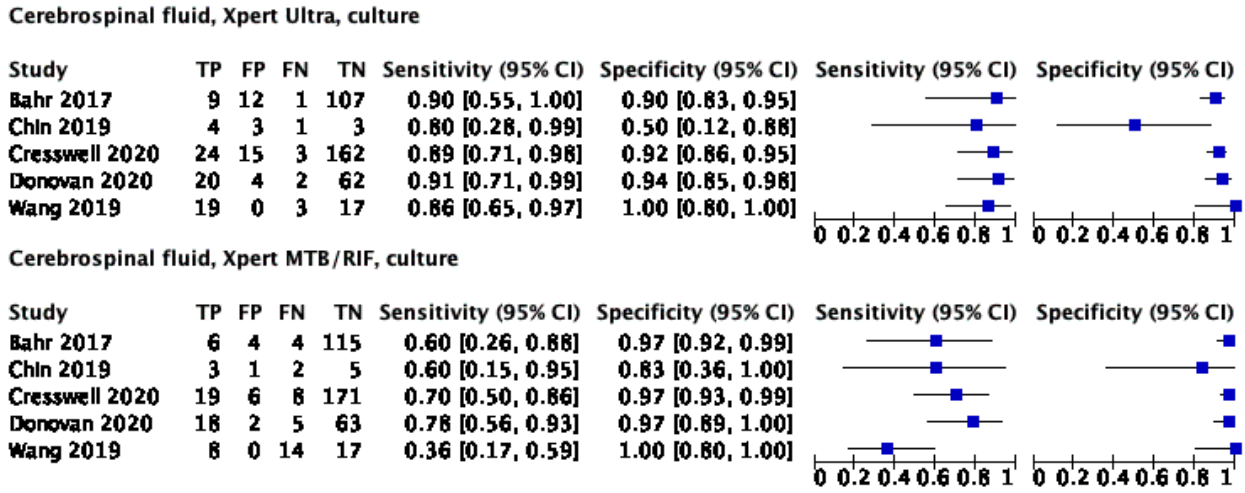


Figure 3.7: Forest plots of Xpert Ultra and Xpert MTB/RIF sensitivity and specificity in cerebrospinal fluid, comparative studies.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

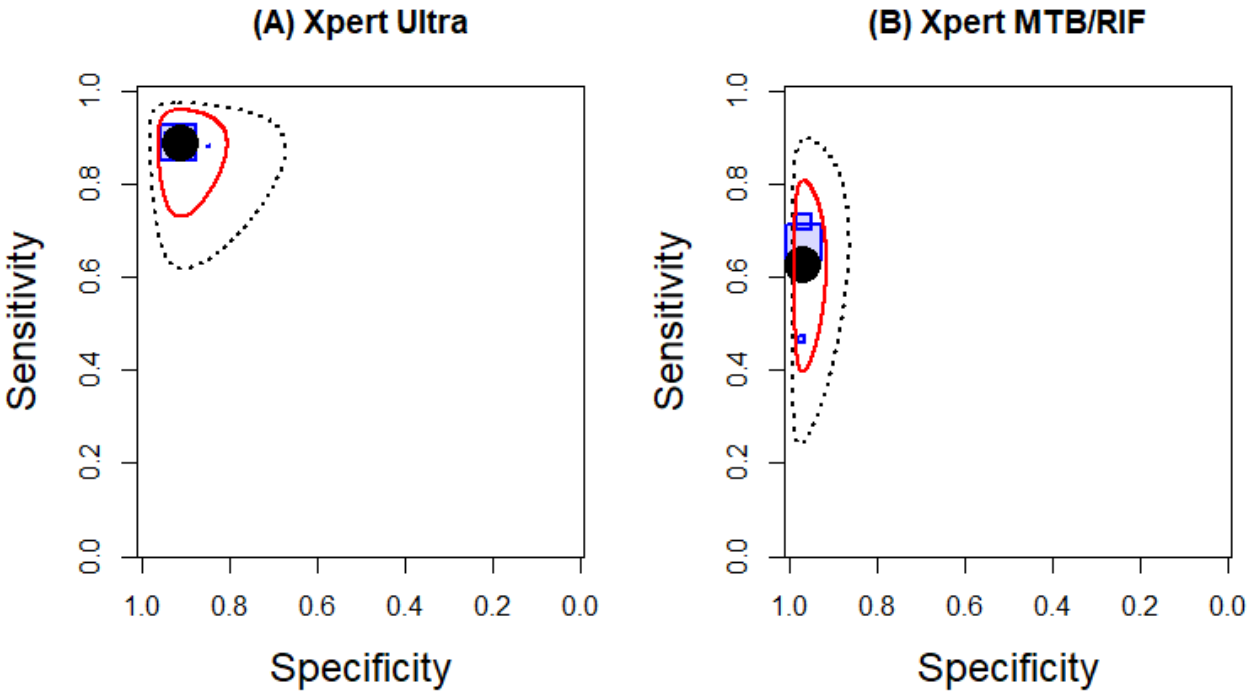


Figure 3.8: Summary plots of the sensitivity and specificity of Xpert Ultra (A) (5 studies) and Xpert MTB/RIF (B) (5 studies) in cerebrospinal fluid for detection of tuberculous meningitis.

Each individual study is represented by a shaded square. The size of the square is proportional to the sample size of the study such that larger studies are represented by larger squares. The filled circle is the median pooled estimate for sensitivity and specificity. The solid curves represent the 95% credible region around the summary estimate; the dashed curves represent the 95% prediction region.

Table 3.3: Accuracy of Xpert Ultra versus Xpert MTB/RIF in cerebrospinal fluid

Detection of tuberculosis in CSF				
Test (studies, participants)	**Xpert Ultra (5, 471)**	**Xpert MTB/RIF (5, 471)**	**Difference (Xpert Ultra minus Xpert MTB/RIF)**	**Probability (Xpert Ultra minus Xpert MTB/RIF)**
Sensitivity	89.0% (77.9 to 95.2)	62.2% (43.7 to 78.1)	26.2% (9.1 to 44.4)	0.997
Specificity	91.0% (82.7 to 95.6)	96.8% (93.4 to 98.6)	−5.6% (−12.9 to −0.1)	0.022

Investigations of heterogeneity

Xpert Ultra versus Xpert MTB/RIF testing in people living with HIV

We identified two studies that directly compared Xpert Ultra and Xpert MTB/RIF, both against culture, in people living with HIV. Sensitivity (95% CI) was 90% (55 to 100) and 89% (71 to 98) for Xpert Ultra and 60% (26 to 88) and 70% (50 to 86) for Xpert MTB/RIF. Specificity (95% CI) was 90% (83 to 95) and 92% (86 to 95) for Xpert Ultra and 97% (95% CI 92 to 99) and 97% (93 to 99) for Xpert MTB/RIF.

Specimen concentration

Xpert Ultra

We found that concentrating CSF improved both Xpert Ultra sensitivity and specificity. Xpert Ultra pooled sensitivity in concentrated specimens was 90.5% (76.7 to 97.0) (3 studies; 421 participants) versus 88.4% (67.8 to 97.5) (3 studies; 54 participants) in unconcentrated specimens. Xpert Ultra pooled specificity in concentrated specimens was 91.9% (84.5 to 96.1) versus 88.6% (58.4 to 99.0) in unconcentrated specimens; Table 3.4. The probability that Xpert Ultra sensitivity and specificity are higher with concentrated CSF compared to unconcentrated CSF were 0.630 and 0.653, respectively.

Xpert MTB/RIF

We found that concentrating CSF improved both Xpert MTB/RIF sensitivity and specificity. Xpert MTB/RIF pooled sensitivity in concentrated specimens was 77.6% (67.2 to 85.9) (14, 2279 participants) versus 59.4% (48.3 to 70.5) (17,1123 participants) in unconcentrated specimens. Xpert MTB/RIF pooled specificity in concentrated specimens was 97.4% (96.1 to 98.4) versus 96.8% (94.0 to 98.7) in unconcentrated specimens, Table 4. The probability that Xpert MTB/RIF sensitivity and specificity are higher with concentrated CSF compared to unconcentrated CSF were 0.989 and 0.696, respectively.

Table 3.4: Impact of concentrating cerebrospinal fluid on Xpert Ultra and Xpert MTB/RIF sensitivity and specificity.

Covariate (number of studies, participants)	Pooled sensitivity (95% CrI)	Pooled specificity (95% CrI)
Concentration step, Xpert Ultra		
Concentrated specimen (3, 421)	90.5% (76.7 to 97.0)	91.9% (84.5 to 96.1)
Unconcentrated specimen (3, 54)	88.4% (67.8 to 97.5)	88.6% (58.4 to 99.0)
Difference (concentrated minus unconcentrated)	2.6% (−13.9 to 24.1)	3.4% (−9.5 to 32.8)
Probability (concentrated minus unconcentrated)	0.630	0.653
Concentration step, Xpert MTB/RIF		
Concentrated specimen (14, 2279)	77.6% (67.2 to 85.9)	97.4% (96.1 to 98.4)
Unconcentrated specimen (17, 1123)	59.4% (48.3 to 70.5)	96.8% (94.0 to 98.7)
Difference (concentrated minus unconcentrated)	18.4% (2.8 to 32.1)	0.6% (−1.7 to 3.6)
Probability (concentrated minus unconcentrated)	0.989	0.696

Cerebrospinal fluid collection volumes

Xpert Ultra

Two studies reported the volume of CSF collected for Xpert Ultra testing, 3 mL in both studies. Sensitivities were similar: 90% (55 to 100) in Bahr 2017 and 89% (71 to 98) in Cresswell 2020[139]. Specificities were also similar 90% (83 to 95) in Bahr 2017 and 92% (86 to 95) in Cresswell 2020[139].

Impact of tuberculosis prevalence on sensitivity and specificity

For Xpert Ultra, we found lower sensitivity in settings with higher tuberculosis prevalence (threshold 30%) than in those with lower tuberculosis prevalence: pooled sensitivity (95% CrI) of 88.3% (68.3 to 97.0) versus 90.8% (77.3 to 96.9). We found lower specificity in settings with higher tuberculosis prevalence than in those with lower tuberculosis prevalence: pooled specificity of 88.0% (64.3 to 97.9) versus 91.9% (82.5 to 96.6). In both analyses, the 95% Crls overlapped; Table 3.5.

Similarly, for Xpert MTB/RIF, we found lower sensitivity in settings with higher tuberculosis prevalence than in those with lower tuberculosis prevalence: pooled sensitivity of 67.0% (49.0 to 81.5) versus 72.0% (62.4 to 81.2). We found lower specificity in settings with higher tuberculosis prevalence than in those with lower tuberculosis prevalence: pooled specificity of 94.1% (86.8 to 97.9) versus 97.3% (95.9 to 98.3). In both analyses, the 95% Crls overlapped; Table 3.6. When we repeated the

analysis at lower tuberculosis prevalence (threshold 10%), in the case of specificity, accuracy in the two groups was significantly different (probability of specificity being lower in the high tuberculosis prevalence group = 0.999); Table 3.6.

Table 3.5: Impact of tuberculosis prevalence on Xpert Ultra and Xpert MTB/RIF sensitivity and specificity

Analysis, (number of studies, specimens)	Pooled sensitivity (95% CrI)	Pooled specificity (95% CrI)
Cerebrospinal fluid, Xpert Ultra		
Among studies with prevalence ≥ 30% (3, 54)	88.3% (68.3 to 97.0)	88.0% (64.3 to 97.9)
Among studies with prevalence < 30% (3, 421)	90.8% (77.3 to 96.9)	91.9% (82.5 to 96.6)
Difference (≥ 30% group minus < 30% group)	−2.2% (−23.0 to 13.5)	−3.8% (−27.7 to 9.8)
Probability (≥ 30% group minus < 30% group)	0.390	0.308
Cerebrospinal fluid, Xpert MTB/RIF		
Among studies with prevalence ≥ 30% (6, 610)	67.0% (49.0 to 81.5)	94.1% (86.8 to 97.9)
Among studies with prevalence < 30% (24, 2785)	72.0% (62.4 to 81.2)	97.3% (95.9 to 98.3)
Difference (≥ 30% group minus < 30% group)	−4.8% (−25.5 to 12.1)	−3.1% (−10.5 to 0.8)
Probability (≥ 30% group minus < 30% group)	0.296	0.071
Cerebrospinal fluid, Xpert MTB/RIF		
Among studies with prevalence ≥ 10% (19, 2190)	68.7% (58.5 to 78.0)	95.1% (92.7 to 96.8)
Among studies with prevalence < 10% (11, 1205)	74.2% (57.4 to 86.6)	98.6% (97.5 to 99.3)
Difference (≥ 10% group minus < 10% group)	-5.5% (-21.2 to 13.3)	-3.5% (-6.0 to -1.5)
Probability (≥ 10% group minus < 10% group)	0.272	0.001
Pleural fluid, Xpert MTB/RIF		
Among studies with prevalence ≥ 50% (6, 627)	20.7% (11.2 to 33.7)	99.6% (97.9 to 99.9)
Among studies with prevalence < 50% (4, 397)	15.5% (6.5 to 30.1)	99.0% (96.9 to 99.8)
Difference (≥ 50% group minus < 50% group)	5.1% (−11.8 to 21.2)	0.5% (−1.2 to 2.7)
Probability (≥ 50% group minus < 50% group)	0.757	0.759
Lymph node aspirate, Xpert MTB/RIF		
Among studies with prevalence ≥ 35 (9, 911)	93.1% (88.9 to 96.3)	83.2% (69.5 to 92.1)

Among studies with prevalence < 35% (5, 677)	72.2% (64.9 to 87.2)	90.0% (78.3 to 95.9)
Difference (≥ 35% group minus < 35% group)	15.7% (5.4 to 28.6)	−6.4% (−21.3 to 76)
Probability (≥ 35% group minus < 35% group)	0.999	0.158

Sensitivity analyses

Overall, the sensitivity analyses made little difference to the findings; Table 3.6.

Table 3.6: Sensitivity analyses, Xpert Ultra in cerebrospinal fluid

Type of specimen	Number of studies (specimens)	Pooled sensitivity (95% CrI)	Pooled specificity (95% CrI)	Predicted sensitivity (95% CrI)	Predicted specificity (95% CrI)
All participants	6 (475)	89.4% (79.1 to 95.6)	91.2% (83.2 to 95.7)	53.0% (36.6 to 69.6)	98.7% (97.5 to 99.5)
Consecutive participant selection	5 (471)	87.9% (76.4 to 94.6)	90.4% (81.1 to 95.1)	88.0% (65.2 to 96.7)	90.5% (65.5 to 97.7)
Reference standard blinding	4 (432)	88.5% (74.7 to 95.6)	89.1% (76.9 to 94.3)	88.6% (63.4 to 97.2)	89.2% (61.0 to 97.1)
Single specimen per participant	4 (432)	88.5% (74.7 to 95.6)	89.1% (76.9 to 94.3)	88.6% (63.4 to 97.2)	89.2% (61.0 to 97.1)

Inconclusive Xpert Ultra and Xpert MTB/RIF results

Xpert Ultra

None of the studies evaluating Xpert Ultra for tuberculous meningitis reported this information.

Xpert MTB/RIF

We previously reported that for CSF, of 2096 tests performed, the pooled proportion of inconclusive Xpert MTB/RIF results was 0.9% (95% CrI 0.3 to 1.9) [60].

B: Pleural tuberculosis: pleural fluid

Xpert Ultra

Culture reference standard

Four studies evaluated Xpert Ultra in pleural fluid with respect to culture [91,92,111]. Xpert Ultra sensitivity ranged from 48% to 84% and specificity ranged from 65% to 100%; Figure 3.9.

In pleural fluid, Xpert Ultra pooled sensitivity and specificity against culture were 75.0% (58.0 to 86.4) and 87.0% (63.1 to 97.9), (4 studies; 398 participants, 158 (39.7%) with tuberculosis); Table 1.

Composite reference standard

Two studies evaluated Xpert Ultra in pleural fluid with respect to a composite reference standard (Meldau 2019; Wang 2019); Figure 9. Sensitivity ranged from 38% to 61%, and specificity ranged from 96% to 99%. We could not explain the variability in the sensitivity estimates and did not perform a meta-analysis.

Latent class meta-analysis

We had insufficient data to obtain robust parameter estimates using the latent class model for Xpert Ultra in pleural fluid.

Xpert MTB/RIF

Culture reference standard

Twenty-eight studies evaluated Xpert MTB/RIF in pleural fluid with respect to culture[91,94,98,103,104,106,109,110,112,115,117,119,120,124–127,129,132,135,141,144,145,153]. Xpert MTB/RIF sensitivity ranged from 0% to 100% and specificity ranged from 82% to 100% (Figure 3.9). Three studies did not contribute data to the meta-analysis because sensitivity was not estimable.

In pleural fluid, Xpert MTB/RIF pooled sensitivity and specificity against culture were 49.5% (39.8 to 59.9) and 98.9% (97.6 to 99.7) (25 studies; 3065 participants, 644 (21.0%) with tuberculosis); Table 1.1.

Composite reference standard

In pleural fluid, Xpert MTB/RIF pooled sensitivity and specificity against a composite reference standard were 18.9% (11.5 to 27.9) and 99.3% (98.1 to 99.8), (10 studies; 1024 participants) Table 1.1; Figure 3.9.

Latent class meta-analysis

Based on the latent class meta-analysis model, Xpert MTB/RIF pooled sensitivity and specificity (95% Crl) in pleural fluid were 53.1% (42.8 to 64.1) and 99.6% (99.3 to 99.8) (25 studies; 3065

participants) Table 3.2. Xpert MTB/RIF pooled sensitivity was slightly higher and its pooled specificity comparable to what was obtained when culture was treated as having perfect accuracy, with pooled sensitivity of 49.5% (39.8 to 59.9) and pooled specificity of 98.8% (97.6 to 99.7). The pooled sensitivity of culture at 89.5% (80.5 to 96.3) was estimated to be lower than 100%. The pooled specificity of culture was estimated to be 99.0% (98.2 to 99.5).

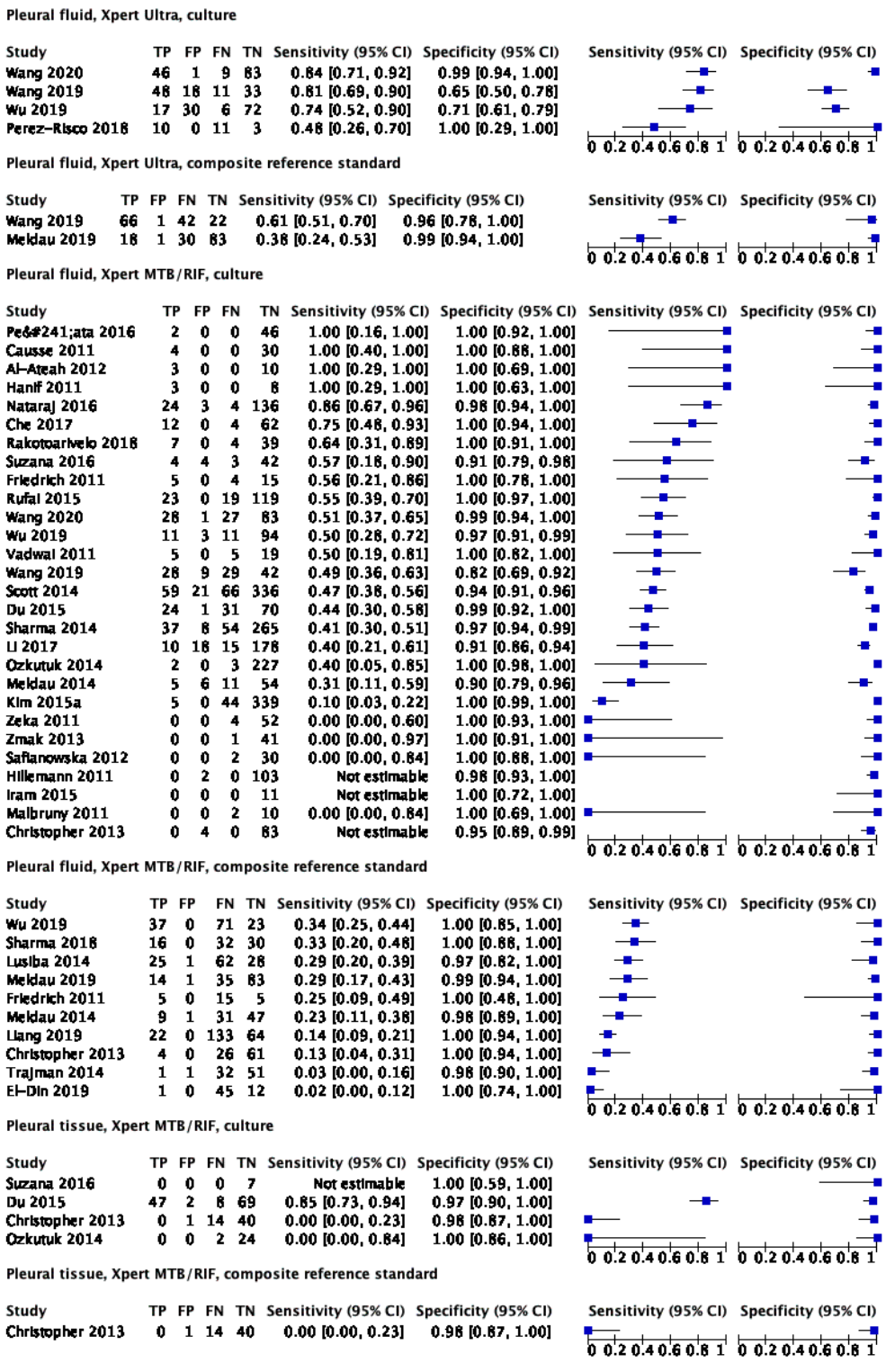


Figure 3.9: Forest plots of Xpert Ultra and Xpert MTB/RIF sensitivity and specificity in pleural fluid and tissue by reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

Xpert Ultra versus Xpert MTB/RIF

We had insufficient data for this analysis.

Impact of tuberculosis prevalence on sensitivity and specificity

For Xpert Ultra, we had insufficient data for this analysis.

For Xpert MTB/RIF, we found higher sensitivity in settings with higher tuberculosis prevalence than in those with lower tuberculosis prevalence: pooled sensitivity (95% CrI) of 20.7% (11.2 to 33.7) versus 15.5% (6.5 to 30.1). We found similar specificity in settings with higher tuberculosis prevalence and in those with lower tuberculosis prevalence: pooled specificity of 99.6% (97.9 to 99.9) versus 99.0% (96.9 to 99.8). In both analyses, the 95% Crls overlapped; Table 5.

Sensitivity analyses

For Xpert Ultra, we had insufficient data for these analyses.

Inconclusive Xpert Ultra and Xpert MTB/RIF results

Xpert Ultra

Of the total 1013 tests performed, the percentage of inconclusive Xpert Ultra results was 0.3%. Only one study reported this information (Wang 2019).

Xpert MTB/RIF

We previously reported that for pleural fluid, of 1416 tests performed the pooled proportion of inconclusive Xpert MTB/RIF results was 1.2% (95% CrI 0.4 to 2.6) [60].

C: Pleural tuberculosis: pleural tissue

Xpert Ultra

Culture reference standard

We did not identify any studies evaluating Xpert Ultra in pleural tissue against culture.

Composite reference standard

We did not identify any studies evaluating Xpert Ultra in pleural tissue against a composite reference standard.

Xpert MTB/RIF

Culture reference standard

Four studies evaluated Xpert MTB/RIF in pleural tissue with respect to culture [98,115,135,141]. Xpert MTB/RIF sensitivity ranged from 0% to 85% and specificity ranged from 97% to 100%; Figure 9. One study reported zero tuberculosis cases [98]. We did not perform a meta-analysis.

Composite reference standard

In pleural tissue, Xpert MTB/RIF sensitivity and specificity against a composite reference standard were 0% (0 to 23) and 98% (87 to 100) (1 study; 55 participants [141]); Figure 9.

D: Lymph node tuberculosis: lymph node aspirate

Xpert Ultra

Culture reference standard

In lymph node aspirates, Xpert Ultra sensitivity and specificity against culture were 78% (40 to 97) and 78% (66 to 87), (1 study; 73 participants; 9 (12.3%) with tuberculosis; Antel 2020[152]); Figure 10.

Composite reference standard

In lymph node aspirates, Xpert Ultra sensitivity and specificity against a composite reference standard were 70% (51 to 85) and 100% (92 to 100), (1 study; 73 participants; 9 (12.3%) with tuberculosis; Antel 2020[152]); Figure 10. Of note, with a composite reference standard, specificity was higher (100%) than that observed when using culture as the reference standard (78%).

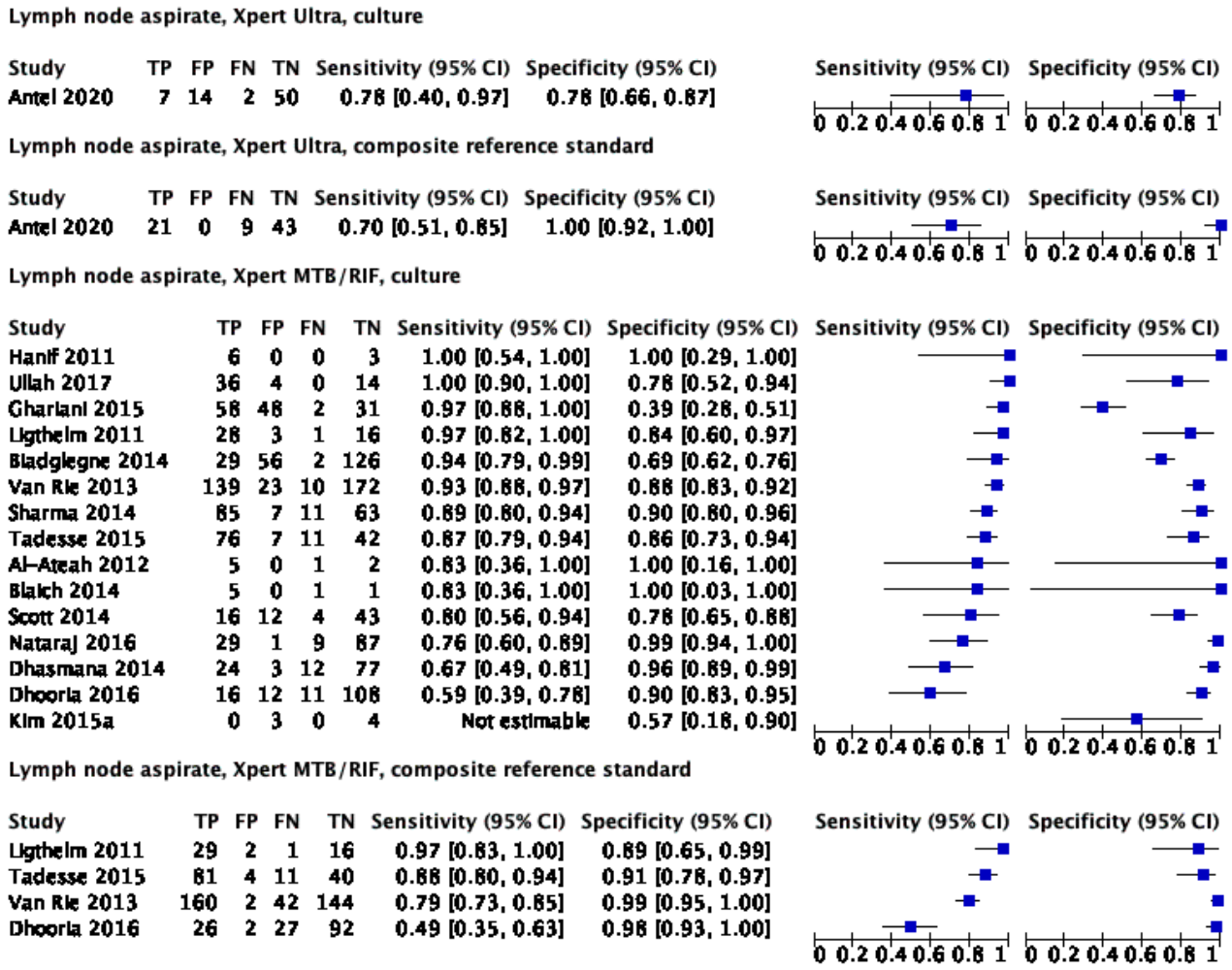


Figure 3.10: Forest plots of Xpert Ultra and Xpert MTB/RIF sensitivity and specificity in lymph node aspirate by reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

Latent class meta-analysis

We had insufficient data to obtain robust parameter estimates using the latent class model for Xpert Ultra in lymph node aspirate.

Xpert MTB/RIF

Culture reference standard

Fifteen studies evaluated Xpert MTB/RIF in lymph node aspirates with for culture [93,95,97,103,104,117,122,125,129,131,137,138,146,147,153]. Xpert MTB/RIF sensitivity ranged from 59% to 100% and specificity from 57% to 100%; Figure 17. Xpert MTB/RIF specificity in lymph node aspirates was considerably more heterogeneous than in CSF and pleural fluid. The variability in Xpert MTB/RIF specificity in lymph node aspirates was unexpected and may be the result of a systematic, unexplained bias in some studies. One study did not contribute data to the meta-analysis because sensitivity was not estimable[125].

In lymph node aspirates, Xpert MTB/RIF pooled sensitivity and specificity against culture were 88.9% (82.7 to 93.6) and 86.2% (78.0 to 92.3) (14 studies; 1588 participants, 627 (39.5%) with tuberculosis); Table 3.1.

Composite reference standard

In lymph node aspirates, Xpert MTB/RIF pooled sensitivity and specificity against a composite reference standard were 81.6% (61.9 to 93.3) and 96.4% (91.3 to 98.6), (4 studies; 679 participants); Table 2; Figure 10. Of note, with a composite reference standard, specificity was less variable and pooled specificity higher than that observed when using culture as the reference standard (86.0%).

Latent class meta-analysis

Based on the latent class meta-analysis model, Xpert MTB/RIF pooled sensitivity and specificity (95% Crl) in lymph node aspirate were 91.3% (84.9 to 96.3) and 99.5% (99.1 to 99.7) (14 studies; 1588 participants); Table 3. Xpert MTB/RIF pooled sensitivity and pooled specificity were higher than when culture was treated as having perfect accuracy, with pooled sensitivity of 88.9% (82.7 to 93.6) and pooled specificity of 86.2% (78.0 to 92.3). The pooled sensitivity of culture at 84.9% (74.0 to 92.8) was estimated to be lower than 100%. The pooled specificity of culture was estimated to be 98.8% (97.7 to 99.4); Table 3. The latent class meta-analysis resulted in high precision in the specificity of Xpert MTB/RIF across studies. This was the result of adjustments for the imperfect and heterogeneous accuracy of culture across studies.

Xpert Ultra versus Xpert MTB/RIF

We had insufficient data for this analysis.

Impact of tuberculosis prevalence on sensitivity and specificity

For Xpert Ultra, we had insufficient data for this analysis.

For Xpert MTB/RIF, we found higher sensitivity in settings with higher tuberculosis prevalence than in those with lower tuberculosis prevalence: pooled sensitivity (95% CrI) of 93.1% (88.9 to 96.3) versus 72.2% (64.9 to 87.2). We found lower specificity in settings with higher tuberculosis prevalence than in those with lower tuberculosis prevalence: pooled specificity of 83.2% (69.5 to 92.1) versus 90.0% (78.3 to 95.9). In the case of sensitivity, accuracy in the two groups was significantly different (probability of sensitivity being lower in the high tuberculosis prevalence group = 0.999); Table 3.6.

Sensitivity analyses

For Xpert Ultra, we had insufficient data for these analyses.

Inconclusive Xpert MTB/RIF and Xpert Ultra results

Xpert Ultra

None of the studies reported this information.

Xpert MTB/RIF

We previously reported that for lymph node aspirates, in the 1134 tests performed, the pooled proportion of inconclusive Xpert MTB/RIF results was 1.0% (95% CrI 0.4 to 2.0) [60].

E: Lymph node tuberculosis: lymph node tissue

Xpert Ultra

Culture reference standard

In lymph node biopsies, Xpert Ultra sensitivity and specificity against culture were 90% (55 to 100) and 87% (77 to 94) (Antel 2020) and 100% (75 to 100) and 38% (22 to 55) (Wu 2019), (2 studies; 131 participants, 23 (17.6%) with tuberculosis); Figure 3.11.

Composite reference standard

In lymph node biopsies, Xpert Ultra sensitivity and specificity against a composite reference standard were 73% (50 to 89) and 96% (88 to 100), (1 study; 81 participants)[152]; Figure 11.

Xpert MTB/RIF

Culture reference standard

Eleven studies evaluated Xpert MTB/RIF in lymph node biopsies against culture [90,92,98,103,105,112,115,125,131,145,146] . Xpert MTB/RIF sensitivity ranged from 50% to 100% and specificity ranged from 0% to 100%; Figure 4.11. We could not explain the heterogeneity in accuracy estimates by study quality, small numbers, or other factors.

In lymph node biopsies, Xpert MTB/RIF pooled sensitivity and specificity against culture were 82.4% (73.5 to 89.7) and 80.3% (60.3 to 91.5), (11 studies; 786 participants, 220 (28.0%) with tuberculosis); Table 4.1.

Composite reference standard

In lymph node biopsies, Xpert MTB/RIF sensitivity and specificity against a composite reference standard were 33% (27 to 40) and 85% (73 to 93) (Sarfaraz 2018)[105] and 76% (50 to 93) and specificity of 100% (66 to 100) (Zeka 2011[90]) (2 studies; 288 participants); Figure 11.

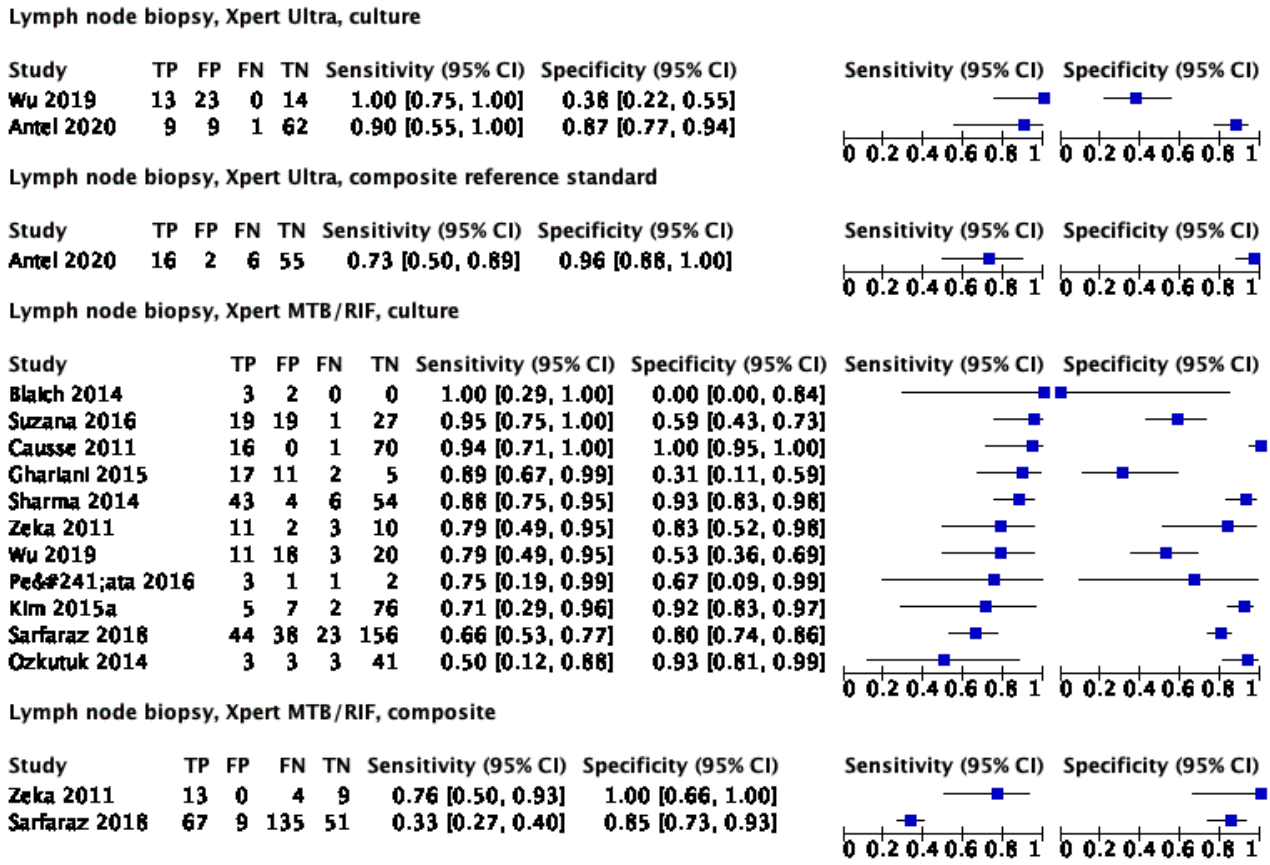


Figure 3.11: Forest plots of Xpert Ultra and Xpert MTB/RIF sensitivity and specificity in lymph node biopsy by reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

F: Genitourinary tuberculosis

Xpert Ultra

Culture reference standard

In urine, Xpert Ultra sensitivity and specificity against culture were 100% (74 to 100) and 100% (74 to 100), (1 study; 24 participants, 12 (50%) with tuberculosis) (Perez-Risco 2018)[111]; Figure 3.12.

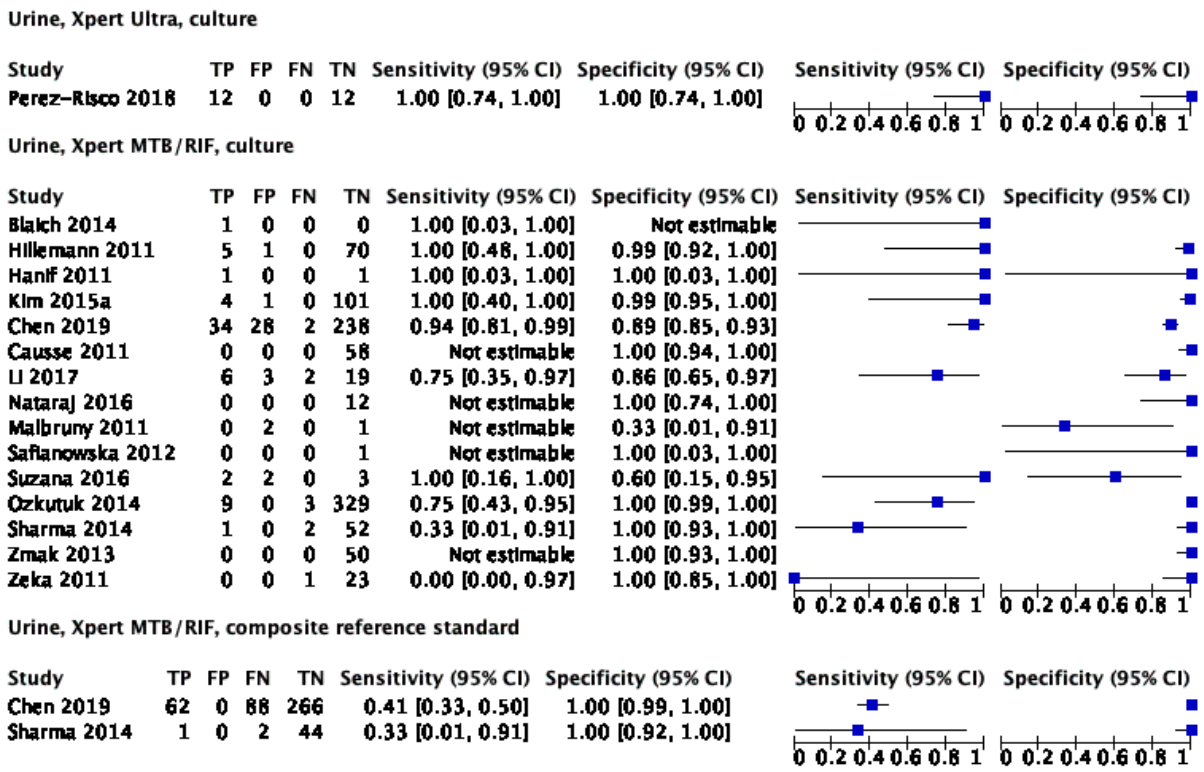


Figure 3.12: Forest plots of Xpert Ultra and Xpert MTB/RIF sensitivity and specificity for rifampicin resistance.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

Composite reference standard

We did not identify any studies that evaluated Xpert Ultra in urine against a composite reference standard.

Xpert MTB/RIF

Culture reference standard

Fifteen studies evaluated Xpert MTB/RIF in urine against culture [89,90,98,103,106,115,117,120,124,125,127,129,143,145,146]. Xpert MTB/RIF sensitivity ranged from 0% to 100% (sensitivity of 0% was reported by Zeka 2011[90] that had only one culture positive, which was Xpert negative) and specificity from 33% to 100% (Figure 12). Six studies did not contribute data to the meta-analysis because either sensitivity or specificity was not estimable.

In urine, Xpert MTB/RIF pooled sensitivity and specificity against culture were 85.9% (71.4 to 94.3) and 98.1% (93.1 to 99.7) (9 studies; 943 participants, 72 (7.6%) with tuberculosis); Table 3.1.

Composite reference standard

In urine, Xpert MTB/RIF sensitivity and specificity against a composite reference standard were 33% (1 to 91) and 100% (92 to 100) (Sharma 2014[103]), and 41% (33 to 50) and 100% (99 to 100) (Chen 2019[143]) (2 studies; 463 participants); Figure 12.

G: Bone or joint tuberculosis: aspirate

Xpert Ultra

Culture reference standard

In bone or joint aspirate, Xpert Ultra sensitivity and specificity against culture were 88% (47 to 100) (specificity was not estimable) (Perez-Risco 2018), and 96% (87 to 100) and 97% (85 to 100) (Sun 2019) (2 studies; 94 participants, 60 (63.8%) with tuberculosis).

Composite reference standard

In bone or joint aspirate, Xpert Ultra sensitivity and specificity against a composite reference standard were 96% (91 to 99) and 97% (85 to 100), (1 study; 145 participants)[99].

Xpert MTB/RIF

Culture reference standard

Twelve studies evaluated Xpert MTB/RIF in bone or joint fluid for culture [98,99,106,112,115,117,120,124,125,130,146,153]. Xpert MTB/RIF sensitivity ranged from 96% to 100% and specificity ranged from 53% to 100%.

In bone or joint aspirate, Xpert MTB/RIF pooled sensitivity and specificity against culture were 97.9% (93.1 to 99.6) and 97.4% (80.2 to 100.0); (6 studies; 471 participants, 110 (23.4%) with tuberculosis); Table 3.1

Composite reference standard

In bone or joint aspirate, Xpert MTB/RIF sensitivity and specificity against a composite reference standard were 82% (69 to 91) and 100% (69 to 100)[130], and 94% (87 to 97) and 100% (90 to 100) (2 studies; 205 participants)[99].

H: Bone or joint tuberculosis: tissue

Xpert Ultra

Culture reference standard

We did not identify any studies that evaluated Xpert Ultra in tissue for bone or joint tuberculosis against culture.

Composite reference standard

We did not identify any studies that evaluated Xpert Ultra in tissue for bone or joint tuberculosis against a composite reference standard.

Xpert MTB/RIF

Culture reference standard

Three studies evaluated Xpert MTB/RIF in bone or joint tissue against culture. Xpert MTB/RIF sensitivity ranged from 50% to 100% and specificity ranged from 94% to 100%.

In bone or joint tissue, Xpert MTB/RIF sensitivity and specificity (95% CI) against culture were 100% (3 to 100) and 100% (48 to 100) (Malbruny 2011), 100% (3 to 100) and 100% (40 to 100) (Ozkutuk 2014), and 50% (1 to 99) and 94% (71 to 100) (Peñata 2016); (3 studies; 30 participants, 4 (13.3%) with tuberculosis).

Composite reference standard

We did not identify any studies that evaluated Xpert MTB/RIF in tissue for bone or joint tuberculosis against a composite reference standard.

J: Peritoneal tuberculosis: fluid

Xpert Ultra

Culture reference standard

In peritoneal fluid, Xpert Ultra sensitivity against culture was 33% (1 to 91) and specificity was not estimable (1 study; 3 participants)[111].

Composite reference standard

We did not identify any studies that evaluated Xpert Ultra in peritoneal fluid against a composite reference standard.

Xpert MTB/RIF

Culture reference standard

Seventeen studies evaluated Xpert MTB/RIF in peritoneal fluid against culture [89,90,94,95,98,103,104,106,109,112,115,120,124–126,145,153]. Four studies (Al-Ateah 2012; Causse 2011; Iram 2015; Safianowska 2012) did not contribute data to the meta-analysis because sensitivity was not estimable. In individual studies, Xpert MTB/RIF sensitivity ranged from 33% to 100% and specificity ranged from 90% to 100%.

In peritoneal fluid, Xpert MTB/RIF pooled sensitivity and specificity against culture were 59.1% (42.1 to 76.2) and 97.6% (95.4 to 98.9), (13 studies; 580 participants, 94 (16.2%) with tuberculosis); Table 3.1.

Composite reference standard

We did not identify any studies that evaluated Xpert MTB/RIF in peritoneal fluid against a composite reference standard.

K: Peritoneal tuberculosis: tissue

Xpert Ultra

Culture reference standard

We did not identify any studies that evaluated Xpert Ultra in peritoneal tissue against culture.

Composite reference standard

We did not identify any studies that evaluated Xpert Ultra in peritoneal tissue against a composite reference standard.

Xpert MTB/RIF

Culture reference standard

In peritoneal tissue, Xpert MTB/RIF sensitivity and specificity against culture were 50% (7 to 93) and 92% (73 to 99) (1 study; 28 participants; Bera 2015).

Composite reference standard

We did not identify any studies that evaluated Xpert MTB/RIF in peritoneal tissue against a composite reference standard.

L: Pericardial tuberculosis: fluid

Xpert Ultra

Culture reference standard

We did not identify any studies that evaluated Xpert Ultra in pericardial fluid against culture.

Composite reference standard

We did not identify any studies that evaluated Xpert Ultra in pericardial fluid against a composite reference standard.

Xpert MTB/RIF

Culture reference standard

Fourteen studies evaluated Xpert MTB/RIF in pericardial fluid against culture [89,90,94,95,98,103,106,112,114,115,125,145,146,153]. Xpert MTB/RIF sensitivity ranged from 0% to 100% and specificity ranged from 69% to 100%. Nine studies did not contribute data to the meta-analysis because either sensitivity or specificity was not estimable.

In pericardial fluid, Xpert MTB/RIF pooled sensitivity and specificity against culture were 61.4% (32.4 to 82.4) and 89.7% (74.9 to 99.0), (5 studies; 181 participants, 57 (31.5%) with tuberculosis); Table 3.1.

Composite reference standard

We did not identify any studies that evaluated Xpert MTB/RIF in pericardial fluid against a composite reference standard.

M: Disseminated tuberculosis: blood

Xpert Ultra

Culture reference standard

We did not identify any studies that evaluated Xpert Ultra in blood against culture.

Composite reference standard

We did not identify any studies that evaluated Xpert Ultra in blood against a composite reference standard.

Xpert MTB/RIF

Culture reference standard

Two studies evaluated Xpert MTB/RIF in blood against culture [89,133]. However, only one of these studies reported tuberculosis culture-positives. Xpert MTB/RIF sensitivity and specificity against culture were 56% (21 to 86) and 94% (85 to 98) (1 study; 74 participants, 9 (12.2%) with tuberculosis[133]).

Composite reference standard

We did not identify any studies that evaluated Xpert MTB/RIF in blood against a composite reference standard.

Nontuberculous mycobacteria

For Xpert Ultra, two studies provided data on a variety of NTMs that grew from the specimens tested to look for evidence of cross-reactivity. Donovan 2020 assessed Xpert Ultra specificity in CSF from more than 100 participants with nontuberculous meningitis and found zero positive Xpert Ultra results in those with a probable or possible diagnosis of tuberculous meningitis and in any participant with a confirmed diagnosis of nontuberculous meningitis. Perez-Risco 2018 assessed Xpert Ultra specificity using 20 culture-positive NTM specimens (covering a total of 18 species) and found that Xpert Ultra was negative for all specimens.

For Xpert MTB/RIF, we previously reported that in 10 studies involving 6975 specimens with 141 NTMs, Xpert MTB/RIF was negative in all specimens[60].

II. Detection of rifampicin resistance

Xpert Ultra and Xpert MTB/RIF testing for rifampicin resistance

Xpert Ultra

Five studies evaluated Xpert Ultra for detection of rifampicin resistance. Xpert Ultra sensitivity estimates varied from 50% to 100%; specificity varied from 93% to 100%; Figure 22. One study reported zero participants with rifampicin resistance and thus sensitivity was not estimable (Chin 2019)[142]. Four studies contributed data to the bivariate meta-analysis [91,92,99]. Xpert Ultra pooled sensitivity and specificity were 100.0% (95.1 to 100.0) and 100.0% (99.0 to 100.0), (4 studies; 129 participants, 24 (18.6%) with rifampicin resistance); Table 3.1.

Xpert MTB/RIF

Xpert MTB/RIF pooled sensitivity and specificity were 96.5% (91.9 to 98.8) and 99.1% (98.0 to 99.7) (19 studies; 970 participants, 148 (15.3%) with rifampicin resistance); Table 3.1; Figure 3.13.

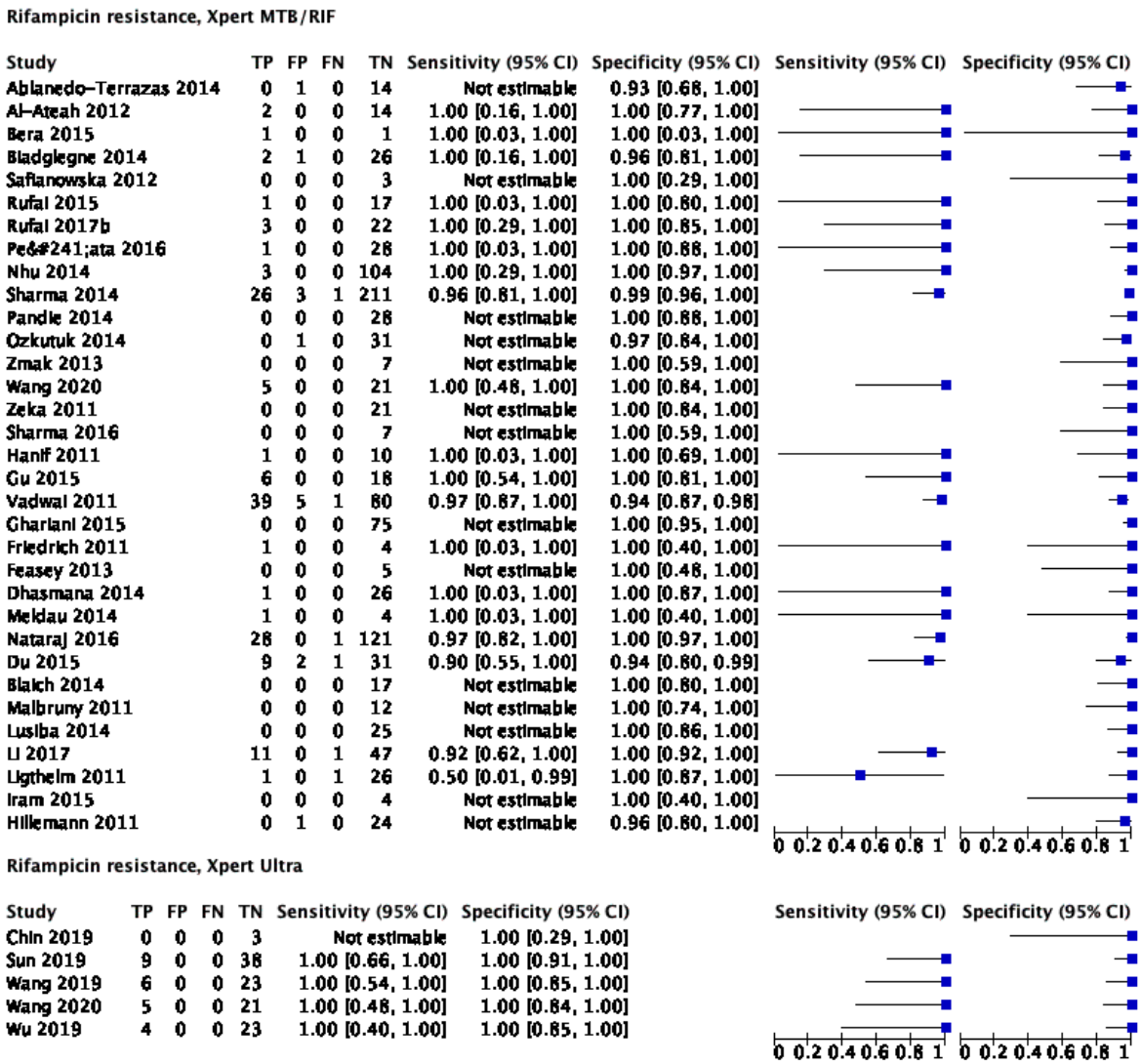


Figure 3.13: Forest plots of Xpert MTB/RIF and Xpert Ultra sensitivity and specificity in urine by reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

Indeterminate Xpert Ultra and Xpert MTB/RIF results for rifampicin resistance

Xpert Ultra

Of the total 391 tests positive by Xpert Ultra, the proportion of indeterminate Xpert Ultra results for RIF resistance was 36.1%. All of these indeterminate results were Xpert Ultra trace-positive.

Xpert MTB/RIF

We previously reported that for rifampicin resistance testing, of 1003 tests performed, the pooled proportion of indeterminate Xpert MTB/RIF results was 2.6% (95% CrI 1.4 to 4.3)[60].

Discussion

Summary of main results

This systematic review update summarizes the current literature and includes 69 unique studies on the accuracy of Xpert Ultra and Xpert MTB/RIF for extrapulmonary tuberculosis and rifampicin resistance. We identified 11 studies evaluating Xpert Ultra, an increase of 10 new studies since the original review [60]. Unlike the original review, we limited inclusion to adults aged 15 years and older. We also include a composite reference standard in addition to a culture reference standard, and have stratified all analyses by type of reference standard. Major findings from our review include the following.

- Xpert Ultra sensitivity for tuberculosis varied across different types of specimens (from 75.0% in pleural fluid to 89.4% in cerebrospinal fluid); Table 3.1.
- Xpert MTB/RIF sensitivity for tuberculosis varied across different types of specimens (from 49.5% in pleural fluid to 97.9% in bone or joint aspirate); Table 3.1.
- Xpert MTB/RIF specificity in cerebrospinal fluid, pleural fluid, urine, bone or joint aspirate, and peritoneal fluid was ≥ 96.9%, against culture; overall, Xpert Ultra specificities were lower than those of Xpert MTB/RIF against culture, but against a composite reference standard results for both index tests were similar; Table 3.1.
- In cerebrospinal fluid, Xpert Ultra sensitivity and specificity were 89.4% (79.1 to 95.6) and 91.2% (83.2 to 95.7) against culture.
- In cerebrospinal fluid, Xpert MTB/RIF sensitivity and specificity were 71.1% (62.8 to 79.1) and 96.9% (95.4 to 98.0) against culture.
- In pleural fluid, Xpert Ultra sensitivity and specificity were 75.0% (58.0 to 86.4) and 87.0% (63.1 to 97.9) against culture.
- In pleural fluid, Xpert MTB/RIF sensitivity and specificity were 49.5% (39.8 to 59.9) and 98.9% (97.6 to 99.7) against culture.
- In lymph node aspirate, Xpert Ultra sensitivity and specificity were 70% (51 to 85) and 100% (92 to 100) against a composite reference standard (1 study).

- In lymph node aspirate, Xpert MTB/RIF sensitivity and specificity were 81.6% (61.9 to 93.3) and 96.4% (91.3 to 98.6) against a composite reference standard.
- For rifampicin resistance, Xpert Ultra sensitivity and specificity were 100.0% (95.1 to 100.0) and 100.0% (99.0 to 100.0).
- For rifampicin resistance, Xpert MTB/RIF sensitivity and specificity were 96.5% (91.9 to 98.8) and 99.1% (98.0 to 99.7).

Xpert Ultra and Xpert MTB/RIF testing in cerebrospinal fluid

Xpert Ultra

Results of these studies indicate that in theory, for a population of 1000 people where 100 have tuberculosis meningitis on culture, 168 would be Xpert Ultra-positive: of these, 79 (47%) would not have tuberculosis (false-positives); and 832 would be Xpert Ultra-negative: of these, 11 (1%) would have tuberculosis (false-negatives).

Xpert MTB/RIF

Results of these studies indicate that in theory, for a population of 1000 people where 100 have tuberculosis meningitis on culture, 99 would be Xpert MTB/RIF-positive: of these, 28 (28%) would not have tuberculosis (false-positives); and 901 would be Xpert MTB/RIF-negative: of these, 29 (3%) would have tuberculosis (false-negatives).

Rapid diagnosis of tuberculous meningitis is critical so that lifesaving treatment can be started promptly. Around 50% of those affected die or experience disabling consequences [156]. Xpert Ultra was designed to improve tuberculosis detection, in particular in people with paucibacillary disease. The limit of detection for MTB is lower with Xpert Ultra (16 bacterial colony-forming units (cfu) per mL) than with Xpert MTB/RIF (131 cfu per mL) [157]. In studies that compared Xpert Ultra and Xpert MTB/RIF in the same participants, we found Xpert Ultra to have higher pooled sensitivity (89.0%) than Xpert MTB/RIF (62.2%), and lower pooled specificity (91.0%) than Xpert MTB/RIF (96.8%) for tuberculous meningitis. In addition, in subgroup analyses we found slightly higher Xpert Ultra accuracy in studies that concentrated the cerebrospinal fluid (CSF): pooled sensitivity of 90.5% in concentrated specimens versus 88.4% in unconcentrated specimens; and pooled specificity of 91.9% in concentrated specimens versus 88.6% in unconcentrated specimens. We note that subgroup findings

should be interpreted with caution, as there were only three studies and a small number of tuberculous meningitis cases included. The Tuberculous Meningitis International Research Consortium has recommended increasing the volume of CSF collected for diagnosis followed by centrifugation as a way of improving Xpert MTB/RIF assay sensitivity [150]; however, we did not have sufficient data to investigate CSF collection volume. Increased Xpert MTB/RIF sensitivity in HIV-positive people compared with HIV-negative people has been reported, with the increased bacterial burden in tuberculosis and HIV co-infection proposed as the reason [113]. We had limited data to investigate this for Xpert Ultra as we identified only two studies in HIV-positive people, with sensitivities of 90% [150]and 89% [140].

Xpert Ultra and Xpert MTB/RIF testing in pleural fluid

Xpert Ultra

Results of these studies indicate that in theory, for a population of 1000 people where 100 have pleural tuberculosis on culture, 192 would be Xpert Ultra-positive: of these, 117 (61%) would not have tuberculosis (false-positives); and 808 would be Xpert Ultra-negative: of these, 25 (3%) would have tuberculosis (false-negatives).

Xpert MTB/RIF

Results of these studies indicate that in theory, for a population of 1000 people where 100 have pleural tuberculosis on culture, 60 would be Xpert MTB/RIF-positive: of these, 10 (17%) would not have tuberculosis (false-positives); and 940 would be Xpert MTB/RIF-negative: of these, 50 (5%) would have tuberculosis (false-negatives).

With the bivariate model, we found Xpert Ultra to have higher pooled sensitivity (75.0%) than Xpert MTB/RIF (49.6%) and lower pooled specificity (87.0%) than Xpert MTB/RIF (98.7%) in pleural fluid against a culture reference standard, between-study comparison. Based on the latent class meta-analysis model, Xpert Ultra pooled sensitivity was comparable (76.0%) and specificity higher (99.5%) than what was obtained when culture was treated as having perfect accuracy. Xpert Ultra pooled sensitivity in pleural fluid was lower than that of CSF. One reason for the lower sensitivity of Xpert Ultra in pleural fluid could be the paucibacillary nature of pleural tuberculosis. Other possible reasons are contamination of blood or the presence of certain polymerase chain reaction (PCR) inhibitors in the pleural fluid [158,159]. However, Theron and colleagues found that extrapulmonary specimens showed

less evidence of PCR inhibition than pulmonary specimens, with bacterial load being more important for a positive Xpert MTB/RIF result [160]. Given that false-negative results were common (low sensitivity), a negative Xpert Ultra or Xpert MTB/RIF result may not be relied on to exclude tuberculosis.

Xpert Ultra testing in lymph node aspirates

Xpert Ultra

Results of these studies indicate that in theory, for a population of 1000 people where 100 have lymph node tuberculosis verified by a composite reference standard, 70 would be Xpert Ultra-positive: of these, 0 (0%) would not have tuberculosis (false-positives); and 930 would be Xpert Ultra-negative: of these, 30 (3%) would have tuberculosis (false-negatives).

Xpert MTB/RIF

Results of these studies indicate that in theory, for a population of 1000 people where 100 have lymph node tuberculosis verified by a composite reference standard, 118 would be Xpert MTB/RIF-positive: of these 37 (31%) would not have tuberculosis (false-positives); and 882 would be Xpert MTB/RIF-negative: of these 19 (2%) would have tuberculosis (false-negatives).

Regarding Xpert testing for lymph node aspirate, it important to point out that although tissue biopsy provides material for histological examination which may be of substantial diagnostic value, a fluid specimen may be collected more easily. In addition, fine-needle aspiration of lymph nodes is well suited for use in resource-limited settings because the procedure is simple, easy to learn, minimally invasive, and inexpensive [161]. Thus clinicians may want to consider fine-needle aspiration of lymph nodes before surgical biopsy.

In our review, using a standard bivariate meta-analysis model, Xpert MTB/RIF pooled specificity (defined by culture) in lymph node aspirate was 86.0%, whereas with a composite reference standard pooled specificity increased to 95.9%. Using a latent class meta-analysis model with informative priors, Xpert MTB/RIF pooled specificity increased to 99.5%. In previous meta-analyses, Xpert MTB/RIF specificity for lymph node tuberculosis (aspirate and tissue) against culture as a reference standard was 94% [162], 93% [163], and 92% [164]. Using a composite reference standard (defined by the primary study authors), Denkinger 2014 found increased Xpert MTB/RIF specificity of 99% for lymph

node tuberculosis (5 studies, 728 specimens). Thus, it appears that accuracy results depend in part on the choice of reference standard. Regarding the use of a composite reference standard, owing to differing definitions and difficulty in interpreting them, there is a risk of bias [165] (see section Strengths and weaknesses of the review).

We considered several reasons why the specificity of Xpert Ultra (78%) and Xpert MTB/RIF (86.0%) in lymph node aspirate against culture would be lower than in other extrapulmonary specimens. Although not always reported, studies may have included participants receiving tuberculosis treatment. We previously reported that in a sensitivity analysis limiting inclusion to studies that involved participants not receiving tuberculosis treatment, specificity increased from 86% to 89%[65]. We considered the type of culture used in the included studies because liquid culture is more sensitive than solid culture [65] . Most studies used liquid culture or a combination of solid and liquid culture. The single study evaluating Xpert Ultra used liquid culture. Only two of the 15 studies (13%) evaluating Xpert MTB/RIF exclusively used solid culture. Culture results may also be negative owing to inefficient specimen collection or errors in sampling, differing bacterial load, and contamination[161] . Negative culture results in lymph node tuberculosis have previously been reported [166].
Another reason for negative culture results is that there may have been a decrease in live tuberculosis bacteria during processing with N-acetyl-L-cysteine-sodium hydroxide, which is routinely used to homogenize, decontaminate, and liquefy non-sterile specimens, such as sputum, for mycobacterial culture. Harsh decontamination practices have been noted to contribute to false-negative culture results, especially in paucibacillary specimens (FIND 2017). Standards specify, "specimens collected from normally sterile sites may be placed directly into the culture medium" [65]. CSF, pleural fluid, and lymph node aspirates are usually considered to be sterile specimens. It is our understanding that some laboratories do decontaminate sterile site specimens as a precaution against non-sterile collection procedures. In this review, 47% of the studies reported decontaminating lymph node aspirates before culture inoculation. We did not have sufficient data to further investigate laboratory practices.

In summary, several factors probably contributed to low Xpert MTB/RIF specificity against culture in lymph node aspirate. The 'true' specificity of Xpert MTB/RIF in lymph node aspirate is likely to be higher owing to the aforementioned reasons. Xpert MTB/RIF specificity was higher against a

composite reference standard and with application of latent class analysis, similar to that found in CSF, pleural fluid, and other specimens (Table 3.2; Table 3.3).

We investigated the prevalence of extrapulmonary tuberculosis (confirmed by culture) as a potential source of heterogeneity because changes in disease prevalence have often been found to be associated with other important changes, such as changes in the disease spectrum, which may affect diagnostic accuracy estimates[167]. For Xpert MTB/RIF, for pleural fluid and lymph node aspirate, we found that pooled sensitivity was higher in settings with higher tuberculosis prevalence. In all analyses, for both Xpert Ultra (CSF) and Xpert MTB/RIF (CSF, pleural fluid, and lymph node aspirate), specificity in settings with higher tuberculosis prevalence was similar or lower than in settings with lower tuberculosis prevalence. Findings from additional analyses are available in the previous version of this review[60].

Xpert Ultra and Xpert MTB/RIF testing for rifampicin resistance

Xpert Ultra

Results of these studies indicate that in theory, for a population of 1000 people where 100 have rifampicin resistance, 100 would be Xpert Ultra-positive (resistant): of these, zero (0%) would not have rifampicin resistance (false-positives); and 900 would be Xpert Ultra-negative (susceptible): of these, zero (0%) would have rifampicin resistance.

Xpert MTB/RIF

Results of these studies indicate that in theory, for a population of 1000 people where 100 have rifampicin resistance, 105 would be Xpert MTB/RIF-positive (resistant): of these, 8 (8%) would not have rifampicin resistance; and 895 would be Xpert MTB/RIF-negative (susceptible): of these, 3 (0.3%) would have rifampicin resistance.

For detection of rifampicin resistance in extrapulmonary specimens, we found the sensitivity of Xpert Ultra (100%) and Xpert MTB/RIF (96.7%) and the specificity of Xpert Ultra (100%) and Xpert MTB/RIF (99.1%), to be comparable to estimates in pulmonary specimens: sensitivity (96%) and specificity (98%) (Horne 2019). We caution that the results for Xpert Ultra are based on only four

studies, involving 129 participants, 24 (18.6%) with rifampicin resistance. Nonetheless, these findings suggest that the use of Xpert Ultra and Xpert MTB/RIF in extrapulmonary specimens could assist in rapid diagnosis of rifampicin-resistant tuberculosis and early initiation of treatment for multidrug-resistant tuberculosis (MDR-TB).

Notably, concerns have been raised about rapid drug susceptibility testing (DST) methods, in particular automated mycobacteria growth indicator tube (MGIT) 960 for tuberculosis drug resistance using the recommended critical concentrations. As a priority, the WHO is planning to re-evaluate the critical concentrations for rifampicin [168].

For Xpert Ultra, we found a high rate (36.1%) of indeterminate rifampicin resistance results, all owing to trace call results. This finding was expected since, for trace call results, rifampicin resistance cannot be determined. Xpert Ultra incorporates two new multi-copy amplification targets (IS*6110* and IS*1081*). Trace call indicates that only the multi-copy targets were detected, and not the tuberculosis-specific regions in the *rpoB* gene. Resistance to rifampicin has mainly been associated mainly with mutations in a limited region of the *rpoB* gene [169].

People-important outcomes, such as mortality, are especially relevant to patients, decision-makers, and the wider tuberculosis community. While performing this systematic review, we did not identify direct evidence of studies linking true-positives, false-positives, true-negatives, and false-negatives to people-important outcomes when either Xpert Ultra or Xpert MTB/RIF was used to diagnose extrapulmonary tuberculosis. To our knowledge, for pulmonary tuberculosis, there have been two systematic reviews of randomized trials on the impact of the use of Xpert MTB/RIF on health outcomes. Both reviews compared the effect of Xpert MTB/RIF and smear microscopy on all-cause mortality; Di Tanna and colleagues summarized the accuracy of Xpert MTB/RIF in an individual patient-level data meta-analysis (3 trials, 8143 participants) [170]and Haraka and colleagues performed a random-effects meta-analysis, (5 trials, 10,409 participants[60,171]. In both reviews, Xpert MTB/RIF did not show a statistically significant effect on all-cause mortality, although the direction of effect was towards mortality reduction. Insufficient power to detect mortality in randomized trials measuring the impact of diagnostic tests on patient-important outcomes has been discussed previously as a limitation of such trials [170,172]. Larger sample sizes are needed to evaluate the effect of Xpert MTB/RIF on mortality, but achieving this is difficult in pragmatic situations. For example, Schumacher

2019 showed that a sample size of 31,000 participants would be needed if researchers were to plan a cluster-randomized diagnostic trial using the baseline mortality and effect size demonstrated by the individual patient data from Di Tanna 2019.

This review represents the most comprehensive review of the diagnostic accuracy of Xpert Ultra and Xpert MTB/RIF for extrapulmonary tuberculosis in adults. For Xpert MTB/RIF, our previous review [60]provides additional findings. These reviews provide evidence that may help countries to make decisions about scaling up the tests for programmatic management of tuberculosis and drug-resistant tuberculosis. Although the information in this review will help to inform such decisions, other factors such as resource requirements and feasibility (including stable electrical power supply, temperature control, and maintenance of the cartridge modules) will also be important considerations.

Strengths and weaknesses of the review

Completeness of evidence

This is a reasonably complete data set. We included any non-English studies that we found from which we could obtain accuracy data. However, we acknowledge that we may have missed some studies despite the comprehensive search and our outreach to investigators. We included eight common forms of extrapulmonary tuberculosis in the review. However, for some of these forms, such as disseminated tuberculosis, data were insufficient to allow us to determine summary accuracy estimates. We did not include less common forms, such as cutaneous tuberculosis, ocular tuberculosis, female genital tuberculosis, and tuberculosis of the breast. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy (PRISMA-DTA) [173].

Accuracy of the reference standards used

In a systematic review of diagnostic test accuracy studies, the reference standard is the best available test to determine the presence or absence of the target condition. In this review, we used two reference standards: culture and a composite reference standard, both of which are known to be imperfect. While the composite reference standard is designed to have improved accuracy compared to culture alone, it may still lead to biased accuracy estimates of the index test, depending on various factors such as the accuracy of the different components; decision rules for combining them; prevalence of the target condition; and conditional dependence between the components and the index test [165]. Conditional

dependence between two imperfect tests arises when both tests make the same false-positive or false-negative errors more often than expected by chance[66]. Hence, conditional dependence may arise between the index test and both reference standards we have used, as they are imperfect. As a consequence, we may over- or underestimate the diagnostic accuracy of the index tests. An additional challenge with including a composite reference standard is that the definition of the composite reference standard may vary across studies, making it difficult to interpret the accuracy estimates. To adjust for the imperfect accuracy of culture, we applied a latent class model when evaluating Xpert MTB/RIF sensitivity and specificity, for which there were a larger number of studies. We added parameters for the sensitivity and specificity of culture and terms for conditional dependence to adjust for the dependence between Xpert MTB/RIF and culture among disease-positive and disease-negative participants. In this way, we were able to improve estimation of both the pooled sensitivity and specificity of Xpert MTB/RIF, as well as between-study variability. An adequate number of studies is needed for a sufficiently robust model to estimate these additional parameters. We therefore found that we were unable to do the same for meta-analyses of the accuracy for Xpert Ultra owing to the small number of studies, many of which had small sample sizes resulting in zero cell counts.

Several factors may have contributed to false-negative culture results for the accuracy of the reference standard for lymph node aspirate in particular, including inefficient specimen collection and overly harsh decontamination. For this particular analysis, we were able to take advantage of the Bayesian estimation approach to incorporate prior information on Xpert MTB/RIF and culture specificity. This allowed us to make the best use of data from the included studies and our knowledge of the performance of Xpert MTB/RIF. We had insufficient data to apply the latent class model to data from the single study evaluating Xpert Ultra in lymph node aspirates.

Establishing a diagnosis of extrapulmonary tuberculosis would ideally include pursuing the diagnosis of pulmonary tuberculosis as well, because participants with tuberculosis may have both pulmonary and extrapulmonary tuberculosis and the lung may be the only site where the presence of tuberculosis can be established. Because of the difficulties involved in diagnosing HIV-associated tuberculosis, it is recommended that multiple cultures from sputum and other types of specimens be evaluated in people living with HIV[174,175]. Given the limitations in the reference standard, we recommend that future studies consider using liquid culture because this is more sensitive than solid culture, and that

researchers obtain multiple specimens for culture to confirm the diagnosis of extrapulmonary tuberculosis [176].

Most studies included in this review used culture-based DST (either Löwenstein-Jensen (LJ) or mycobacteria growth indicator tube (MGIT) 960) as the reference standard for detection of rifampicin resistance. Concerns have been raised about rapid DST methods, in particular automated MGIT 960, for tuberculosis drug resistance using the recommended critical concentrations. The WHO is planning to prioritise a re-evaluation of the critical concentrations for rifampicin (WHO 2018).

We assessed the number of specimens with nontuberculous mycobacteria (NTMs) that were Xpert Ultra-positive. In two studies that reported 120 NTMs, Xpert Ultra was negative in all specimens. In the previous version of this review, among 10 studies that reported information comprising 141 NTMs, Xpert MTB/RIF was negative in all specimens [168].

Quality and quality of reporting of the included studies

Risk of bias was low for the participant selection, index test, and flow and timing domains and was high or unclear for the reference standard domain (most of these studies performed specimen decontamination before culture inoculation). A limitation was that several studies included more than one specimen per participant, which artificially inflated the sample size of the study and may have led to overestimation or underestimation of the accuracy estimates. In general, studies were fairly well reported, although we corresponded with almost all primary study authors to ask for additional data and missing information. In several studies, accuracy data by site of extrapulmonary disease were not reported, and in a minority of studies, blinding was not reported. We strongly encourage the authors of future studies to follow the recommendations provided in the updated Standards for Reporting Diagnostic Accuracy (STARD) statement to improve the quality of reporting[177].

Interpretability of subgroup analyses

We investigated potential sources of heterogeneity in the different extrapulmonary specimens. Importantly, we found slightly higher Xpert Ultra accuracy in studies with concentrated cerebrospinal fluid (CSF) in comparison to unconcentrated specimens. We note that subgroup findings should be interpreted with caution, as there were only three studies and a small number of tuberculous meningitis cases included in these analyses.

Comparison with other systematic reviews

We are aware of several systematic reviews previously published on this topic that estimated summary accuracy of Xpert MTB/RIF for distinct forms of extrapulmonary tuberculosis, as well as different forms of extrapulmonary tuberculosis combined (Table 8). We identified one systematic review that estimated summary accuracy of Xpert Ultra that found, for all forms of extrapulmonary tuberculosis combined, pooled sensitivity and specificity of 85.1% (95% CI 76.7 to 90.8%) and 95.7% (95% CI 87.9 to 98.6%), (7 studies; 1500 specimens)[178].

Compared with previous systematic reviews, our review extends the date of the search for potential studies for inclusion. Our strict inclusion criteria, e.g. excluding case-control studies, meant that some of the studies included in other reviews were excluded from ours.

Applicability of findings to the review question

For the participant selection domain, most studies had high or unclear concern for applicability because either participants were evaluated exclusively as inpatients in tertiary care or we were not sure about the clinical settings. We therefore cannot be sure about the applicability of our findings to primary care. Studies that take place in referral settings may include participants whose condition is more difficult to diagnose than are seen at lower levels of the health system. However, we recognize that classifying studies as primary, secondary, or tertiary care may not adequately account for differences in disease spectrum [167]. For the index and reference test domains, most studies had low concern for applicability.

Acknowledgements

The Academic Editors on this review were Dr Michael Eisenhut (Cochrane Infectious Diseases Group; CIDG) and Dr Mia Schmidt-Hansen (Cochrane Diagnostic Test Accuracy).

We are grateful to Vittoria Lutje, the CIDG Information Specialist, for help with the search strategy. The CIDG editorial base is funded by UK aid from the UK government for the benefit of low- and middle-income countries (project number 300342-104). The views expressed do not necessarily reflect the UK government's official policies.

We acknowledge Eleanor Ochodo, Centre for Evidence-based Health Care Department of Global Health, Stellenbosch, and Fredrick Haraka, Ifakara Health Institute, Bagamoyo, Tanzania, for the section on people-important outcomes. We thank Hannah Ryan, from the Tropical and Infectious Diseases Unit, Royal Liverpool University Hospital, for her help with the protocol. We thank Chunli Lu, from the Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, and Marcela Perlwitz, Ivy Tech Community College of Indiana, West Lafayette, for help with translation. We thank all authors of the included studies for providing answers to our questions along with additional data. We thank the referees for their helpful comments.

Appendix for manuscript

I. Detailed search strategies

MEDLINE (OVID)
1 Mycobacterium tuberculosis/
2 Tuberculosis/ or "Tuberculosis, Multidrug-Resistant"/ or Extensively Drug-Resistant Tuberculosis/
3 (Tuberculosis or MDR-TB or XDR-TB or "Multidrug Resistant Tuberculosis" or "Extensively Drug Resistant Tuberculosis" or tuberculous).ti. ab .
4 (extrapulmonary or extra-pulmonary or EPTB).ti. ab .
5 (lymphadenitis or disseminated or miliary or pleur* or skeletal or spine or mening* or intracranial or intra-ocular or ocular or abdominal or splenic or genitourinary or pericardial).ti. ab .
6 "Tuberculosis, Central Nervous System"/ or "Tuberculosis, Urogenital"/ or "Tuberculosis, Splenic"/ or "Tuberculosis, Spinal"/ or "Tuberculosis, Renal"/ or "Tuberculosis, Pleural"/ or "Tuberculosis, Osteoarticular"/ or "Tuberculosis, Oral"/ or "Tuberculosis, Ocular"/ or "Tuberculosis, Meningeal"/ or "Tuberculosis, Lymph Node"/ or "Tuberculosis, Laryngeal"/ or "Tuberculosis, Hepatic"/ or "Tuberculosis, Gastrointestinal"/ or "Tuberculosis, Female Genital"/ or "Tuberculosis, Endocrine"/ or "Tuberculosis, Cutaneous"/ or "Tuberculosis, Cardiovascular"/ or Tuberculosis, Miliary/ or Tuberculosis, Male Genital/
7 1 or 2 or 3
8 4 or 5
9 7 and 8
10 9 or 6
11 Xpert*.ti. ab .
12 (GeneXpert or cepheid).ti.ab .
13 (near* patient or near-patient).ti.ab
14 11 or 12 or 13
15 10 and 14
Embase (OVID)
1 Tuberculosis, Multidrug-Resistant/ or Extensively Drug-Resistant Tuberculosis/ or Tuberculosis/ or tuberculosis.mp. or Mycobacterium tuberculosis/
2 (MDR-TB or XDR-TB).mp.
3 1 or 2
4 (extrapulmonary or extra-pulmonary or EPTB).ti. or (extrapulmonary or extra-pulmonary or EPTB).ab.
5 (lymphadenitis or disseminated or miliary or pleur* or skeletal or spine or mening* or intracranial or intra-ocular or ocular or abdominal or genitourinary or pericardial).ti. or (lymphadenitis or disseminated or miliary or pleur* or skeletal or spine or mening* or intracranial or intra-ocular or ocular or abdominal or genitourinary or pericardial).ab.
6 tuberculous.ti. or tuberculous.ab.
7 3 or 6
8 Tuberculosis, Central Nervous System/ or Tuberculosis, Hepatic/ or Tuberculosis, Male Genital/ or Tuberculosis, Spinal/ or Tuberculosis, Cutaneous/ or Tuberculosis, Urogenital/ or Tuberculosis, Osteoarticular/ or Tuberculosis, Endocrine/ or Tuberculosis, Renal/ or Tuberculosis, Splenic/ or Tuberculosis, Ocular/ or Tuberculosis, Laryngeal/ or Tuberculosis, Gastrointestinal/ or Tuberculosis/ or Tuberculosis, Meningeal/ or Tuberculosis, Oral/ or Tuberculosis, Pleural/ or Tuberculosis, Lymph Node/ or Tuberculosis, Female Genital/ or Tuberculosis, Miliary/ or Tuberculosis, Cardiovascular/
9 4 or 5 or 8
10 7 and 9
11 xpert*TB.mp.
12 Xpert* MTB RIF.ti. or Xpert* MTB RIF.ab.

13 (GeneXpert or cepheid).ti. or (GeneXpert or cepheid).ab.
14 (near* patient or near-patient).ti. or (near* patient or near-patient).ab.
15 12 or 13 or 14
16 10 and 15
Indexes=SCI-EXPANDED, CPCI-S, Biosis previews

TOPIC

(tuberculosis or tuberculous) *AND* **TOPIC:** (extrapulmonary or extra-pulmonary or EPTB or lymphadenitis or disseminated or miliary or pleur* or skeletal or spine or mening* or intracranial or intra-ocular or ocular or abdominal or genitourinary or pericardial) *AND* **TOPIC:** (Xpert* or Genexpert or cepheid)

LILACS

tuberculosis or tuberculous [Words] and Xpert$ or Genexpert or cepheid [Words]

SCOPUS

(TITLE-ABS-KEY (tuberculosis OR tuberculous) AND TITLE-ABS-KEY (extrapulmonary OR extra-pulmonary OR eptb OR lymphadenitis OR disseminated OR miliary OR pleur* OR skeletal OR spine OR mening* OR intracranial OR intra-ocular OR ocular OR abdominal OR genitourinary OR pericardial) AND TITLE-ABS-KEY (xpert* OR genexpert OR cepheid))

Cochrane Infectious Diseases Group Specialist Register; ClinicalTrials.gov, WHO ICTRP, ISRCTN registry, ProQuest Dissertations & Theses A&I

tuberculosis and Xpert$; tuberculosis and Genexpert; tuberculosis and Cepheid.

II. Data extraction form

Data extractor	MK KRS
First study author	
Corresponding study author and email	
Title of paper	
Journal	
Language if other than English	
Year	

I. Study details

Type of study: randomized controlled trial; cross-sectional cohort (with follow-up); case-control (exclude); unclear/not reported

Study data collection: prospective; retrospective; unclear/not reported

Participant selection: convenience; consecutive; random; other; unclear/not reported

Country:

Country income status: low; middle; high

II. Presenting signs and symptoms, setting

Presenting signs and symptoms?

Clinical setting: inpatient; outpatient; both; unclear/not reported

Level of laboratory running Xpert? peripheral; intermediate; central (reference)

Comments, describe exclusions

(Tests at laboratory levels)

Peripheral: AFB (Ziehl-Neelsen, Auramine-rhodamine, Auramine-O staining) and Xpert MTB/RIF

Intermediate: peripheral laboratory tests and culture on solid media and line probe assay (LPA) from smear-positive sputum

Central: intermediate laboratory tests and culture on liquid media and DST (first- and second-line anti-TB drugs) on solid or in liquid media and LPA on positive cultures and rapid speciation tests

III. Other demographics

HIV patients included? yes; no; unclear/not reported; if yes ## and percentage? (*denominator is number tested, when possible*)

Age? Median age in years (IQR); mean (SD); range; unclear/not reported

Children (< 15 years old) included: yes; no; unclear/not reported; if yes, percentage?

Percentage female included? Unclear/not reported

Past history of TB? yes; no; unclear/not reported; if yes, percentage?

Only patients who received TB treatment for ≤ 7 days were included? yes; no; unclear/not reported; if no, percentage on treatment included?

IV. Reference standard

A. Reference standard for TB detection

Solid culture (specify): LJ 7H10 7H11; other

Liquid culture (specify): MGIT Bactec 460; other

Solid and liquid culture (indicate which kind above)

Were reference standard results interpreted without knowledge of index test results? yes; no; unclear/not reported

B. Composite reference standard for pleural TB

Solid culture (specify): LJ 7H10 7H11; other

Liquid culture (specify): MGIT Bactec 460; other

Solid and liquid culture (indicate which kind above)

Histopathology (specify): granulomas; caseating granulomas
Were reference standard results interpreted without knowledge of index test results? yes; no; unclear/not reported
Did all patients receive the same reference standard? yes; no; unclear/not reported; if no, describe
C. Reference standard for rifampicin resistance
LJ DST MGIT DST MTBDR*plus*
Were reference standard results interpreted without knowledge of index test results? yes; no; unclear/not reported

V. Sites with more than five specimens (check all that apply)

A. Lymph node TB fluid; tissue; both fluid and tissue
B. Pleural TB fluid; tissue; both fluid and tissue
C. TB meningitis CSF
D. Bone or joint TB fluid; tissue; both fluid and tissue
E. Genitourinary TB urine; other, specify
F. Peritoneal TB fluid; tissue; both fluid and tissue
G. Pericardial TB fluid; tissue; both fluid and tissue
H. Disseminated TB blood
I. Other, specify

VI. Specimen processing for Xpert

Condition of specimens: fresh frozen
If frozen for > 7 days, indicate WHO not followed
For a given site, how many specimens were collected per patient? one; multiple; unclear/not reported
A. Lymph node tissue, other tissue
Was the WHO standard operating procedure (SOP) followed for each specimen type?
1a. Lymph node tissue WHO followed: yes; no; unclear
1b. Lymph node tissue homogenization step for tissue specimens: yes; no; unclear/not reported
2a. Other tissue, specify WHO followed: yes; no; unclear
2b. Other tissue homogenization step for tissue specimens: yes; no; unclear/not reported
(For tissue, if WHO SOP not followed, briefly describe specimen processing in comments.)

WHO SOPs for specimen processing; lymph node and other tissue; sterile specimen

Cut the tissue specimen into small pieces in a sterile mortar.
Add approximately 2 mL of sterile phosphate buffered saline (PBS).
Grind solution of tissue and PBS until homogeneous suspension has been obtained.
Place approximately 0.7 mL of the homogenized tissue in a sterile, conical screw-capped tube.
Double volume of specimen with Xpert® Sample Reagent (1.4 mL Sample Reagent to 0.7 mL of homogenized tissue).
Shake tube vigorously 10 to 20 times or vortex for at least 10 seconds.
Incubate specimen for 10 minutes at room temperature, and again shake specimen 10 to 20 times or vortex for at least 10 seconds.
Incubate specimen at room temperature for an additional 5 minutes.
Transfer 2mL to Xpert® MTB/RIF cartridge.
Load into GeneXpert and per manufacturer's instructions.

(Note: For specimens **not collected in a sterile manner,** WHO SOP suggests an NaOH decontamination/concentration protocol similar to that used for sputum.)

B. CSF
3a. CSF WHO followed: yes; no; unclear
3b. CSF concentration step: yes; no; unclear/not reported
3c. CSF sample input volume: specify; unclear/not reported
(For CSF, if WHO SOP not followed, briefly describe specimen processing in comments.)

WHO SOPs for CSF

If more than 5 mL of CSF is available for testing.

Transfer all of the CSF specimen to a conical centrifuge tube and concentrate the specimen at 3000 × g for 15 minutes.
Resuspend the pellet to a final volume of 2 mL by adding Xpert® MTB/RIF Sample Reagent.
Transfer 2 mL of the resuspended CSF sample to the Xpert® MTB/RIF cartridge.
Load the cartridge into the GeneXpert instrument according to the manufacturer's instructions.
If 1 mL to 5 mL of CSF is available.
Add an equal volume of Sample Reagent to the CSF.
Mix the specimen and the Sample Reagent by vortexing as described above. After seven to eight minutes at room temperature, vortex the sample as above a second time.
Incubate for an additional seven to eight minutes (15 minutes total incubation) at room temperature.
Add 2 mL of the sample mixture directly to the Xpert® MTB/RIF cartridge.
Load the cartridge into the GeneXpert instrument according to the manufacturer's instructions.
C. Body fluids, other than CSF
4a. Body fluid: specify; processed as per manufacturer for sputum
Yes/No/Unclear
4b. Body fluid: specify; sample input volume: specify; unclear/not reported
5a. Body fluid: specify; processed as per manufacturer for sputum (WHO followed)
Yes/No/Unclear
5b. Body fluid: specify; sample input volume: specify; unclear/not reported
(Add additional specimens as needed.)
(For body fluids other than CSF, if manufacturer's instructions not followed, briefly describe specimen processing in comments.)
Manufacturer's instructions for sputum
Raw specimen
Pour or pipette (pipette not provided) approximately 2 times the volume of Sample Reagent into the specimen (2:1 dilution, Sample Reagent: specimen).
Shake vigorously 10 to 20 times or vortex for at least 10 seconds.
Incubate sample for a total of 15 minutes at 20°C to 30°C.
Between 5 and 10 minutes into the incubation period, shake vigorously 10 to 20 times or vortex for at least 10 seconds.
Specimen sediment
Assay requires at least 0.5 mL of resuspended specimen sediment after digestion, decontamination, and concentration.
Use the method of Kent and Kubica and resuspend the sediment in a 67 mM phosphate/H_2O buffer.
After resuspension, keep at least 0.5 mL of the resuspended sediment for the Xpert® MTB/RIF assay.
Add 1.5 mL of Sample Reagent to 0.5 mL of the resuspended sediment (3:1 dilution, Sample Reagent: specimen)
Follow steps 2 to 4 above.
Comments on specimen processing.
VII. Specimen processing for culture
Specimen collected from sterile site: Yes/No/Unclear
Specimen processed for culture as per American Thoracic Society Diagnostic Standards? Yes/No/Unclear
(ATS guidelines: specimens collected from normally sterile sites may be placed directly into the culture medium.)
Note: All specimens such as CSF, pleura, lymph node aspirates and tissues, peritoneal fluid, pericardial fluid, bone or joint fluid and tissue, and urine are considered sterile.
VIII. Results
TB detection: number error or invalid or both Xpert® MTB/RIF results over total number of cultures performed. The denominator includes contaminated cultures and results that were inconclusive.
Unclear/not reported.
RIF resistance: number indeterminate Xpert results (over total number of cultures performed).

Unclear/not reported.
Non-tuberculous mycobacteria (NTM): number of cultures with NTM (over total number of cultures performed).
Unclear/not reported.
IX. Tables
(Non-tuberculous mycobacteria (NTM) should be included as not TB)
Tuberculosis detection (example for Xpert Ultra against culture reference standard; provide additional tables Xpert MTB/RIF and for other extrapulmonary specimens; extract trace results for Xpert Ultra)

Xpert Ultra in lymph node aspirate		**Definite TB**		
		Yes	**No**	**Total**
Xpert Ultra result	**Positive**			
	Negative			
	Total			
	Error/invalid			

Rifampicin resistance detection (for all culture-positive, extrapulmonary specimens)

Rifampicin resistance detection		**Rifampicin resistance**		
		Yes	**No**	**Total**
Xpert result	**Positive**			
	Negative			
	Total			
	Indeterminate			

Abbreviation: TB: tuberculosis.

III. Rules for QUADAS-2

Domain 1: patient selection
Risk of bias: could the selection of patients have introduced bias?
Signalling question 1: was a consecutive or random sample of patients enrolled?
We scored "yes" if the study enrolled a consecutive or random sample of eligible patients, "no" if the study selected patients by convenience, and "unclear" if the study did not report the manner of patient selection or we could not tell.
Signalling question 2: was a case-control design avoided?
We did not include in the review studies using a case-control design because this study design, especially when used to compare results in severely ill patients versus those in relatively healthy individuals, may lead to overestimation of accuracy in diagnostic studies. We scored "yes" for all studies.
Signalling question 3: did the study avoid inappropriate exclusions?
We scored "yes" if the study included both smear-positive and smear-negative specimens or included only smear-negative specimens. We judged "no" if the study included only smear-positive specimens or excluded specimens based on physical appearance (such as purulence) or a biochemical analysis (e.g. adenosine deaminase (ADA), cytology (cell analysis)). We scored "unclear" if we could not tell.
Applicability: are there concerns that the included patients and setting do not match the review question?
We were interested in how the index tests performed in patients presumed to have extrapulmonary tuberculosis who were evaluated as they would be in routine practice. We scored "low concern" if patients were evaluated at local hospitals or primary care centres. We scored "high concern" if patients were evaluated

exclusively as inpatients at tertiary care centres, except for tuberculous meningitis (we judged low concern) where we would expect patients to be evaluated in tertiary care settings. We scored "unclear concern" if the clinical setting was not reported or if information was insufficient to allow a decision. We also scored "unclear concern" if Xpert testing was done at a reference laboratory and the clinical setting was not reported for the following reason. It was difficult to tell if a given reference laboratory provided services mainly to very sick patients (inpatients in tertiary care) or to all patients, including very sick patients and those with less severe disease (primary, secondary, and tertiary care).

Domain 2: index test

Risk of bias: could the conduct or interpretation of the index test have introduced bias?

Signalling question 1: were the index test results interpreted without knowledge of results of the reference standard?

We answered this question "yes" for all studies because Xpert test results are automatically generated and the user is provided with printable test results. Thus, there is no room for subjective interpretation of test results.

Signalling question 2: If a threshold was used, was it pre-specified?

As the threshold is pre-specified in all versions of Xpert, we answered this question "yes" for all studies.

Applicability: are there concerns that the index test, its conduct, or its interpretation differ from the review question?

We note that variations in execution of the test might affect accuracy estimates. We judged "low concern" if the test was performed according to WHO standard operating procedures (WHO 2014), or if the index test was performed as recommended by the manufacturer for sputum. We scored "high concern" if the test was performed in a way that deviated from these recommendations. We scored "unclear concern" if we could not tell. In studies that evaluated several different types of specimens, we used the following rule: if ≥ 75% of the specimen types were processed per WHO standard operating procedure (SOP) or as per the manufacturer's instructions, we judged "low concern"; if < 50% of the specimen types were processed per WHO SOP or as per the manufacturer's instructions, we scored "high concern"; and if at least 50% to 74% of the specimen types were processed per WHO SOP or as per the manufacturer's instructions, or if we could not tell, we scored "unclear concern".

Domain 3: reference standard

Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias?

We considered this domain separately for the reference standard for detection of extrapulmonary tuberculosis and the reference standard for detection of rifampicin resistance.

Signalling question 1: is the reference standard likely to correctly classify the target condition?

For detection of extrapulmonary tuberculosis, culture is generally considered the best reference standard. However, limitations are associated with culture; bacterial load is usually low in extrapulmonary tuberculosis, leading to a reduction in the sensitivity of culture. Concerning the conduct of the reference standard (preparation of the specimen for culture), N-acetyl-L-cysteine-sodium hydroxide is routinely used to homogenize, decontaminate, and liquefy non-sterile specimens for TB culture (American Thoracic Society 2000). However, CSF, pleural fluid, and lymph node aspirates are usually considered sterile, and standards specify, "specimens collected from normally sterile sites may be placed directly into the culture medium" (American Thoracic Society 2000). Overly processing (sterile) specimens with N-acetyl-L-cysteine-sodium hydroxide may lead to a decrease in viable TB bacteria and consequently false-negative cultures. We scored "yes" if studies did not use N-acetyl-L-cysteine-sodium hydroxide for processing specimens and "unclear" if studies used N-acetyl-L-cysteine-sodium hydroxide. We discussed this further under Discussion and Strengths and weaknesses of the review.

For detection of rifampicin resistance, culture-based drug susceptibility testing (DST, also called conventional phenotypic method) is considered to be the best reference standard. Line probe assays are also WHO-recommended tests for rifampicin resistance. As we extracted data only for studies that used culture-based DST or line probe assays (most often MTBDRplus), we answered this question "yes" for all studies.

Signalling question 2: were the reference standard results interpreted without knowledge of results of the index test?

We scored "yes" if the reference test provided an automated result (e.g. MGIT 960), if blinding was explicitly stated, or if it was clear that the reference standard was performed at a separate laboratory and/or was performed by different people. We scored "no" if the study stated that the reference standard result was interpreted with knowledge of the Xpert Ultra or Xpert MTB/RIF test result. We scored "unclear" if we could not tell.

Signalling question 3: (rifampicin resistance) were the reference standard results interpreted without knowledge of results of the index test?

We added a signalling question for rifampicin resistance detection. We scored "yes" if the reference test provided an automated result (e.g. MGIT 960), if solid culture was performed followed by speciation, if blinding was explicitly stated, or if it was clear that the reference standard was performed at a separate laboratory or was performed by different people, or both. We scored "no" if the study stated that the reference standard result was interpreted with knowledge of the Xpert test result. We scored "unclear" if we could not tell. Not all studies evaluated detection of rifampicin resistance; therefore this question was not applicable to all studies.

Applicability: are there concerns that the target condition as defined by the reference standard does not match the question?

We judged "high concern" if included studies did not speciate mycobacteria isolated in culture, "low concern" if speciation was performed, and "unclear concern" if we could not tell. If a study performed sequencing, we considered the speciation yes. If the study only used a composite reference standard, we considered applicability unclear.

Domain 4: flow and timing

Risk of bias: could the patient flow have introduced bias?

Signalling question 1: was there an appropriate interval between the index test and the reference standard?

In most included studies, we expected that specimens for index test and verification by culture (or a composite reference standard) would be obtained at the same time, when patients were evaluated for presumptive extrapulmonary tuberculosis. However, even if there were a delay of several days between index test and reference standard, tuberculosis is a chronic disease, and we considered misclassification of disease status to be unlikely, as long as treatment was not initiated in the interim. We judged "yes" if the index test and the reference standard were performed at the same time or if the time interval was less than or equal to seven days, "no" if the time interval was greater than seven days, and "unclear" if we could not tell.

Signalling question 2: did all patients receive the same reference standard?

For the diagnosis of any form of extrapulmonary tuberculosis we answered this question "yes" if all participants in the study or a subset of participants in the study (for whom we extracted data) received the acceptable reference standard either culture or a composite reference standard. Regarding culture, we acknowledge that it is possible that some specimens could undergo solid culture and others liquid culture as the reference standard. This could potentially result in variations in accuracy, but we think the variation would be minimal.

For rifampicin resistance detection, we answered "yes" if all participants received the same reference standard (either culture-based DST or MTBDR*plus*), "no" if not all participants received the same reference standard, and "unclear" if we could not tell.

Signalling question 3: were all patients included in the analysis?

We determined the answer to this question by comparing the number of patients enrolled with the number of patients included in the 2 × 2 tables. We answered "yes" if the numbers matched and "no" if there were patients enrolled in the study who were not included in the analysis. We answered "unclear" if we could not tell.

Judgements for overall ‘Risk of bias' assessments.

If we answered all signalling questions for a domain "yes", then we scored risk of bias as "low".

If we answered all or most signalling questions for a domain "no", then we scored risk of bias as "high".

If we answered only one signalling question for a domain "no", we discussed further the "risk of bias" judgement.

If we answered all or most signalling questions for a domain "unclear", then we scored risk of bias as "unclear".
If we answered only one signalling question for a domain "unclear", we discussed further the "risk of bias" judgement for the domain

IV. OpenBugs for analyses

In this section we provide OpenBUGS models for the bivariate meta-analysis as well as the latent class meta-analysis. Any alternative prior distributions used are provided in the comments within each model.

BIVARIATE MODEL ASSUMING PERFECT CULTURE REFERENCE TEST

```
model {
for(i in 1:N) { # N is the number of studies
############################## LIKELIHOOD
logit(TPR[i]) <- l[i,1]
logit(FPR[i]) <- -l[i,2]
pos[i]<-TP[i]+FN[i]
neg[i]<-TN[i]+FP[i]
TP[i] ~ dbin(TPR[i],pos[i])
FP[i] ~ dbin(FPR[i],neg[i])
se[i] <- TPR[i]
sp[i] <- 1-FPR[i]
l[i,1:2] ~ dmnorm(mu[1:2], T[1:2,1:2])

}
############################## HYPER PRIOR DISTRIBUTIONS
mu[1] ~ dnorm(0,0.25) # replaced by mu[1] ~ dnorm(0, 0.01) in sensitivity analysis to check impact of less informative prior
mu[2] ~ dnorm(0,0.25) # replaced by mu[2] ~ dnorm(0, 0.01) in sensitivity analysis to check impact of less informative prior
T[1:2,1:2]<-inverse(TAU[1:2,1:2])

#### BETWEEN-STUDY VARIANCE-COVARIANCE MATRIX
TAU[1,1] <- tau[1]*tau[1]
TAU[2,2] <- tau[2]*tau[2]
TAU[1,2] <- rho*tau[1]*tau[2]
TAU[2,1] <- rho*tau[1]*tau[2]
tau[1]<-pow(prec[1],-0.5) # replaced by tau[1] ~ dunif(0,3) in sensitivity analysis to check impact of less informative prior
tau[2]<-pow(prec[2],-0.5) # replaced by tau[2] ~ dunif(0,3) in sensitivity analysis to check impact of less informative prior
sigma.sq[1] <- pow(tau[1], 2)
sigma.sq[2] <- pow(tau[2], 2)

#### prec = between-study precision in the logit(sensitivity) and logit(specificity)

prec[1] ~ dgamma(2,0.5) # replaced by prec[1] <- 1/pow(tau[1],-2) in sensitivity analysis to check impact of less informative prior
prec[2] ~ dgamma(2,0.5) # replaced by prec[2] <- 1/pow(tau[2],-2) in sensitivity analysis to check impact of less informative prior
rho ~ dunif(-1,1)
```

```
############################## OTHER PARAMETERS OF INTEREST

#### POOLED SENSITIVITY AND SPECIFICITY

Pooled_S<-1/(1+exp(-mu[1]))
Pooled_C<-1/(1+exp(-mu[2]))

#### POOLED POSITIVE AND NEGATIVE LIKELIHOOD RATIOS

PLR <- Pooled_S/(1-Pooled_C)
NLR <- (1-Pooled_S)/Pooled_C

#### PREDICTED SENSITIVITY AND SPECIFICITY IN A FUTURE STUDY

l.new[1:2] ~ dmnorm(mu[],T[,])
sens.new <- 1/(1+exp(-l.new[1]))
spec.new <- 1/(1+exp(-l.new[2]))

} #### END OF PROGRAM
```

LATENT CLASS META-ANALYSIS MODEL

```
# WinBUGS PROGRAM FOR ESTIMATING A BIVARIATE HIERARCHICAL META-ANALYSIS
MODEL
# FOR SENSITIVITY AND SPECIFICITY ALLOWING FOR HETEROGENEITY BETWEEN STUDIES

model {

############################## ##############################

for(i in 1:N) {# N is the number of studies

############################## LIKELIHOOD
logit(p[1, i]) <- l[i,1]
logit(p[2, i]) <- -l[i,2]


prob[i,1] <- pi[i]*(p[1,i]* s2[i] + covp[i]) + (1-pi[i])*(p[2,i]*(1-c2[i]) + covn[i])
prob[i,2] <- pi[i]*(p[1,i]* (1-s2[i]) - covp[i]) + (1-pi[i])*(p[2,i]*c2[i] - covn[i])
prob[i,3] <- pi[i]*((1-p[1,i])*s2[i] - covp[i]) + (1-pi[i])*((1-p[2,i])*(1-c2[i]) - covn[i])
prob[i,4] <- pi[i]*((1-p[1,i])*(1-s2[i]) + covp[i]) + (1-pi[i])*((1-p[2,i])*c2[i] + covn[i])

n[i] <- sum(cell[i,1:4])
cell[i,1:4] ~ dmulti(prob[i,1:4],n[i])

pi[i] ~ dbeta(1,1)

se[i] <- p[1,i]
sp[i] <- 1-p[2,i]


l[i,1:2] ~ dmnorm(mu[1:2], T[1:2,1:2])
```

```
#============================================================
# CONDITIONAL DEPENDENCE
#============================================================

#=====================================
# upper limits of covariance parameters
#=====================================
us[i]<-min(se[i],s2[i])-(se[i]*s2[i]);
uc[i]<-min(sp[i],c2[i])-(sp[i]*c2[i]);

ls[i]<- -(1-se[i])*(1-s2[i])
lc[i]<- -(1-sp[i])*(1-c2[i])


#============================================================
# prior distribution of transformed covariances on (0,1) range
#============================================================
covp[i]~dunif(ls[i],us[i]);
covn[i]~dunif(lc[i],uc[i]);
#covn[i]<-0


}

# ==================================
# NON-INFORMATIVE HIERARCHICAL PRIOR DISTRIBUTION OVER REF STD PROPERTIES
# ==================================

for(j in 1:29) {

logit(s2[j]) <- l2[j,1]
logit(c2[j]) <- l2[j,2]
l2[j,1:2] ~ dmnorm(mu2[1:2], T2[1:2,1:2])

}

############################## HYPER PRIOR DISTRIBUTIONS ##############################

#############################################################
#############################################################
###
### XPERT TEST
###
#############################################################
#############################################################

mu[1] ~ dnorm(0,0.25)
mu[2] ~ dnorm(0,0.25) #dnorm(4.59512,10)
T[1:2,1:2]<-inverse(TAU[1:2,1:2])
```

```
#### BETWEEN-STUDY VARIANCE-COVARIANCE MATRIX
TAU[1,1] <- tau[1]*tau[1]
TAU[2,2] <- tau[2]*tau[2]
TAU[1,2] <- rho*tau[1]*tau[2]
TAU[2,1] <- rho*tau[1]*tau[2]

tau[1]<-pow(prec[1],-0.5)
tau[2]<-pow(prec[2],-0.5)

sigma.sq[1] <- pow(tau[1], 2)
sigma.sq[2] <- pow(tau[2], 2)

#### prec = between-study precision in the logit(sensitivity) and logit(specificity)
prec[1] ~ dgamma(2,0.5)
prec[2] ~ dgamma(2,0.5)
rho ~ dunif(-1,1)

############################## OTHER PARAMETERS OF INTEREST
#### POOLED SENSITIVITY AND SPECIFICITY OF XPERT
Pooled_S<-1/(1+exp(-mu[1]))
Pooled_C<-1/(1+exp(-mu[2]))

#### POOLED POSITIVE AND NEGATIVE LIKELIHOOD RATIOS
PLR <- Pooled_S/(1-Pooled_C)
NLR <- (1-Pooled_S)/Pooled_C

#### PREDICTED SENSITIVITY AND SPECIFICITY OF XPERT IN A FUTURE STUDY
l.new[1:2] ~ dmnorm(mu[],T[,])
sens.new <- 1/(1+exp(-l.new[1]))
spec.new <- 1/(1+exp(-l.new[2]))


############################################################
############################################################
###
### CULTURE TEST
###
############################################################
############################################################


mu2[1] ~ dnorm(0,0.25)
mu2[2] ~ dnorm(0,0.25)
T2[1:2,1:2]<-inverse(TAU2[1:2,1:2])

#### BETWEEN-STUDY VARIANCE-COVARIANCE MATRIX
TAU2[1,1] <- tau2[1]*tau2[1]
TAU2[2,2] <- tau2[2]*tau2[2]
TAU2[1,2] <- rho2*tau2[1]*tau2[2]
TAU2[2,1] <- rho2*tau2[1]*tau2[2]
```

```
tau2[1] <-pow(prec2[1],-0.5)
tau2[2] <-pow(prec2[2],-0.5)

sigma.sq2[1] <- pow(tau2[1], 2)
sigma.sq2[2] <- pow(tau2[2], 2)

#### prec = between-study precision in the logit(sensitivity) and logit(specificity)
prec2[1] ~ dgamma(2,0.5)
prec2[2] ~ dgamma(2,0.5)
rho2 ~ dunif(-1,1)



#### POOLED SENSITIVITY AND SPECIFICITY OF CULTURE
S2<-1/(1+exp(-mu2[1]))
C2<-1/(1+exp(-mu2[2]))

s2.new <- 1/(1+exp(-ls2.new))
c2.new <- 1/(1+exp(-lc2.new))
ls2.new ~ dnorm(mu2[1],prec2[1])
lc2.new ~ dnorm(mu2[2],prec2[2])

}
```

Chapter 4- Manuscript

Diagnostic accuracy of centralised assays for TB detection and detection of resistance to rifampicin and isoniazid: a systematic review and meta-analysis

Mikashmi Kohli [1,2], Emily MacLean [1,2], Madhukar Pai [1,2], Samuel G. Schumacher [3,5] and Claudia M. Denkinger[3,4,5]

[1] Dept of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada.
[2] McGill International TB Centre, McGill University, Montreal, QC, Canada. [3] Foundation for Innovative New Diagnostics, Geneva, Switzerland. [4] Centre for Infectious Diseases, University Heidelberg, Heidelberg, Germany. [5]

Published in European Respiratory Journal 2021 Feb4;57(2):2000747.

Introduction

Tuberculosis, caused by Mycobacterium tuberculosis complex (MTBC), has surpassed HIV/AIDS as the world's leading infectious cause of death. The World Health Organization (WHO) estimates that, in 2018, 10 million people became ill with tuberculosis, and approximately 1.45 million died of the disease. In 2018, only half of all confirmed tuberculosis patients underwent drug susceptibility testing[1]. The introduction and rollout of nucleic acid amplification tests (NAATs) has significantly improved the area of tuberculosis diagnosis by providing rapid tuberculosis and drug resistance detection. The principal behind these assays is amplification of a targeted region of the M. tuberculosis genome by PCR. NAATs are used for both tuberculosis detection (particularly the Xpert MTB/RIF) and identification of mutations that confer resistance to anti-tuberculosis drugs (for example, Bruker-Hain and Nipro line probe assays (LPAs), most commonly rifampicin (RIF) and isoniazid (INH))[179,180] Globally, INH mono-resistant tuberculosis is more prevalent than multidrug-resistant tuberculosis (MDR-TB), and WHO guidelines advocate for universal testing for both RIF and INH resistance before commencing tuberculosis treatment [181]. Recently, several companies have developed molecular tests for tuberculosis and RIF/INH resistance detection on centralised platforms, many of which have already been established as multi-disease platforms, primarily for detection of HIV, human papillomavirus and hepatitis C virus. This systematic review intended to evaluate the diagnostic accuracy of five of these tests for M. tuberculosis and RIF/INH resistance detection to assess their diagnostic accuracy. The tests included were Abbott RealTime MTB, Abbott RealTime RIF/INH, FluoroType MTB, FluoroType MTDBR and BD Max MDR-TB assay.

Methods

Search strategy, information sources and eligibility criteria

We followed standard guidelines and methods for systematic review and meta-analyses of diagnostic test accuracy[68,182]. A comprehensive search of databases (PubMed, EMBASE, BIOSIS, Web of Science, LILACS, Cochrane) for relevant citations, without language restrictions was performed. An example search strategy is provided in the supplementary methods. The time period was restricted to January 2009 to June 2018 and another scoping search was done till May 2020 to look for published studies for these platforms. We also contacted the developers of these tests to provide available data and lists of studies they are aware of. Cross-sectional, case–control, cohort studies and randomised

controlled trials of any of the index tests (listed above) were included if at least 25 specimens were tested. Abstracts and unpublished studies were excluded. Patients of all age groups with presumed or confirmed pulmonary tuberculosis or MDR-TB, in all settings and any country, were included. Our search strategy also included terms for two assays by Roche and Bioneer that are comparable to the assays reviewed, however, we did not find any studies for these assays.

Citation screening and study selection

Two authors (M. Kohli and E. MacLean) independently screened and reviewed the full texts. Any discrepancies were resolved by discussion, and in case of disagreement, a third author was consulted (C.M. Denkinger). If a study contributed data to more than one analysis (e.g. two different index tests in one study), it was considered as two or more datasets. Disagreements in extracted information were resolved by discussion with third author (C.M. Denkinger). Study authors were contacted in cases of missing data. In cases of papers without extractable diagnostic accuracy data, the study was excluded if after three attempts the study author did not reply.

Reference standards

For tuberculosis detection, solid or liquid culture was the reference standard. For resistance detection, phenotypic drug susceptibility testing (DST) was the primary reference standard. However, if the studies provided information on sequencing, we analysed the data using a phenotypic DST reference standard, a sequencing reference standard, and a composite reference standard (CRS). For a CRS, if phenotypic DST showed drug sensitivity but sequencing identified mutations recognised to be associated with resistance, the CRS was considered resistant when the mutations were associated with high or moderate confidence of resistance as per Miotto et al[183]. If phenotypic DST showed resistance but sequencing did not identify mutations associated with resistance, the CRS was considered resistant (as mutations could be outside of the region sequenced).

Head-to-head comparisons

When possible, the index tests were also compared to other well-characterised, WHO-recommended molecular test: Xpert MTB/RIF for both TB detection and rifampicin resistance. Such head-to-head comparisons are preferred, as using a WHO-recommended comparator test with known diagnostic accuracy serves as an easily understood benchmark for the index test's performance[184]. It can allow

flagging of studies with particularly strong or weak results for the index test, which may help explain some between-study heterogeneity. Assessment of methodological quality The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool, a validated quality assessment tool for diagnostic studies [72], was used to assess the included studies' risk of bias. Statistical analysis and data synthesis For each index test, meta-analyses were performed of sensitivity and specificity of TB detection, as well as RIF resistance and INH resistance when at least four studies were available. Studies were pooled using bivariate random effect hierarchical models to calculate sensitivity and specificity, with associated 95% confidence intervals, of each index test against the relevant reference standard. When there were fewer than four studies for an index test or evident heterogeneity between studies, a descriptive analysis only was performed.

Results

From the literature search, 750 citations were identified, 81 full-text articles were reviewed, and 21 studies were included in the systematic review (see figure 4.1). The 21 studies contributed 26 datasets, as four provided data for more than one index test. All studies were conducted in central level laboratories, which was expected as these assays require sophisticated laboratory infrastructure and skilled laboratory workers. As most studies were laboratory-based, there were limited demographic data available, such as age, HIV status, and past tuberculosis history of the included patient population. Tables 4.1a and 4.1b show the results of all the index tests analysed separately for both tuberculosis detection and resistance detection. Table 4.2 provides data for head-to-head comparisons of the index tests with Xpert MTB/RIF.

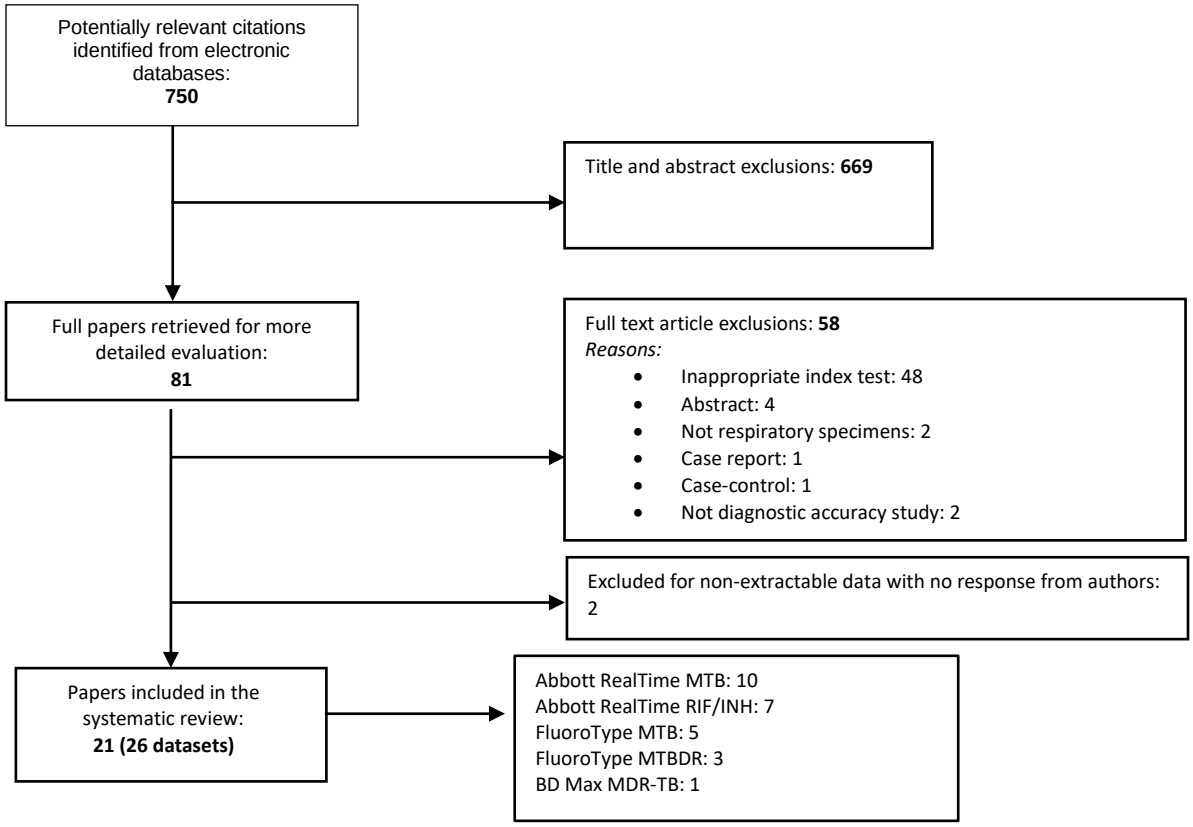


Figure 4.1: PRISMA diagram of included studies.

Table 4.1a. Diagnostic accuracy of each index test- TB detection

Index test	Smear status	# Datasets (#specimens)	Sensitivity (95% CI)	Specificity (95% CI)
Abbott MTB	All	10 (4858)	96.2% (90.2-98.6)	97.1% (93.7-98.7)
	Positive	10 (765)	99.0% (97.7-100)	-
	Negative	10 (4056)	88.4% (74.0-95.3)	98.3% (96.3-99.2)
FluoroType MTB	All	5 (2660)	92.1% (87.6-93.3)	98.9% (64.0-99.9)
	Positive*	3 (174)	Range: 100% (92-100)	-
	Negative*	3 (1754)	Range: 30%-85%	Range:62%-98%
FluoroType MTBDR*	All	2 (782)	Range: 91%-96%	Range: 100% (97-100)
	Positive	2 (288)	Range: 98%-100%	-
	Negative	2 (494)	Range: 69%-98%	Range: 100% (97-100)
BD Max MDR-TB*	All	1 (892)	93% (89.0- 96.0)	97% (96.0- 98.0)
	Positive	1 (176)	100% (98–100)	-
	Negative	3 (713)	81% (73–88%)	98% (96–99%)

CI:Confidence interval; # = number of;
*These datasets were not analysed as the number of studies were less than four and could not be analyzed.

Table 4.1b. Diagnostic accuracy of each index test - resistance detection

Index test	# Datasets (# specimens)	Sensitivity (95% CI)	Specificity (95% CI)
Abbott RIF/INH			
Rifampicin resistance	7 (1008)	94% (89-99)	100% (99-100)
Isoniazid resistance	7 (1013)	89% (86-92)	99% (98-100)
FluoroType MTBDR*			
Rifampicin resistance	2 (231)	Range: 97%-99%	Range: 100% (85-100)
Isoniazid resistance	2 (207)	Range: 70%-92%	Range: 100% (84-100)
BD Max MDR-TB*			
Rifampicin resistance	1 (232)	90% (55-100)	95% (91-97)
Isoniazid resistance	1 (232)	82% (63–92)	100% (98-100)

CI: Confidence interval; # = number of

* These datasets were not meta-analyzed as the number of studies were less than four and could not be analyzed.

Table 4.2: Head to head comparisons of the index test with Xpert MTB/RIF

Index test	Smear status	# Datasets (# specimens)	Sensitivity (95% CI)	Specificity (95% CI)
Head to Head comparisons (Abbott RealTime MTB and Xpert MTB/RIF)				
Berhanu 2018				
Abbott RealTime MTB	All	1 (237)	79% (66-88)	97% (93-99)
Xpert MTB/RIF	All		82% (70-91)	100% (98-100)
Scott 2017				
Abbott RealTime MTB	All	1 (193)	85% (74-93)	92% (86-96)
Xpert MTB/RIF	All		92% (82-97)	98% (93-100)
Wang 2016				
Abbott RealTime MTB	All	1 (255)	100% (98-100)	84% (76-91)
Xpert MTB/RIF	All		97% (93-99)	90% (82-95)
Head to head comparisons (FluoroType MTB and Xpert MTB/RIF)				
Obasanya 2017				
FluoroType MTB	All	1 (296)	89% (79-95)	60% (53-66)
Xpert MTB/RIF	All		79% (67-87)	94% (90-97)
Head to head comparisons (BD Max MDR-TB and Xpert MTB/RIF)				
Shah 2019				
BD Max MDR-TB	All	1 (889)	91% (87-94)	96% (94-97)
Xpert MTB/RIF	All		90% (86-93)	98% (97-99)

Risk of bias by QUADAS-2 assessment

The overall methodological quality of the included studies for each index test is summarised in Supplementary Figures S1- S10. For all assays except BD Max MDR-TB, the studies had applicability concerns in the domain of participant selection, as the studies were not conducted in high TB or MDR-TB burden settings. Similarly, for risk of bias, some studies had concerns in the patient selection domain and also the reference standard domains. In all other domains, risk of bias was low.

Index tests

Abbott RealTime MTB

Ten studies with 4858 respiratory specimens were included in the meta-analysis that evaluated Abbott RealTi*me* MTB assay for TB detection [10-19] (Figure 4.2). In all studies, the assay was run directly on specimens, as opposed to positive culture isolates. Most studies (6/10) used fresh specimens, while four used frozen specimens. The median sample size was 389 (interquartile range [IQR]: 242 to 599). In individual studies, the sensitivity point estimates of Abbott MTB assay varied from 79% to 100% with specificity varied from 84% to 99% (Figure 4.2). Pooled sensitivity and specificity were 96.2% (95% Confidence Interval (CI): 90.2- 98.6) and 97.1% (95%CI: 93.7–98.7), respectively.

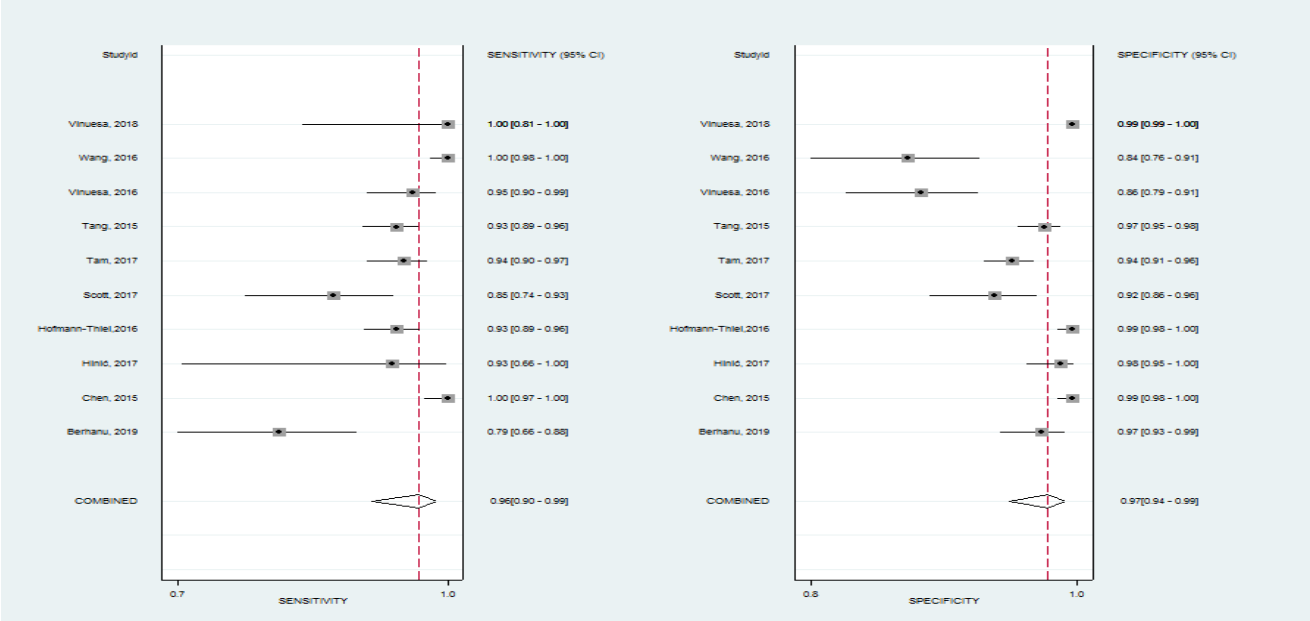


Figure 4.2: Forest plots for TB detection by Abbott RealTime MTB

Comparator test for TB detection: Xpert MTB/RIF

In addition to the RealTime MTB assay, three studies [185–187] performed Xpert MTB/RIF on the same specimens [185,187] or on different specimens obtained from the patient on the same visit [186] (Figure 3a). In the study by Wang et al. a lower overall specificity was observed for both Xpert (90%) and RealTime MTB (84%) than would be expected. In contrast, Scott et al, showed Xpert specificity of 97%, and specificity for RealTime MTB was 92% in the study. Berhanu et al also evaluated Xpert Ultra on same specimens. The study showed an increased sensitivity of 89%, but with a trade-off for lower specificity of 96% (Figure 3b).

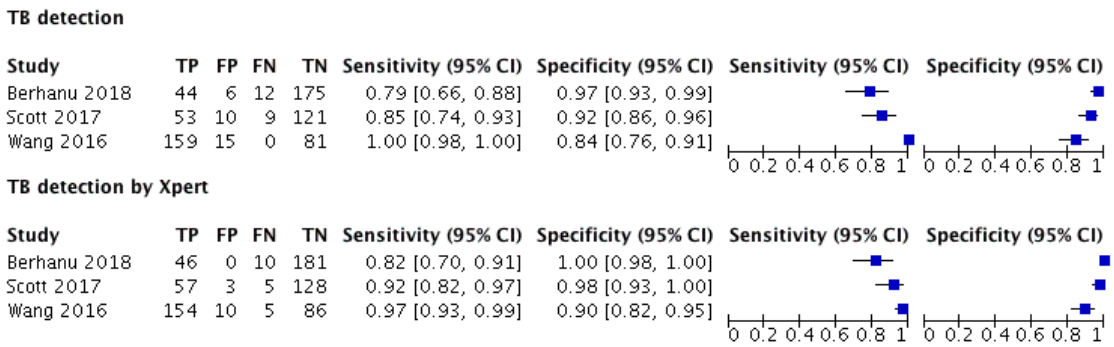


Figure 4.3a: Forest plot for TB detection with Abbott RealTime MTB assay and Xpert MTB/RIF with culture as reference standard

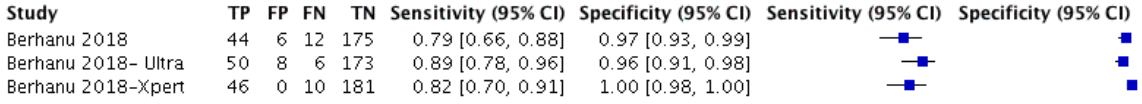


Figure 4.3b: Forest plot for TB detection with Abbott RealTime MTB assay, Xpert MTB/RIF and Xpert Ultra with culture as reference standard

Sub-group analyses: smear status

All ten studies provided data allowing stratification by smear status. For smear-positive specimens, the sensitivity of RealTime MTB assay varied from 95% to 100%. Pooled sensitivity was 99.0% (95%CI: 97.7-100) (10 studies, 765 specimens).

For smear-negative specimens, the sensitivity in these specimens varied from 41% to 100%. Pooled sensitivity was 88.4% (95%CI: 74.0-99.3) (10 studies, 4056 specimens). The study by Berhanu et al.[185] demonstrated very low sensitivity of 41.0% (95%CI: 18.0-67.0), which may partially be explained by the high prevalence of HIV in their study population, meaning that a high proportion of cases suffered from paucibacillary disease. We were not able to explore test performance by HIV status further as most of the studies (60%) did not report HIV prevalence.

Abbott RealTime MTB RIF/INH

Seven studies provided data for RIF and INH resistance detection by RealTime MTB RIF/INH, with phenotypic DST as the reference standard in both use cases [185,186,188–191]. Six studies performed the index test directly on known TB positive specimens or as an accompanying drug susceptibility test with RealTime MTB. One study [189] used TB positive culture isolates for the index test specimen. Four studies used fresh specimens while others used bio-banked specimens.

RIF resistance detection

The pooled sensitivity and specificity for RIF resistance were 94% (95%CI: 89.0-99.0) and 100% (95%CI: 99.0- 100), respectively, from seven studies and a total of 1008 specimens (Figure 4.4). There was little heterogeneity across studies.

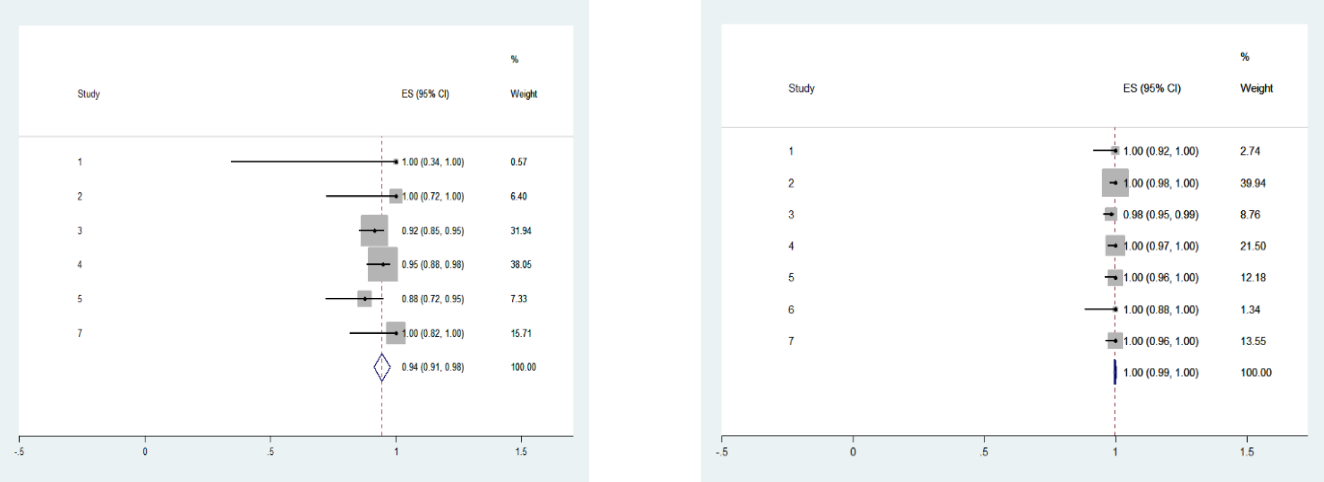


Figure 4.4: Forest plots for rifampicin resistance detection by Abbott RIF/INH assay using phenotypic DST as reference standard

Additionally, three studies provided sequencing data for RIF resistance, so we compared RealTime MTB RIF/INH performance against sequencing and a composite reference standard (CRS) in these instances (Figure S11). In the paper by Hoffman-Thiel et al., three specimens were classified as resistant by RealTime MTB RIF/INH due to a L511P mutation in the *rpoB* gene, but were sensitive on phenotypic DST [189]. These three specimens were reclassified as true positives with CRS. In the same study, 10 specimens that were susceptible to RIF by index test were resistant by both sequencing and culture (6 specimens with the high confidence mutation H526R and 4 with the moderate confidence mutation L533P mutations). In the paper by Kostera et al., four specimens were classified as susceptible wildtype by the index test and sequencing, but were classified as false negatives by the CRS, as phenotypically they were resistant to RIF [190]. In the smaller study by Tam et al., the index test and reference standards had complete concordance[192]. Thus overall, given the limited number of discordances between the phenotypic and genotypic DST, the results in reference to the different reference standards hardly changed (Figure S11).

INH resistance detection

For INH resistance detection, the pooled sensitivity and specificity were 89% (95%CI: 86.0-92.0) and 99% (95%CI: 98.0-100), respectively, from seven studies and a total of 1013 specimens (Figure 5). There was little heterogeneity across studies.

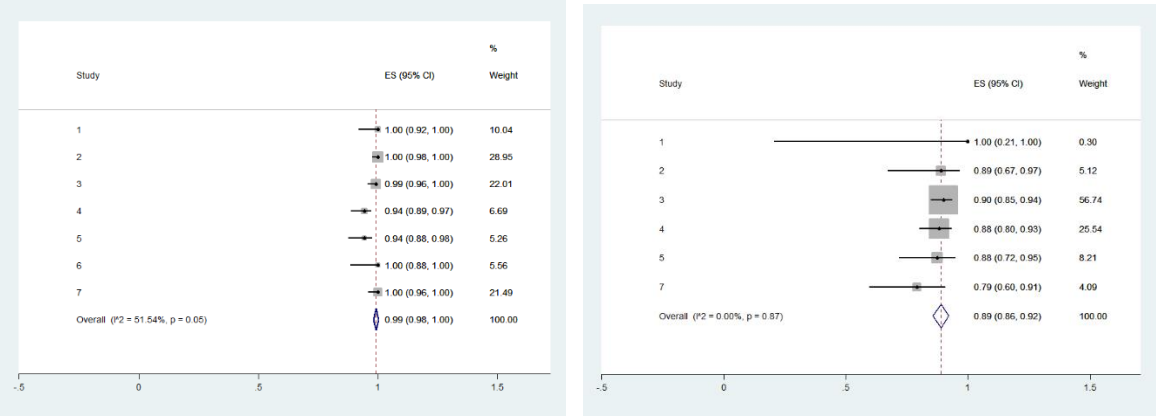


Figure 4.5: Forest plots for isoniazid resistance detection by Abbott RIF/INH assay using phenotypic DST as reference standard

The same three studies provided data for INH resistance against sequencing. RealTime MTB RIF/INH displayed better accuracy when compared against the sequencing reference standard than against the phenotypic DST. For Hofmann-Thiel, there were 18 specimens that were susceptible by index test but resistant by phenotypic DST[189]. These 18 specimens did not show any mutations in the *katG* or *inhA* target regions using sequencing, so by the CRS we classified them as resistant, since these mutations could have been outside the target regions. Hence the accuracy estimates with CRS in the study were identical to the phenotypic DST. In the study by Kostera et al. 2016[190], seven discordant specimens that were classified as susceptible phenotypically but INH resistant by index test were confirmed to be resistant by sequencing. This was due to the presence of the *katG* mutation S315T in three cases and an *inhA* protomer region mutation, c -15t, in four cases. These 7 specimens were correctly identified as resistant by the index test but were missed by conventional phenotypic DST (Figure S12).

FluoroType MTB

Five studies with 2660 respiratory specimens were included in the meta-analysis[188,193–196]. Median sample size was 608 (IQR: 296 to 661). The assay was performed directly on specimens in all studies, with all but one (4/5, 80%) studies reporting use of fresh specimens. One study used biobanked specimens[188]. Individual sensitivities ranged from 87% to 95%, while specificities ranged from 60% to 100% (Figure 6). Pooled sensitivity and specificity were 92.1% (95%CI: 87.6-93.3) and 98.9% (95%CI: 64.0-99.9), respectively. Obasanya et al observed relatively low specificity of 60% (95%CI: 53.0-66.0), which may be partially explained by the study being conducted in a low resource setting with higher potential for sample contamination, the use of Petroff's method for sputum decontamination, and Löwenstein-Jensen solid culture as the reference standard[196] .

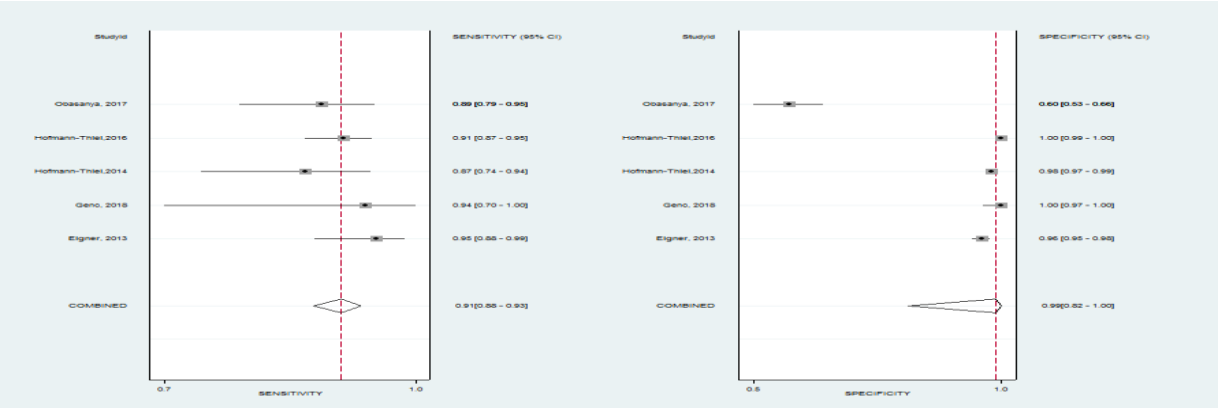


Figure 4.6: Forest plots for TB detection by FluoroType MTB assay

Comparator test for TB detection: Xpert MTB/RIF

In assessing Xpert as a comparator test in the same study[196], a substantially higher specificity was observed (94% for Xpert versus 60% for the FluoroType) (Figure 7). However, the specificity of Xpert was lower than the observed specificity of the test for PTB in a large meta-analysis[30]. The same study observed Xpert MTB/RIF sensitivity of 79% and FluoroType MTB sensitivity of 89%.

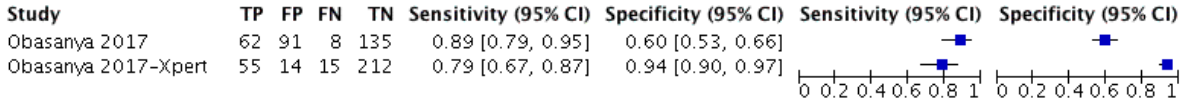


Figure 4.7: Forest plots of TB detection by FluoroType MTB and Xpert MTB/RIF

FluoroType MTBDR

Two studies[197,198] evaluated FluoroType MTBDR for TB detection using 782 frozen specimens (Table 2). The study by de Vos et al reported a sensitivity of 96% (95%CI: 93-98) and a specificity of 100% (95%CI: 97-100)[197]. Haasis et al reported a sensitivity of 91% (95%CI: 82-97) and specificity of 100% (95%CI: 98-100)[198]. The de Vos study only included Xpert-positive specimens, which could have introduced spectrum bias and an inflated sensitivity estimate.

RIF resistance detection

Two studies[198,199] assessed the performance of the test for RIF resistance detection using a phenotypic DST. Hillemann et al[199] used culture isolates for FluoroType MTBDR while Haasis performed the testing directly on specimens[198]. Sensitivity was 97% (95%CI: 82.0-100) for Haasis and 99% (95%CI: 96.0-100) for Hillemann and specificity was 100% in both studies. No comparison to sequencing was performed.

INH resistance detection

For isoniazid resistance detection, phenotypic culture was also the reference standard. In Haasis[29] and Hillemann[199], sensitivities were 70% (95%CI: 46.0-88.0) and 92% (95%CI: 84.0-97.0), respectively, and specificity was 100% in both studies. No comparison to sequencing was performed.
For the Hillemann et al[199] study, the use of culture isolates for testing might have resulted in better resistance detection than in Haasis et al[198].

BD Max MDR-TB

One recently published multicentre study provided data for this assay[37]. The assay was run on fresh sputum specimens. It reported a sensitivity of 93% (95%CI: 89.0-96.0) with specificity of 97% (95%CI: 96.0-98.0) on raw sputum specimens. For decontaminated sputum specimens, the sensitivity was 91% (95%CI: 87.0-94.0) and specificity was 95% (95%CI: 93.0-97.0).

Comparator test for TB detection: Xpert MTB/RIF

The study performed Xpert on the same processed sputum specimens as a comparator test. It reported similar sensitivities of 91% and 90% and specificities of 96% and 98% for BD Max and Xpert, respectively (Figure 8).

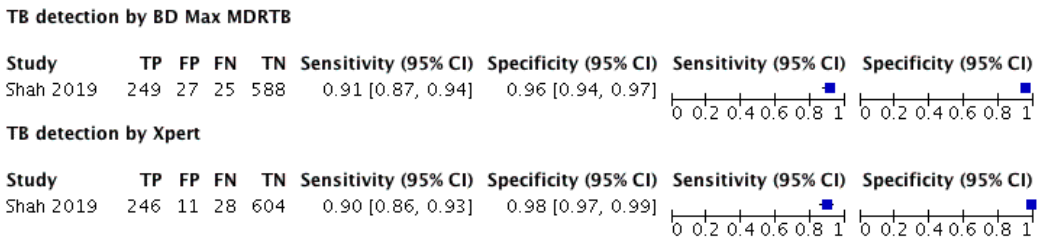


Figure 4.8: Forest plot of TB detection by BD Max MDR-TB and Xpert MTB/RIF

RIF resistance detection

For RIF resistance, the sensitivity and specificity with phenotypic DST as reference standard were 90% (95%CI: 55-100) and 95% (95%CI: 91-97), respectively (1 study, 232 specimens). However, six of eleven specimens classified as false positives by phenotypic DST had *rpoB* mutations identified by Sanger sequencing. Two specimens each had D516Y and L511P mutations, while one specimen each had D516F and L533P mutations, all of which are considered to confer resistance with high or moderate confidence[7]. Based on this reclassification, specificity increased from 95% (211/222) against phenotypic DST to 98% (211/216) with the sequencing and CRS reference standards[37].

INH resistance detection

For INH resistance, the sensitivity and specificity were 82% (95%CI: 63.0-92.0) and 100% (95%CI: 98.0-100), respectively, against phenotypic DST.

Discussion

In this systematic review, we summarise the performance of five diagnostic test for tuberculosis and RIF/ INH resistance detection: Abbott RealTime MTB, Abbott RealTime MTB RIF/INH, FluoroType MTB, FluoroType MTBDR and BD Max MDR-TB. Overall, the tests show similar performance to tests currently recommended by WHO. Sensitivity across tests was in the range of 90% and above with markedly low observed variability for all assays. For specificity in tuberculosis detection, there was more variability across studies and tests and further research needs to be conducted to understand whether this variability is related to test characteristics.

For some studies, accuracy estimates were low for both the index test and the comparator (Xpert), which helped in understanding that decreased accuracy could be due to some confounders or study characteristics not stated explicitly [187,196]. Contrastingly, other studies were well conducted and there was more confidence in the diagnostic accuracy of the index tests as the comparators had accuracy estimates which were in-line with WHO estimates[37,186]. Conceivably, the different tests might perform differently when it comes to detection of viable and non-viable bacteria depending on the extraction methods and the methods used to enrich whole cell bacteria (e.g. filters)[200,201]. Therefore, studies recruiting individuals with recent tuberculosis history that compare index tests to existing WHO recommended tests (such as Xpert MTB/RIF) would be useful. As well, manual extraction methods, such as those employed by OBASANYA et al. [196]for Fluorotype MTB, in the hands of less experienced users might have contributed to contamination and thus false positive results.

For RIF and INH resistance detection, the sensitivity and specificity estimates were also in the range of the published accuracy estimates for Xpert [30] and LPA [34]. Although data was limited and variability was observed, which might relate to how the tests were performed (e.g. from isolates or sample) or the study populations. Three assays were evaluated for the detection of RIF and INH resistance. Abbott RealTime RIF/INH assay was the only assay that had sufficient data for meta-analysis, with pooled sensitivity and specificity for RIF resistance of 94% and 100%, respectively, and for INH resistance 89% and 99%, respectively. For the other two assays, data was insufficient to meta-analyze, but overall diagnostic accuracy for RIF and INH resistance detection at this point appeared comparable to that of the WHO-recommended LPA test (90%)[34]. The use of CRS increased the specificity in some studies due to the identification of disputed mutations by sequencing that went undetected by phenotypic DST [183]. All studies that provided sequencing information performed targeted Sanger sequencing, which is a limitation as only targeted sequences can be identified, compared to whole genome sequencing which would provide information on the entire genome and thus identify resistance conferring mutations outside of target regions such as rpoB.

A concerning finding to be noted was that in a study [189] of RealTime MTB RIF/INH where six specimens were identified as susceptible to RIF by index test despite the presence of the high confidence mutation H526R. This finding needs to be further assessed in additional studies. S315T is

a frequent katG mutation and arises typically before all other drug mutations. It is also one of the mutations termed as "harbinger mutations". Its early detection may help in preventing multidrug resistance transmission[202]. In the current systematic review, Abbott RealTime RIF/INH assay picked up this mutation correctly in three specimens in comparison to the phenotypic DST [190]. There was insufficient data to assess these mutations in other assays included in the review. Only for the BD Max MDR-TB, a single well-conducted study provided information across a well-characterised and representative population. For other tests, HIV status, sex, tuberculosis history and tuberculosis treatment status were not available for 70% of the datasets included in the analyses, making generalisability to specific settings difficult. For these tests, additional studies are needed that provide more demographic information for the samples tested to allow for further generalisability of the data.

Operational characteristics are also a critical component for the use of testing platforms in different settings. The throughput of all of the mostly automated platforms assessed in this study is large. Specifically, the number of specimens that can be processed in these platforms vary from 24 (BD Max) to 94 specimens (Abbott RealTime, Hain Fluorotype, Roche Cobas). The turnaround time vary from 3 to 5 h as available from the company manufacturers' package inserts. All of the platforms can be connected to central laboratory information management systems, which is beneficial for disseminating reports to clinicians and patients without delay. Furthermore, the platforms are able to run a large portfolio of assays for different diseases, with Abbott having the largest among the tests evaluated. As such the assays are suited for centralised settings and can provide results to many patients with minimal hands-on manipulation. This limits infection risk to healthcare workers and laboratory technicians, as well as the risk of sample contamination. All tests demonstrated sensitivity for smear-negative cases comparable to Xpert MTB/RIF assay, making them good contenders for this frequently difficult-to-diagnose use case. However, the tests are not suited for use in lower levels of the healthcare system where patients first present for care. And for the platforms to have the same impact than near-patient platforms, specimen transport needs to optimised. In addition, without reliable systems in place to deliver test results to patients, the impact of these centralised platforms will be very limited, despite their high performance.

An important strength of our systematic review and meta-analysis was that we provided head-to-head comparisons of the index test with Xpert MTB/RIF, a WHO recommended molecular test [8].

Additionally, we also used multiple reference standards for evaluating drug resistance, which provided information on the mutations captured or missed by the index tests. However, the review and meta-analysis also had some limitations. As most of these tests are very new to market, there was minimal data to perform more detailed analyses. Most of the studies were laboratory-based studies, and therefore demographic data of the included participants were not provided. Thus, the generalisability of the performances of all tests (with the exception of BD Max) is uncertain. Another potential concern was that most of the studies had test manufacturers' involvement. In summary, for patients with pulmonary tuberculosis, these centralised molecular assays demonstrate promising diagnostic accuracy for tuberculosis, RIF resistance, and INH resistance detection. While data were limited, the performance of these assays appears similar to that of WHO-recommended Xpert and LPA assays. The assays might prove to have operational advantages in some settings, but further research is necessary.

Acknowledgments

We would like to thank Genevieve Gore, McGill Librarian for helping with the literature search for this review. We would also like to thank the Government of the United Kingdom for providing financial support.

Supplementary information

I.Search Strategy

Database: Embase <1996 to 2018 Week 26>

Search Strategy:

---

1 (rifampin* or rifampicin* or Isoniazid*).mp. (73472)

2 (MDR TB or MDRTB or RRTB or RR TB or DRTB or DR TB).mp. (4878)

3 exp tuberculosis/ or mycobacterium tuberculosis/ or tuberculosis control/ or rifampicin/ or isoniazid/ (189174)

4 (tubercul* or antitubercul* or tb).mp. (190575)

5 1 or 2 or 3 or 4 (231179)

6 (((Abbott or RealTime* or Real Time*) adj (mtb* or rif* or inh*)) or fluorotype* or bd max* or bdmax* or cobas* taqman* or bioneer*).mp. (1612)

7 *real time polymerase chain reaction/ (10598)

8 ((real time or realtime or rt or direct) and (pcr or polymerase chain reaction)).ti. (18450)

9 6 or 7 or 8 (22380)

10 5 and 9 (574)

11 limit 10 to yr="2009 -Current" (447)

12 limit 11 to dc=20171205-20180628 (17)

13 remove duplicates from 12 (17)

II. QUADAS-2 Assessment

Domain 1 Patient Selection:

Risk of Bias: Could the selection of patients have introduced bias?

- Signaling question 1: Was a consecutive or random sample of patients or specimens enrolled?
    - We scored ‘yes’ if the study enrolled a consecutive or random sample of eligible patients; ‘no’ if the study selected patients by convenience, and ‘unclear’ if the study did not report the manner of patient selection or this cannot be discerned.

- Signaling question 2: Was a case-control design avoided?
    - We scored ‘yes’ if the study enrolled only patients presumed of drug-resistant TB, including patients with confirmed TB. We scored ‘no’ if the study enrolled patients for whom resistance status was already known, and ‘unclear’ if the study did not report the design or this cannot be discerned.
- Signaling question 3: Did the study avoid inappropriate exclusions?
    - We scored ‘yes’ if no inappropriate exclusions were noted. We scored ‘no’ if studies note specific exclusions. Inappropriate exclusions could potentially occur if patients were excluded based on prior knowledge or testing about them or if the technician does not record performed test results but this was not anticipated for research studies in this review.

Applicability: Are there concerns that the included patients and setting do not match the review question?

We were interested in how the index tests (centralized molecular DST assays) performed in patients presumed of having TB who are evaluated. We judged ‘low concern’ when the specimens included in the study were from the patients with presumptive pulmonary TB and was conducted in high TB and/or high MDR-TB burden country as per the WHO list. We judged ‘high concern’ if the specimens were collected from patients in a low TB and/or MDR-TB burden country. We will judge ‘unclear concern’ if the study included specimens from both high and low TB/MDR-TB burden settings or we could not tell.

Domain 2: Index Test

Risk of Bias: Could the conduct or interpretation of the index test have introduced bias?

- Signaling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?
    - We scored ‘yes’ for all studies because all the centralized molecular DST assay results are automatically generated and the user is provided with printable test results. Thus, there was no room for subjective interpretation of test results.

- Signaling question 2: If a threshold was used, was it prespecified?
    - As the threshold is prespecified in all centralized molecular DST assay in this

review, we answered this question "yes" for all studies.

Applicability: Are there concerns that the index test, its conduct, or its interpretation differ from the review question? Variations in test technology, execution, or interpretation may affect estimates of the diagnostic accuracy of a test.

We judged ‘low concern’ if the test was done as per recommendation of the manufacturer for PTB specimens. We judged ‘high concern’ it was stated and/or if additional steps were used for sample preparation and ‘unclear concern’ if we could not tell.

Domain 3: Reference Standard

Risk of Bias: Could the reference standard, its conduct, or its interpretation have introduced bias?

- Signaling question 1: Is the reference standard likely to correctly classify the target condition?
    - For detection of TB, culture is generally considered the best reference standard. We scored ‘yes’ if the studies used MGIT 960 as the reference standard (higher quality reference standard). We scored ‘no’ if the studies used only solid media-based culture (lower quality reference standard) as all these index tests are for centralized settings, we expect the laboratory settings to have liquid culture for detecting TB. LJ culture has lower diagnostic accuracy than liquid culture and would over or under-estimate the diagnostic accuracy of the index test. We scored ‘unclear’ if we could not tell.
    - For detection of rifampicin resistance, culture-based drug susceptibility testing (DST, also called conventional phenotypic method) is considered to be the best reference standard. As we extracted data for studies that used culture-based DST, we will score “ yes” for all studies.

- Signaling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?
    - We scored 'yes' if the reference test provided was culture e.g. MGIT 960 DST where an automated result is generated (except for LJ with confirmation of

MTB by a NAAT-based test), if blinding was explicitly stated, or if it was clear that the reference standard was performed at a separate laboratory and/or performed by different people. We will score ‘no’ if the study stated that the reference standard was interpreted with knowledge of the index test result. We scored 'unclear' if this was not stated or answered inadequately.

- Signaling question 3: (Rifampicin resistance) Were the reference standard results interpreted without knowledge of the results of the index test?
    - We added a signaling question for rifampicin resistance detection. We scored "yes" if the reference test provided an automated result (for example, MGIT 960), blinding was explicitly stated, or it was clear that the reference standard was performed at a separate laboratory or performed by different people, or both. We scored "no" if the study stated that the reference standard result was interpreted with knowledge of the index test result. We scored "unclear" if we could not tell.

Applicability: Are there concerns that the target condition as defined by the reference standard does not match the question?

We judged applicability to be of ‘low concern’ for all studies.

Domain 4: Flow and Timing

Risk of Bias: Could the patient flow have introduced bias?

- Signaling question 1: Was there an appropriate interval between the index test and reference standard?
    - We scored ‘‘yes’ if the tests were paired or separated by less than 48 hours after treatment initiation. We scored ‘no’ if the reference and index tests were not performed on paired specimens or were separated by more than a week. We scored ‘unclear’ if this was not stated in the paper or answered inadequately. In the majority of included studies, we expected specimens for index tests and culture to be obtained at the same time (i.e. to be performed on paired specimens for the majority of studies), when patients are presumed of having TB or MDR-TB.

- Signaling question 2: Did all patients receive the same reference standard?
    - For the diagnosis of TB, we scored this question "yes" if all participants in the study or a subset of participants in the study (for whom we will extract data) received the acceptable reference standard (solid culture, liquid culture, or both), which we specified as a criterion for inclusion in the review. However, we acknowledge that it is possible that some specimens could undergo solid culture and others liquid culture as the reference standard. This variation was recorded.
    - For rifampicin resistance detection, we scored "yes" if all participants received the same reference standard (either culture-based DST or MTBDR*plus*), "no" if not all participants received the same reference standard, and "unclear" if we could not tell.

- Signaling question 3: Were all patients included in the analysis?
- The answer to this question was determined by comparing the number of patients enrolled with the number of patients included in the two-by-two tables. We noted if authors record the number of indeterminate results. We scored ‘yes’ if the number of participants enrolled was clearly stated and corresponded to the number presented in the analysis or if exclusions were adequately described. We scored 'no’ if there were participants missing or excluded from the analysis and there was no explanation given; and 'unclear ' if not enough information was given to assess whether participants were excluded from the analysis

III. QUADAS-2 summaries ─ Risk of bias and applicability concerns

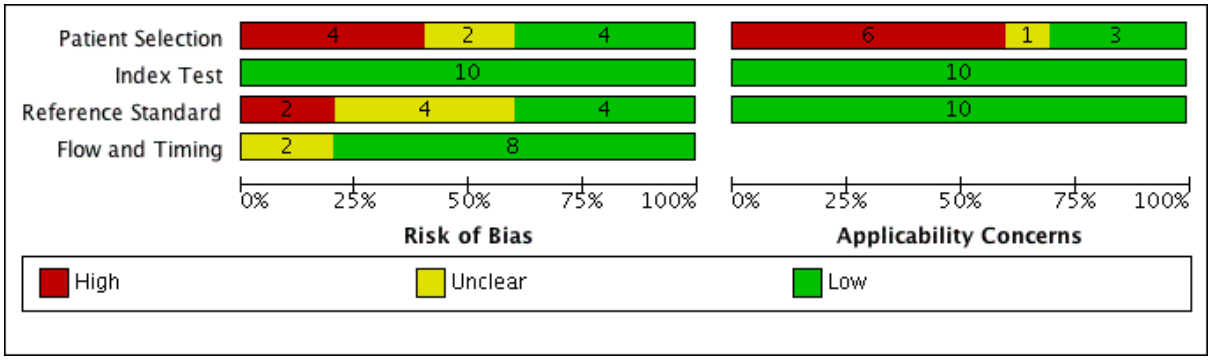


Figure S1. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain for Abbott RealTime MTB assay

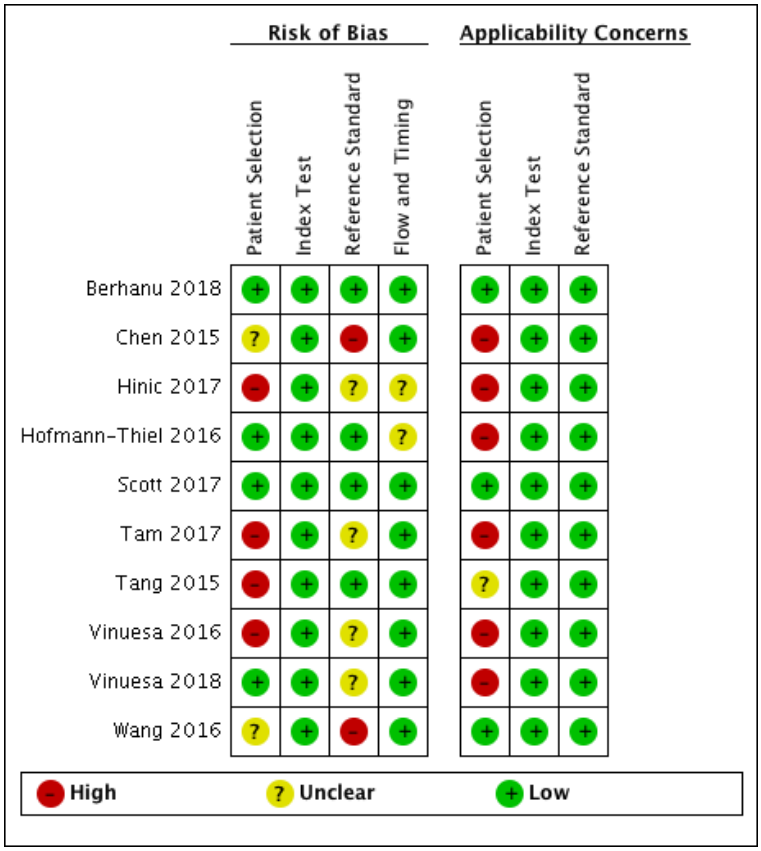


Figure S2. Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating Abbott RealTime MTB assay

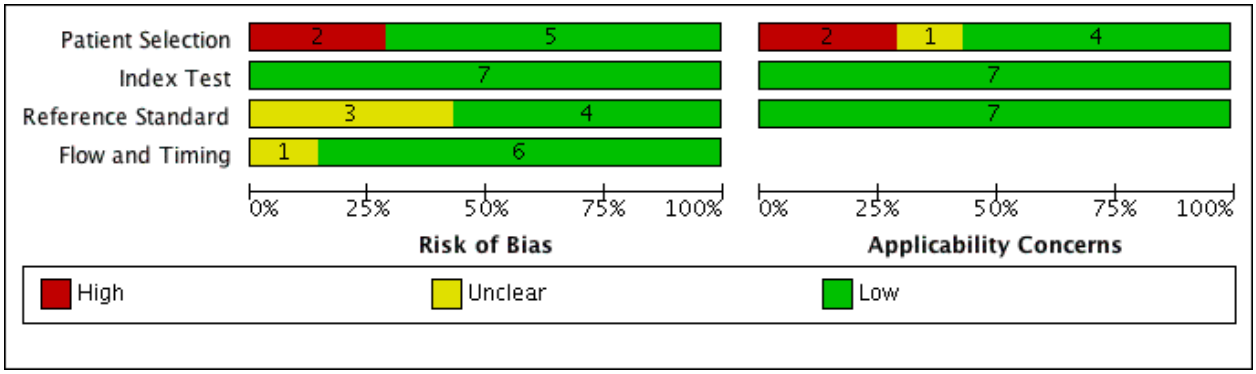


Figure S3. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for Abbott RealTime MTB RIF/INH assay

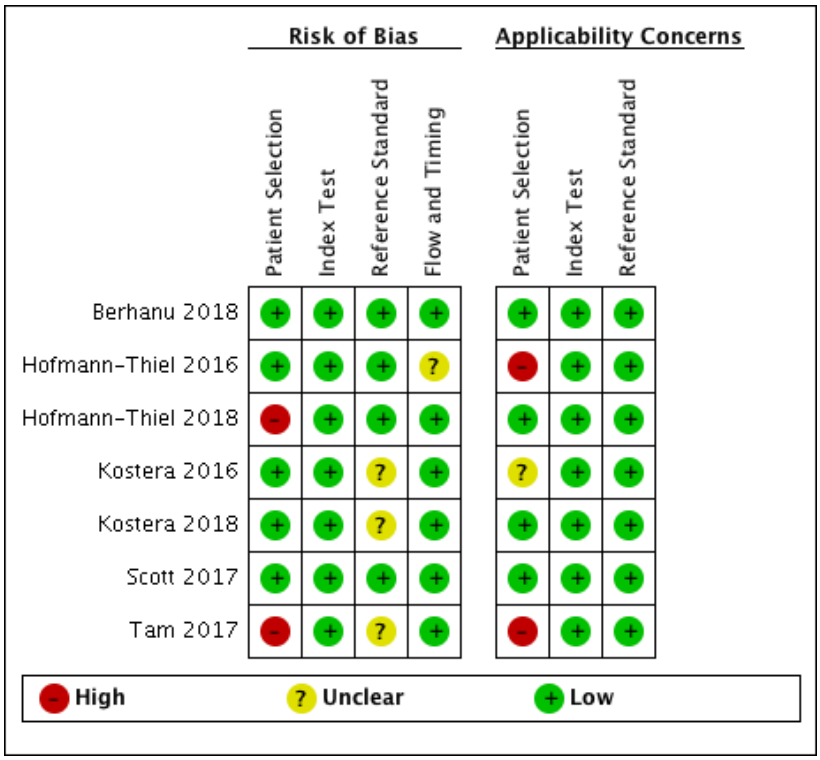


Figure S4. Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating Abbott RealTime MTB RIF/INH assay

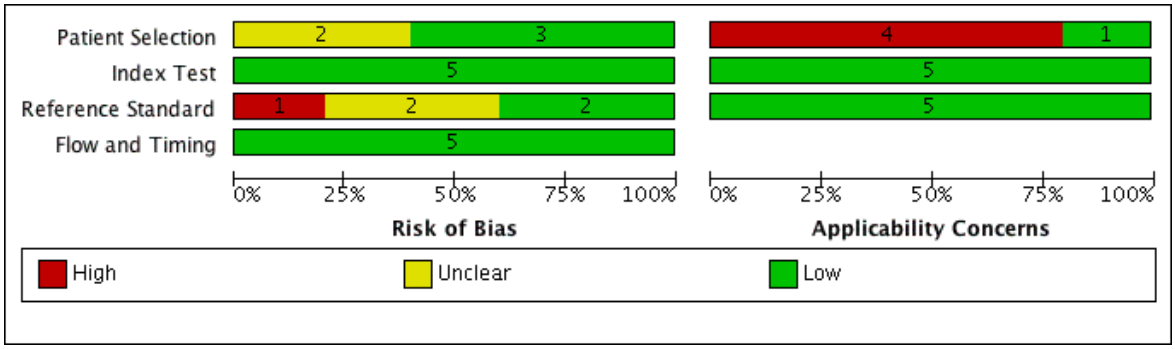


Figure S5. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for FluoroType MTB assay

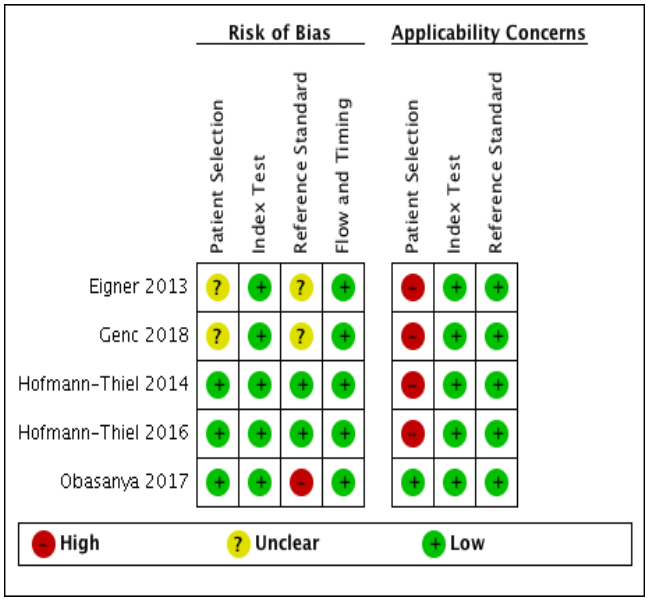


Figure S6: Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating FluoroType MTB assay

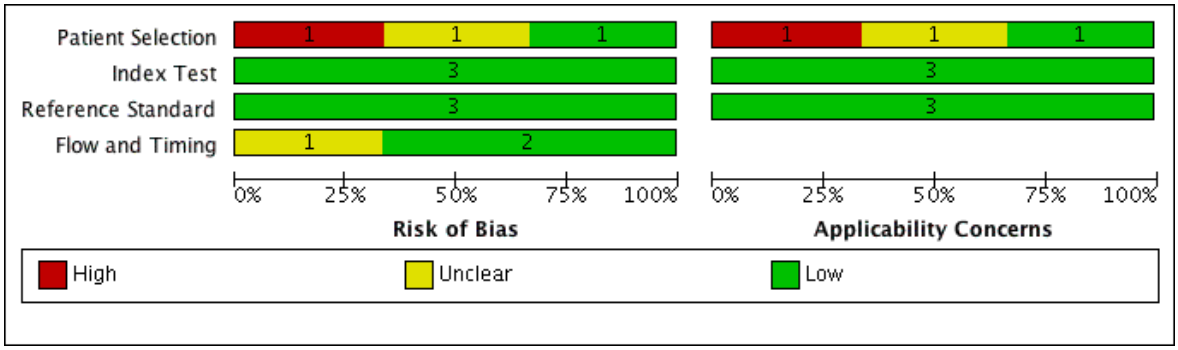


Figure S7. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for FluoroType MTBDR assay

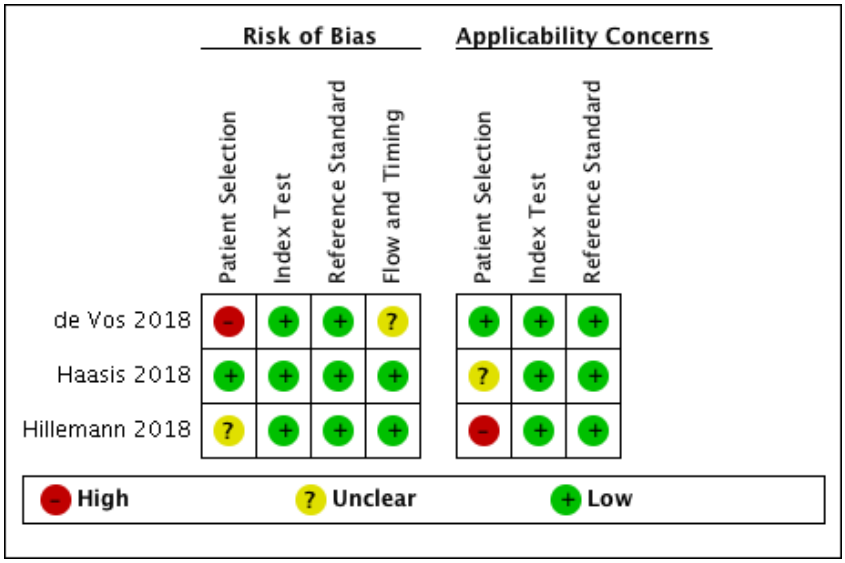


Figure S8. Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating FluoroType MTBDR assay

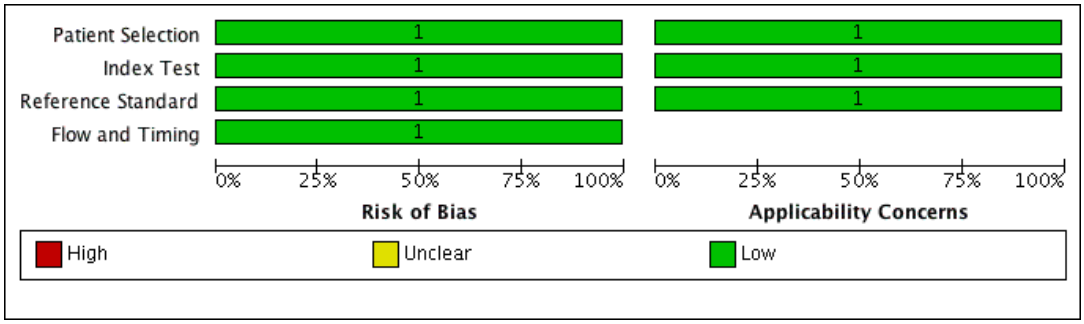


Figure S9: Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for BD Max MDR-TB assay

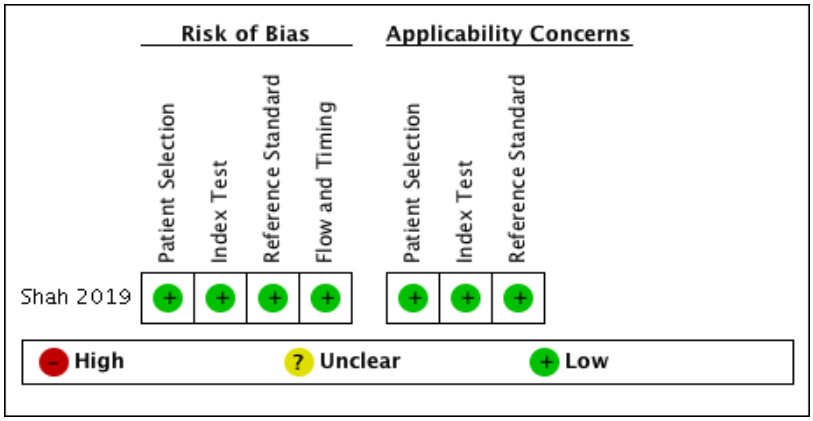


Figure S10: Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating BD Max MDR-TB assay

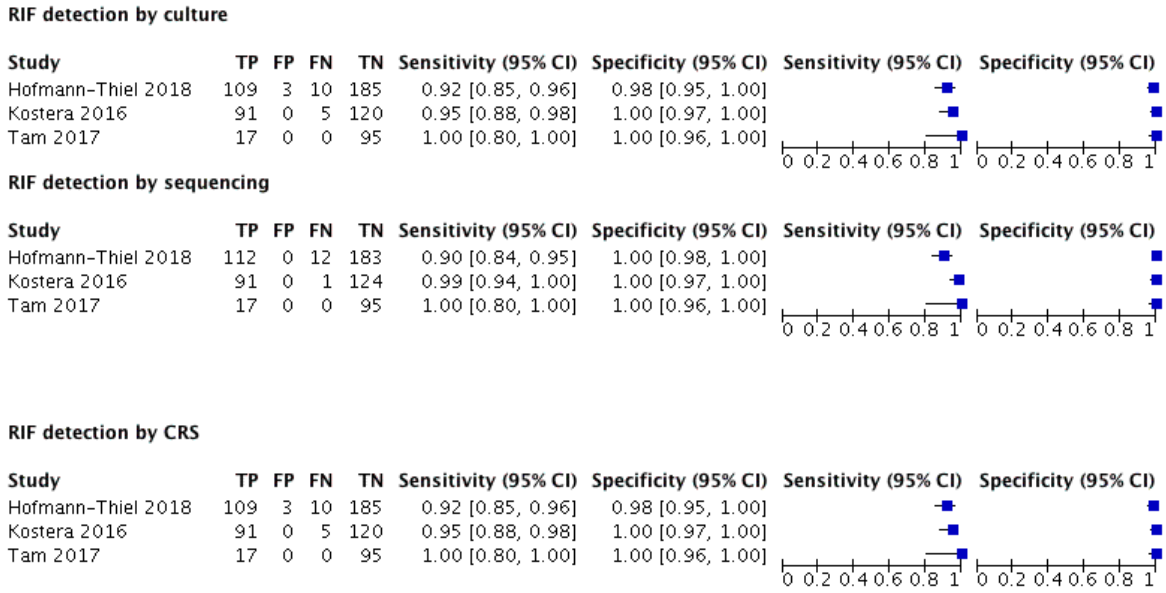


Figure S11: Forest plots for rifampicin resistance detection by Abbott RIF/INH assay using phenotypic DST, sequencing and composite reference standard

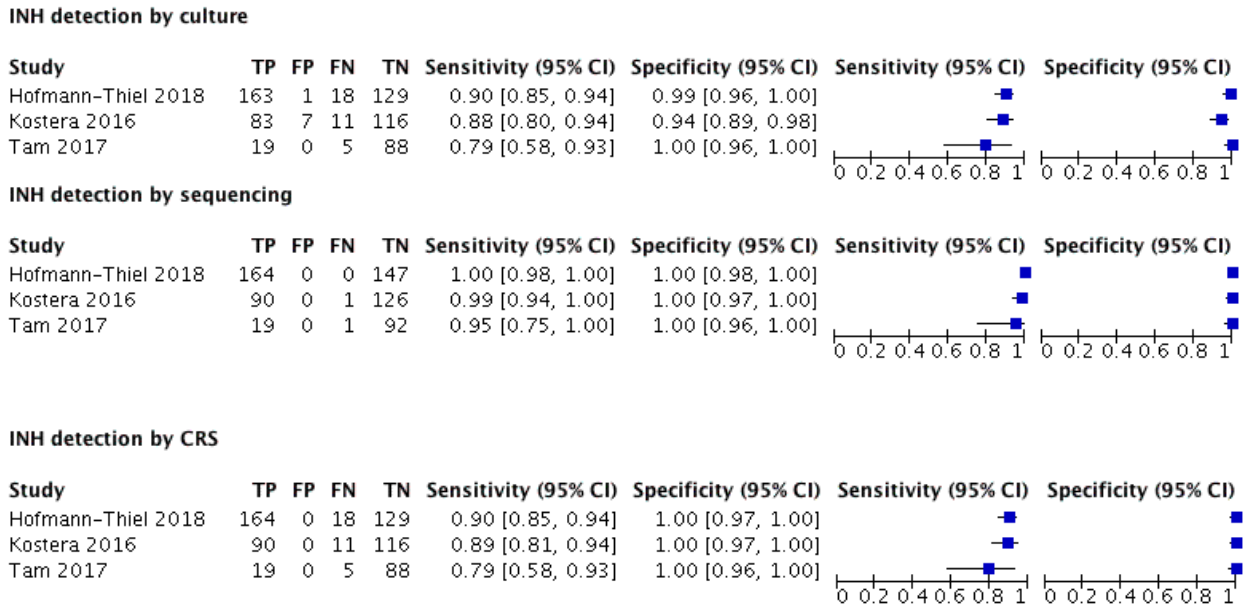


Figure S12: Forest plots for isoniazid resistance detection by Abbott RIF/INH assay using phenotypic DST, sequencing and composite reference standard

The political declaration at the first United Nations high-level meeting (UNHLM) on tuberculosis (TB) held on 26 September 2018 included commitments by Member States to achieve several global targets[203]. One of these targets was to diagnose and treat 40 million people with TB in the 5-year period 2018–2022. By end of 2021, only 26 million people were treated. The next UNHLM is set for 2023 and will set new global targets for TB detection and treatment. Sadly, due to 3 years of the pandemic, and slow progress even before the pandemic, none of the UNHLM or End TB targets are likely to be met (Figure 5.1).

Figure 5.1: UN Global targets to End TB

Source: WHO, Global TB Report 2022

The proportion of notified cases that are bacteriologically confirmed needs to be urgently increased. The microbiological detection of TB is critical as it allows people to be correctly diagnosed and started on the most effective treatment regimen as clinical assessment of the disease alone can lead to incorrect diagnosis leading to unnecessary treatment. The aim should be to increase the percentage of TB cases confirmed bacteriologically, based on scaling up the use of WHO recommended rapid diagnostics that are more sensitive than smear microscopy.

WHO has endorsed a range of new diagnostic technologies during the past 10 years[22]. The amplification and detection of *M. tuberculosis* complex (MTBC) nucleic acids is a technology that has proven to be highly sensitive and specific. During the Covid-19 pandemic, all countries rapidly scaled up their ability to perform molecular (PCR) testing for SARS-CoV2, and there is no reason to believe this capacity cannot be used for improving TB detection.

Systematic reviews and meta-analyses provide robust evidence on diagnostic performance of these tests. The two systematic reviews on low complexity and moderate complexity assays in this thesis were used to develop WHO guidelines and recommendations for use of these NAATs for TB detection[22].

In people presumed to have extrapulmonary tuberculosis, Xpert Ultra and Xpert MTB/RIF may be helpful in confirming the diagnosis. Sensitivity varies across different extrapulmonary specimens, while for most specimens specificity is high, the test rarely yielding a positive result for people without tuberculosis. For tuberculous meningitis, Xpert Ultra had higher pooled sensitivity and lower pooled specificity than Xpert MTB/RIF against culture. Xpert Ultra and Xpert MTB/RIF had similar sensitivity and specificity for rifampicin resistance.

Future studies should perform comparisons of different tests, including Xpert Ultra, as this approach will reveal which tests (or strategies) yield superior diagnostic accuracy. For these studies, the preferred study design is one in which all participants receive all available diagnostic tests or are randomly assigned to receive one or another of the tests. Studies should include children and people living with HIV. Future research should acknowledge the concern associated with culture as a reference standard in paucibacillary specimens, and should consider ways to address this limitation.

Rapid point-of-care diagnostic tests for extrapulmonary tuberculosis are critically needed. Research groups should focus on developing diagnostic tests and strategies that use readily-available clinical specimens, such as urine, rather than specimens that require invasive procedures for collection.

For moderate complexity assays, all the tests had accuracy estimates similar to the WHO recommended molecular tests like Xpert Ultra and Xpert MTB/RIF for pulmonary TB. These tests can be very helpful when placed in high TB burden and in settings where high throughput and multi-disease nature of these platforms could be used to their fullest. This review was also used by the WHO to support policy recommendation on these platforms for diagnosing pulmonary TB.

The strengths of these reviews are the completeness of the evidence. All of these reviews included all recent publications in order to get a comprehensive picture of the research question. Additionally, the use of multiple reference standards in these reviews helps in addressing the imperfect nature of each reference standard. We also performed latent class meta-analyses which is a new statistical method being used for meta-analyses to account for imperfect nature of culture as the reference standard for extrapulmonary TB detection. Performing head-to-head comparisons wherever applicable (moderate complexity assays) helped in understanding if there is an issue with the accuracy of the index test or related to the design of the study and the target population included in the study. Such analyses help in understanding the accuracy estimates better and helpful in determining recommendations for future study designs evaluating such platforms.

The WHO recommends these molecular tests for diagnosis of TB and these reviews provided the evidence base for these recommendations. Molecular WHO-recommended rapid diagnostic tests (mWRDs) should be made available to all individuals with signs or symptoms of TB, to meet the End TB Strategy targets.[17] Additionally, all bacteriologically confirmed TB patients should receive DST for at least RIF (in 2018, only about 61% of such patients were tested for RIF resistance), and all patients with rifampicin resistant-TB should receive DST for at least fluoroquinolones.

These reviews provide how accurate these tests are, however, the decision on where to place a specific test is an important one that can lead to either success or failure in achieving the desired outcome of diagnosing a TB patient correctly. A test alone is not a silver bullet and it needs to be viewed in a broader ecosystem of tests, network of labs, and timely reporting of the results for further clinical management. To attain public health impact and have sustainability, national TB programmes need to assess where to place these tests, cost implications of these tests on the health budgets, appropriate resources, human resources for implementation, local availability of technical support for running these tests etc.

The WHO operational handbook on tuberculosis for diagnostic tests[204] provides guidance to countries on where these mWRDs should be placed based on the tiered healthcare system and infrastructure of the facilities. Figure 5.1 shows the organization of a TB diagnostic network suggesting different mWRDs placements across health system tiers. Diagnostic network optimization (DNO) is a geospatial network analytics approach to plan diagnostic networks

and ensure greatest access to and coverage of services, while maximizing the overall efficiency of the system, and there is now evidence to support the use of DNO to improve access and efficiency.

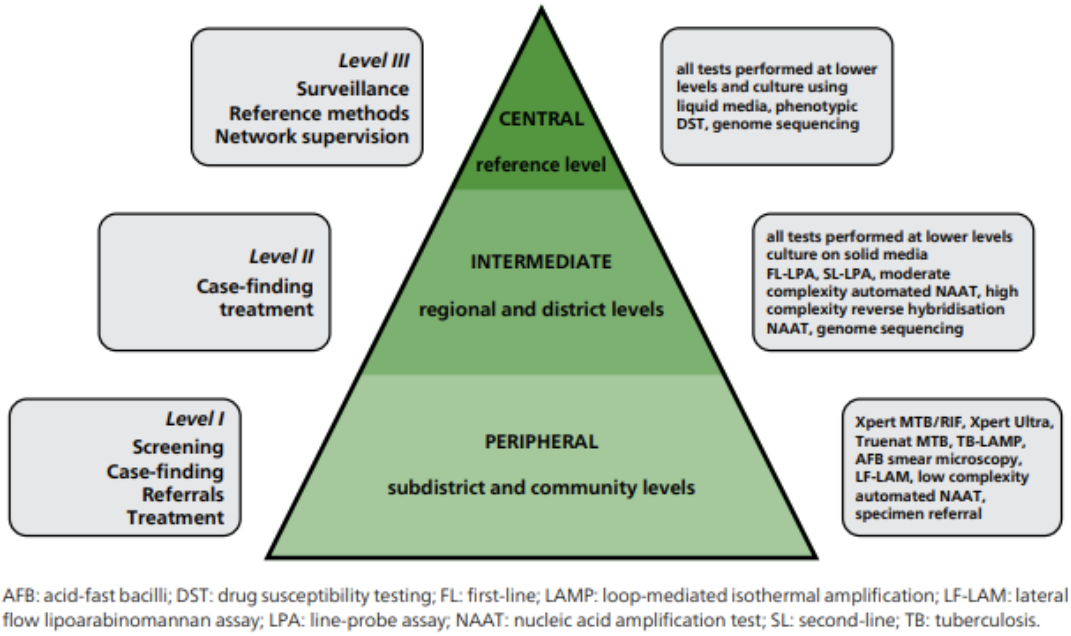


Figure 5.2 TB diagnostic network

Source: TB operational handbook on tuberculosis: Module 3: diagnosis

After placing the diagnostic test at an appropriate level in the healthcare system, it is imperative to place them correctly in the diagnostic cascade. Effective and efficient TB diagnostic algorithms are key components of a diagnostic cascade designed to ensure that patients with TB are diagnosed accurately and rapidly and are promptly placed on appropriate therapy. The WHO operational guide suggests four model algorithms. However, these are not prescriptive, and it is important that each country adapts them based on the country context and need.

When selecting a diagnostic algorithm to implement, it is important to consider the characteristics of the population being served. Thus, the four model algorithms are as follows:

• **Algorithm 1** relies on *mWRDs as the initial diagnostic tests* and is appropriate for all settings, although the choice of which mWRD to use may differ in a setting with high MDR/RR-TB prevalence

• **Algorithm 2** incorporates the *use of the LF-LAM as an aid in the diagnosis of TB in PLHIV* and is most relevant to settings with a high HIV prevalence.

• **Algorithm 3 and Algorithm 4** *are for follow-up testing*, after TB is diagnosed, to detect additional drug resistance. These algorithms are appropriate for all settings; however, the resource requirements for follow-up testing may differ strongly between settings with a high burden of DR-TB and settings with a low burden of DR-TB.

Conclusion:

The results of these studies provided evidence base for WHO policy recommendations on low and moderate complexity TB molecular tests. Our reviews also emphasize the critical role of systematic reviews and meta-analyses in generating quality evidence to make informed decisions about the newer diagnostic tests and their usability in countries. This is the first step for recommending the mWRDs and consequently every country based on their needs, disease prevalence and healthcare infrastructure should deploy these tests to achieve favourable outcomes of diagnosing and treating TB patients without losing them through the care cascade.

CHAPTER 6: REFERENCES

1. World Health Organization. *Global TB Report 2022*. (2022).
2. Boyd, R., Ford, N., Padgen, P. & Cox, H. Time to treatment for rifampicin-resistant tuberculosis: systematic review and meta-analysis. *The International Journal of Tuberculosis and Lung Disease* **21**, 1173–1180 (2017).
3. Theron, G. *et al.* Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *The Lancet* **383**, 424–435 (2014).
4. Sharma, S. K. & Mohan, A. Extrapulmonary tuberculosis. *Indian Journal of Medical Research* vol. 120 Preprint at https://doi.org/10.56834/26631504_2022_1_32 (2004).
5. World Health Organization. Automated Real-Time Nucleic Acid Amplification Technology for Rapid and Simultaneous Detection of Tuberculosis and Rifampicin Resistance: Xpert MTB/RIF Assay for the Diagnosis of Pulmonary and Extrapulmonary TB in Adults and Children. (2013).
6. United States Food, D. Administration. Decision summary K131706.
7. Migliori, G. B. *et al.* The definition of tuberculosis infection based on the spectrum of tuberculosis disease. *Breathe* **17**, 210079 (2021).
8. Lönnroth, K. *et al.* Towards tuberculosis elimination: an action framework for low-incidence countries. *European Respiratory Journal* **45**, 928–952 (2015).
9. European Centre for Disease Prevention and Control. Framework Action Plan to Fight Tuberculosis in the European Union.Stockholm, ECDC. (2008).
10. Pai, M. *et al.* Primer on Tuberculosis. *Nat Rev Dis Primers* **2**, (2016).
11. Barry, C. E. *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* **7**, 845–855 (2009).
12. Sharma, S. K. *et al.* Index-TB guidelines: Guidelines on extrapulmonary tuberculosis for India. *Indian Journal of Medical Research* **145**, (2017).
13. REUTER, H. Overview of extrapulmonary tuberculosis in adults and children. in *Tuberculosis* 377–390 (Elsevier, 2009). doi:10.1016/B978-1-4160-3988-4.00034-2.
14. ISEMAN, M. Extrapulmonary tuberculosis. *Clinician's Guide to Tuberculosis* (2000).
15. Harries, A. D. & Kumar, A. M. v. Challenges and Progress with Diagnosing Pulmonary Tuberculosis in Low- and Middle-Income Countries. *Diagnostics (Basel)* **8**, (2018).
16. Denkinger, C. M. *et al.* Defining the Needs for Next Generation Assays for Tuberculosis. *J Infect Dis* **211**, S29–S38 (2015).
17. World Health Organization. WHO | WHO End TB Strategy. *World Health Organization* (2015).
18. Desikan, P. Sputum smear microscopy in tuberculosis: is it still relevant? *Indian J Med Res* **137**, 442–4 (2013).
19. Lawn, S. D. Advances in Diagnostic Assays for Tuberculosis. *Cold Spring Harb Perspect Med* **5**, (2015).
20. Elliott, A. M. *et al.* Negative sputum smear results in HIV-positive patients with pulmonary tuberculosis in Lusaka, Zambia. *Tubercle and Lung Disease* **74**, 191–194 (1993).
21. Cattamanchi, A. *et al.* Sensitivity of direct versus concentrated sputum smear microscopy in HIV-infected patients suspected of having pulmonary tuberculosis. *BMC Infect Dis* **9**, 53 (2009).

22. World Health Orgnaization. *WHO consolidated guidelines on tuberculosis Rapid diagnostics for tuberculosis detection, 2021 update*. *World Health Organization* vol. Module 3 (2021).
23. WHO. Molecular line probe assays for rapid screening of patients at risk of multidrug-resistant tuberculosis (MDR-TB). *Policy statement* (2008).
24. World Health Organization (WHO). The use of molecular line probe assays for the detection of resistance to second-line anti-tuberculosis drugs THE STRATEGY END TB. *World Health Organization* (2016).
25. World Health Organization (WHO). *The use of loop-mediated isothermal amplification (TB-LAMP) for the diagnosis of pulmonary tuberculosis*. *WHO Library Cataloguing-in-Publication Data* (2016).
26. Shete, P. B., Farr, K., Strnad, L., Gray, C. M. & Cattamanchi, A. Diagnostic accuracy of TB-LAMP for pulmonary tuberculosis: a systematic review and meta-analysis. *BMC Infect Dis* **19**, 268 (2019).
27. World Health Organization. WHO endorses new rapid tuberculosis test. *WHO Media Centre* (2010).
28. World Health Organization. *Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB*. (2013).
29. WHO. WHO | Next-generation Xpert® MTB/RIF Ultra assay recommended by WHO. *WHO* Preprint at (2017).
30. Horne, D. J. *et al.* Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* (2019) doi:10.1002/14651858.CD009593.pub4.
31. 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).
32. Pai, M. Global health technologies: time to re-think the 'trickle down'model. *Forbes.[Google Scholar]* (2020).
33. Indian Council of Medical Research. India moving forward in research towards new tools for tuberculosis. 1–2 (2017).
34. Nathavitharana, R. R. *et al.* Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal* **49**, 1601075 (2017).
35. Hofmann-Thiel, S. & Hoffmann, H. Evaluation of Fluorotype MTB for detection of Mycobacterium tuberculosis complex DNA in clinical specimens from a low-incidence country. *BMC Infect Dis* **14**, 59 (2014).
36. Rocchetti, T. T., Silbert, S., Gostnell, A., Kubasek, C. & Widen, R. Validation of a Multiplex Real-Time PCR Assay for Detection of Mycobacterium spp., Mycobacterium tuberculosis Complex, and Mycobacterium avium Complex Directly from Clinical Samples by Use of the BD Max Open System. *J Clin Microbiol* **54**, 1644–1647 (2016).
37. Shah, M. *et al.* Multicenter Study of the Accuracy of the BD MAX Multidrug-resistant Tuberculosis Assay for Detection of Mycobacterium tuberculosis Complex and Mutations Associated With Resistance to Rifampin and Isoniazid. *Clinical Infectious Diseases* **71**, 1161–1167 (2020).
38. FIND. Diagnostics pipeline tracker: tuberculosis. (2020).
39. Wang, G. *et al.* Accuracy of Xpert MTB/RIF Ultra for the Diagnosis of Pleural TB in a Multicenter Cohort Study. *Chest* **157**, 268–275 (2020).

40. Deng, S. *et al.* Accuracy of Commercial Molecular Diagnostics for the Detection of Pulmonary Tuberculosis in China: A Systematic Review. *Sci Rep* **9**, 4553 (2019).
41. Sun, Y. *et al.* Accuracy of molecular diagnostic tests for drug-resistant tuberculosis detection in China: a systematic review. *The International Journal of Tuberculosis and Lung Disease* **23**, 931–942 (2019).
42. Yan, L. *et al.* A Large Cohort Study on the Clinical Value of Simultaneous Amplification and Testing for the Diagnosis of Pulmonary Tuberculosis. *Medicine* **95**, e2597 (2016).
43. Pang, Y. *et al.* Rapid diagnosis of MDR and XDR tuberculosis with the MeltPro TB assay in China. *Sci Rep* **6**, 25330 (2016).
44. Guo, Y. *et al.* Rapid, accurate determination of multidrug resistance in M. tuberculosis isolates and sputum using a biochip system. *Int J Tuberc Lung Dis* **13**, 914–20 (2009).
45. Lee, R. S. & Pai, M. Real-Time Sequencing of Mycobacterium tuberculosis: Are We There Yet? *J Clin Microbiol* **55**, 1249–1254 (2017).
46. Georghiou, S. *et al. The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide*. *WHO Global TB Programme* vol. 54 (2018).
47. World Health Organization. Relational sequencing TB data platform. (2019).
48. Myc-TB, G. D. Deeplex Myc-TB: From clinical samples to drug resistance profiles Lille, France: Genoscreen;[cited 2020 14/08]. Preprint at (2020).
49. Zignol, M. *et al.* Genetic sequencing for surveillance of drug resistance in tuberculosis in highly endemic countries: a multi-country population-based surveillance study. *Lancet Infect Dis* **18**, 675–683 (2018).
50. Cabibbe, A. M. *et al.* Application of Targeted Next-Generation Sequencing Assay on a Portable Sequencing Platform for Culture-Free Detection of Drug-Resistant Tuberculosis from Clinical Samples. *J Clin Microbiol* **58**, (2020).
51. Colman, R. E. *et al.* Whole-genome and targeted sequencing of drug-resistant Mycobacterium tuberculosis on the iSeq100 and MiSeq: A performance, ease-of-use, and cost evaluation. *PLoS Med* **16**, e1002794 (2019).
52. Kohli, M. *et al.* Xpert MTB/RIF Ultra and Xpert MTB/RIF assays for extrapulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* vol. 2021 Preprint at https://doi.org/10.1002/14651858.CD012768.pub3 (2021).
53. 2020, W. Global TB Report 2020. *World Health Organization* **1**, (2020).
54. Sandgren, A., Hollo, V. & van der Werf, M. J. Extrapulmonary tuberculosis in the European Union and European Economic Area, 2002 to 2011. *Eurosurveillance* **18**, (2013).
55. Pang, Y. *et al.* Epidemiology of Extrapulmonary Tuberculosis among Inpatients, China, 2008–2017. *Emerg Infect Dis* **25**, 457–464 (2019).
56. Golden, M. P. & Vikram, H. R. Extrapulmonary tuberculosis: an overview. *Am Fam Physician* **72**, 1761–8 (2005).
57. Perkins, M. D. & Cunningham, J. Facing the Crisis: Improving the Diagnosis of Tuberculosis in the HIV Era. *J Infect Dis* **196**, S15–S27 (2007).
58. Webster, A. & Shandera, W. The extrapulmonary dissemination of tuberculosis: A meta-analysis. *Int J Mycobacteriol* **3**, 9 (2014).
59. Nelson, L. J. & Wells, C. D. Global epidemiology of childhood tuberculosis [Childhood TB]. *The International journal of Tuberculosis and lung Disease* **8**, 636–647 (2004).

60. Kohli, M. *et al.* Xpert ® MTB/RIF assay for extrapulmonary tuberculosis and rifampicin resistance. *Cochrane Database of Systematic Reviews* (2018) doi:10.1002/14651858.CD012768.pub2.
61. 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* (2019) doi:10.1002/14651858.CD013359.
62. Chakraborty, A., Ramaswamy, S., Shivananjiah, A. J., Puttaswamy, R. B. & Chikkavenkatappa, N. The Role of Genexpert in the Diagnosis of Tubercular Pleural Effusion in India. *Adv Respir Med* **87**, 276–280 (2019).
63. Cepheid. Xpert® MTB/RIF Ultra. (2018).
64. Centers for Disease Control and Prevention. Reported tuberculosis in the United States. (2018).
65. Lewinsohn, D. M. *et al.* Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. *Clinical Infectious Diseases* vol. 64 Preprint at https://doi.org/10.1093/cid/ciw694 (2017).
66. Naaktgeboren, C. A. *et al.* Value of composite reference standards in diagnostic research. *BMJ* **347**, (2013).
67. Veritas Health Innovation. Covidence systematic review software. . (2017).
68. Moher, D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* vol. 6 Preprint at https://doi.org/10.1371/journal.pmed.1000097 (2009).
69. The World Bank. *World Bank list of economies (June 2020). World Bank list of economies* (2020).
70. World Health Organization. Xpert MTB/RIF implementation manual Technical and operational 'how-to': practical considerations. *Geneva, Switzerland* **ISBN: 978**, (2014).
71. Begg, C. B. & Greenes, R. A. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* **39**, 207–15 (1983).
72. Whiting, P. F. *et al.* Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* vol. 155 Preprint at https://doi.org/10.7326/0003-4819-155-8-201110180-00009 (2011).
73. Stata. Statistical Software: Release 15. StataCorp. College Station, TX: StataCorp LP. (2017).
74. Collaboration, T. C. Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014. *Cochrane Community* Preprint at (2014).
75. Chu, H. & Cole, S. R. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* **59**, 1331–1332 (2006).
76. Takwoingi, Y., Leeflang, M. M. G. & Deeks, J. J. Empirical Evidence of the Importance of Comparative Studies of Diagnostic Test Accuracy. *Ann Intern Med* **158**, 544 (2013).
77. Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. The BUGS project: Evolution, critique and future directions. *Stat Med* **28**, 3049–3067 (2009).
78. van Smeden, M., Naaktgeboren, C. A., Reitsma, J. B., Moons, K. G. M. & de Groot, J. A. H. Latent Class Models in Diagnostic Studies When There is No Reference Standard--A Systematic Review. *Am J Epidemiol* **179**, 423–431 (2014).
79. Gopinath, K. & Singh, S. Non-Tuberculous Mycobacteria in TB-Endemic Countries: Are We Neglecting the Danger? *PLoS Negl Trop Dis* **4**, e615 (2010).

80. Blakemore, R. *et al.* Evaluation of the Analytical Performance of the Xpert MTB/RIF Assay. *J Clin Microbiol* **48**, 2495–2501 (2010).
81. Helb, D. *et al.* Rapid Detection of *Mycobacterium tuberculosis* and Rifampin Resistance by Use of On-Demand, Near-Patient Technology. *J Clin Microbiol* **48**, 229–237 (2010).
82. Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R. & Takwoingi, Y. Analysing and Presenting Results. in *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (2018).
83. Balshem, H. *et al.* GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* **64**, 401–406 (2011).
84. McMaster University. GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. *Evidence Prime, Inc.* **Available**, (2015).
85. Schünemann, H. J. *et al.* Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336**, 1106–1110 (2008).
86. Schünemann, H. J. *et al.* GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol* **76**, 89–98 (2016).
87. Schünemann, H. J. *et al.* GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *J Clin Epidemiol* **122**, 129–141 (2020).
88. Schünemann, H. J. *et al.* GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J Clin Epidemiol* **122**, 142–152 (2020).
89. Zmak, L., Jankovic, M. & Jankovic, V. K. Evaluation of Xpert MTB/RIF assay for rapid molecular diagnosis of tuberculosis in a two-year period in Croatia. *Int J Mycobacteriol* **2**, (2013).
90. Zeka, A. N., Tasbakan, S. & Cavusoglu, C. Evaluation of the GeneXpert MTB/RIF assay for rapid diagnosis of tuberculosis and detection of rifampin resistance in pulmonary and extrapulmonary specimens. *J Clin Microbiol* **49**, (2011).
91. Wang, G. *et al.* Xpert MTB/RIF Ultra improved the diagnosis of paucibacillary tuberculosis: A prospective cohort study. *Journal of Infection* **78**, (2019).
92. Wu, X. *et al.* Assessment of the Xpert MTB/RIF Ultra assay on rapid diagnosis of extrapulmonary tuberculosis. *International Journal of Infectious Diseases* **81**, (2019).
93. van Rie, A. *et al.* Diagnostic accuracy and effectiveness of the Xpert MTB/RIF assay for the diagnosis of HIV-associated lymph node tuberculosis. *European Journal of Clinical Microbiology and Infectious Diseases* **32**, (2013).
94. Vadwai, V. *et al.* Xpert MTB/RIF: A new pillar in diagnosis of extrapulmonary tuberculosis? *J Clin Microbiol* **49**, (2011).
95. Ullah, I. *et al.* Rapid detection of mycobacterium tuberculosis and rifampicin resistance in extrapulmonary tuberculosis and sputum smearnegative pulmonary suspects using Xpert MTB/RIF. *J Med Microbiol* **66**, (2017).
96. Trajman, A. *et al.* Accuracy of polimerase chain reaction for the diagnosis of pleural tuberculosis. *Respir Med* **108**, (2014).
97. Tadesse, M. *et al.* GeneXpert MTB/RIF assay for the diagnosis of tuberculous lymphadenitis on concentrated fine needle aspirates in high tuberculosis burden settings. *PLoS One* **10**, (2015).
98. Suzana, S. *et al.* Xpert MTB/Rif for the diagnosis of extrapulmonary tuberculosis - an experience from a tertiary care centre in South India. *Tropical Medicine and International Health* **21**, (2016).

99. Sun, Q. *et al.* Diagnostic value of Xpert MTB/RIF Ultra for osteoarticular tuberculosis. *Journal of Infection* **79**, (2019).
100. Siddiqi, O. K. *et al.* Prospective cohort study on performance of cerebrospinal fluid (CSF) Xpert MTB/RIF, CSF lipoarabinomannan (LAM) lateral flow assay (LFA), and urine LAM LFA for diagnosis of tuberculous meningitis in Zambia. *J Clin Microbiol* **57**, (2019).
101. Sharma, K. *et al.* Comparative evaluation of Xpert MTB/RIF assay with multiplex polymerase chain reaction for the diagnosis of tuberculous meningitis. *Tuberculosis* **113**, (2018).
102. Sharma, J. B. *et al.* Role of Gene Xpert in diagnosis of female genital tuberculosis: a preliminary report. *European Journal of Obstetrics and Gynecology and Reproductive Biology* vol. 207 Preprint at https://doi.org/10.1016/j.ejogrb.2016.10.045 (2016).
103. Sharma, S. K. *et al.* Evaluation of Xpert MTB/RIF assay performance in diagnosing extrapulmonary tuberculosis among adults in a tertiary care centre in India. *European Respiratory Journal* vol. 44 Preprint at https://doi.org/10.1183/09031936.00059014 (2014).
104. Scott, L. E. *et al.* Diagnostic accuracy of xpert MTB/RIF for extrapulmonary tuberculosis specimens: Establishing a laboratory testing algorithm for South Africa. *J Clin Microbiol* **52**, (2014).
105. Sarfaraz, S. *et al.* Histopathological and microbiological findings and diagnostic performance of GeneXpert in clinically suspected tuberculous lymphadenitis. *International Journal of Infectious Diseases* **76**, (2018).
106. Safianowska, A., Walkiewicz, R., Nejman-Gryz, P. & Grubek-Jaworska, H. [Two selected commercially based nucleic acid amplification tests for the diagnosis of tuberculosis]. *Pneumonol Alergol Pol* **80**, (2012).
107. Rufai, S. B. *et al.* Diagnostic usefulness of Xpert MTB/RIF assay for detection of tuberculous meningitis using cerebrospinal fluid. *Journal of Infection* **75**, (2017).
108. Rufai, S. B. *et al.* Performance of Xpert MTB/RIF on Ascitic Fluid Samples for Detection of Abdominal Tuberculosis. *J Lab Physicians* **9**, (2017).
109. Rufai, S. B., Singh, A., Kumar, P., Singh, J. & Singh, S. Performance of xpert MTB/RIF assay in diagnosis of pleural tuberculosis by use of pleural fluid samples. *J Clin Microbiol* **53**, (2015).
110. Rakotoarivelo, R. *et al.* Evaluation of the Xpert MTB/RIF assay for the diagnosis of smear-negative pulmonary and extrapulmonary tuberculosis in Madagascar. *International Journal of Infectious Diseases* **69**, (2018).
111. Perez-Risco, D., Rodriguez-Temporal, D., Valledor-Sanchez, I. & Alcaidea, F. Evaluation of the Xpert MTB/RIF Ultra Assay for Direct Detection of Mycobacterium tuberculosis Complex in Smear-Negative Extrapulmonary Samples. *J Clin Microbiol* **56**, (2018).
112. Peñata, A., Salazar, R., Castaño, T., Bustamante, J. & Ospina, S. Diagnóstico molecular de tuberculosis extrapulmonar y sensibilidad a rifampicina con un método automatizado en tiempo real. *Biomédica* **36**, (2016).
113. Patel, V. B. *et al.* Diagnostic Accuracy of Quantitative PCR (Xpert MTB/RIF) for Tuberculous Meningitis in a High Burden Setting: A Prospective Study. *PLoS Med* **10**, e1001536 (2013).
114. Pandie, S. *et al.* Diagnostic accuracy of quantitative PCR (Xpert MTB/RIF) for tuberculous pericarditis compared to adenosine deaminase and unstimulated interferon-γ in a high burden setting: a prospective study. *BMC Med* **12**, 101 (2014).

115. ÖZKÜTÜK, N. & SÜRÜCÜOĞLU, S. Evaluation of the Xpert MTB/RIF Assay for the Diagnosis of Pulmonary and Extrapulmonary Tuberculosis in an Intermediate-Prevalence Setting. *Mikrobiyol Bul* **48**, 223–232 (2014).
116. Nhu, N. T. Q. *et al.* Evaluation of GeneXpert MTB/RIF for Diagnosis of Tuberculous Meningitis. *J Clin Microbiol* **52**, 226–233 (2014).
117. Nataraj, G., Kanade, S. & Mehta, P. Xpert® MTB/RIF for improved case detection of extra-pulmonary TB in a tertiary care setting in urban India. *The International Journal of Tuberculosis and Lung Disease* **20**, 890–894 (2016).
118. Meldau, R. *et al.* Same-Day Tools, Including Xpert Ultra and IRISA-TB, for Rapid Diagnosis of Pleural Tuberculosis: a Prospective Observational Study. *J Clin Microbiol* **57**, (2019).
119. Meldau, R. *et al.* Comparison of same day diagnostic tools including Gene Xpert and unstimulated IFN-γ for the evaluation of pleural tuberculosis: A prospective cohort study. *BMC Pulm Med* **14**, (2014).
120. Malbruny, B., le Marrec, G., Courageux, K., Leclercq, R. & Cattoir, V. Rapid and efficient detection of Mycobacterium tuberculosis in respiratory and non-respiratory samples. *International Journal of Tuberculosis and Lung Disease* **15**, (2011).
121. Lusiba, J. K. *et al.* Evaluation of Cepheid's Xpert MTB/RIF test on pleural fluid in the diagnosis of pleural tuberculosis in a high prevalence HIV/TB setting. *PLoS One* **9**, (2014).
122. Ligthelm, L. J. *et al.* Xpert MTB/RIF for rapid diagnosis of tuberculous lymphadenitis from fine-needle-aspiration biopsy specimens. *J Clin Microbiol* **49**, (2011).
123. Liang, Q. *et al.* An improved algorithm for rapid diagnosis of pleural tuberculosis from pleural effusion by combined testing with GeneXpert MTB/RIF and an anti-LAM antibody-based assay. *BMC Infect Dis* **19**, (2019).
124. Li, Y. *et al.* Rapid diagnosis of extrapulmonary tuberculosis with Xpert Mycobacterium tuberculosis/rifampicin assay. *J Med Microbiol* **66**, (2017).
125. Kim, Y. W. *et al.* Accuracy of the Xpert® MTB/RIF assay for the diagnosis of extra-pulmonary tuberculosis in South Korea. *The International Journal of Tuberculosis and Lung Disease* **19**, 81–86 (2015).
126. Iram, S., Zeenat, A., Hussain, S., Yusuf, N. W. & Aslam, M. Rapid diagnosis of tuberculosis using Xpert MTB/RIF assay - report from a developing country. *Pak J Med Sci* **31**, (2015).
127. Hillemann, D., Rüsch-Gerdes, S., Boehme, C. & Richter, E. Rapid molecular detection of extrapulmonary tuberculosis by the automated genexpert MTB/RIF system. *J Clin Microbiol* **49**, (2011).
128. Heemskerk, A. D. *et al.* Improving the microbiological diagnosis of tuberculous meningitis: A prospective, international, multicentre comparison of conventional and modified Ziehl–Neelsen stain, GeneXpert, and culture of cerebrospinal fluid. *Journal of Infection* **77**, (2018).
129. Hanif, S. N. M., Eldeen, H. S., Ahmad, S. & Mokaddas, E. GeneXpert® MTB/RIF for rapid detection of *Mycobacterium tuberculosis* in pulmonary and extra-pulmonary samples [Correspondence]. *The International Journal of Tuberculosis and Lung Disease* **15**, 1274–1275 (2011).
130. Gu, Y. *et al.* Xpert MTB/RIF and GenoType MTBDRplus assays for the rapid diagnosis of bone and joint tuberculosis. *International Journal of Infectious Diseases* **36**, (2015).
131. Ghariani, A. *et al.* Diagnosis of lymph node tuberculosis using the GeneXpert MTB/RIF in Tunisia. *Int J Mycobacteriol* **4**, (2015).

132. Friedrich, S. O., von Groote-Bidlingmaier, F. & Diacon, A. H. Xpert MTB/RIF assay for diagnosis of pleural tuberculosis. *J Clin Microbiol* **49**, (2011).
133. Feasey, N. A. *et al.* Evaluation of Xpert MTB/RIF for detection of tuberculosis from blood samples of HIV-infected adults confirms Mycobacterium tuberculosis bacteremia as an indicator of poor prognosis. *J Clin Microbiol* **51**, (2013).
134. Galal El-Din, M., Sobh, E., Adawy, Z. & Farghaly, N. Diagnostic utility of gene X-pert in the diagnosis of tuberculous pleural effusion. *Infectious Diseases* vol. 51 Preprint at https://doi.org/10.1080/23744235.2018.1532105 (2019).
135. Du, J. *et al.* Rapid diagnosis of pleural tuberculosis by Xpert MTB/RIF assay using pleural biopsy and pleural fluid specimens. *Journal of Research in Medical Sciences* **20**, (2015).
136. Donovan, J. *et al.* Xpert MTB/RIF Ultra versus Xpert MTB/RIF for the diagnosis of tuberculous meningitis: a prospective, randomised, diagnostic accuracy study. *Lancet Infect Dis* **20**, (2020).
137. Dhooria, S. *et al.* Role of Xpert MTB/RIF in differentiating tuberculosis from sarcoidosis in patients with mediastinal lymphadenopathy undergoing EBUS-TBNA: A study of 147 patients. *Sarcoidosis Vasculitis and Diffuse Lung Diseases* **33**, (2016).
138. Dhasmana, D. J. *et al.* Performance of Xpert MTB/RIF in the diagnosis of tuberculous mediastinal lymphadenopathy by endobronchial ultrasound. *Ann Am Thorac Soc* **11**, (2014).
139. Cresswell, F. v. *et al.* Xpert MTB/RIF Ultra for the diagnosis of HIV-associated tuberculous meningitis: a prospective validation study. *Lancet Infect Dis* **20**, (2020).
140. Cresswell, F. v. *et al.* Can improved diagnostics reduce mortality from Tuberculous meningitis? Findings from a 6.5-year cohort in Uganda. *Wellcome Open Res* **3**, (2018).
141. Christopher, D. J. *et al.* Performance of Xpert MTB/RIF on pleural tissue for the diagnosis of pleural tuberculosis. *European Respiratory Journal* vol. 42 Preprint at https://doi.org/10.1183/09031936.00103213 (2013).
142. Chin, J. H. *et al.* Xpert MTB/RIF Ultra for Detection of Mycobacterium tuberculosis in Cerebrospinal Fluid. *Journal of Clinical Microbiology* vol. 57 Preprint at https://doi.org/10.1128/JCM.00249-19 (2019).
143. Chen, C. *et al.* Community-based active case finding for tuberculosis in rural western China: A cross-sectional study. *International Journal of Tuberculosis and Lung Disease* **21**, (2017).
144. Che, N., Yang, X., Liu, Z., Li, K. & Chen, X. Rapid detection of cell-free Mycobacterium tuberculosis DNA in tuberculous pleural effusion. *J Clin Microbiol* **55**, (2017).
145. Causse, M., Ruiz, P., Gutiérrez-Aroca, J. B. & Casal, M. Comparison of two molecular methods for rapid diagnosis of extrapulmonary tuberculosis. *J Clin Microbiol* **49**, (2011).
146. Blaich, A. & Frei, R. Performance of the Xpert MTB/RIF assay on nonrespiratory specimens and accuracy of this assay for detection of rifampin resistance in a low-prevalence setting. *Journal of Clinical Microbiology* vol. 52 Preprint at https://doi.org/10.1128/JCM.02591-13 (2014).
147. Biadglegne, F., Mulu, A., Rodloff, A. C. & Sack, U. Diagnostic performance of the Xpert MTB/RIF assay for tuberculous lymphadenitis on fine needle aspirates from Ethiopia. *Tuberculosis* **94**, (2014).
148. Bera, C. *et al.* Tissue Xpert^{TM} MTB/Rif assay is of limited use in diagnosing peritoneal tuberculosis in patients with exudative ascites. *Indian Journal of Gastroenterology* **34**, 395–398 (2015).

149. Bahr, N. C. *et al.* Diagnostic accuracy of Xpert MTB/RIF Ultra for tuberculous meningitis in HIV-infected adults: a prospective cohort study. *Lancet Infect Dis* **18**, 68–75 (2018).
150. Bahr, N. C. *et al.* Improved diagnostic sensitivity for tuberculous meningitis with Xpert® MTB/RIF of centrifuged CSF. *International Journal of Tuberculosis and Lung Disease* **19**, (2015).
151. Azevedo, R. G., Dinallo, F. S., de Laurentis, L. S., Boulware, D. R. & Vidal, J. E. Xpert MTB/RIF® assay for the diagnosis of HIV-related tuberculous meningitis in São Paulo, Brazil. *International Journal of Tuberculosis and Lung Disease* vol. 22 Preprint at https://doi.org/10.5588/ijtld.18.0191 (2018).
152. Antel, K. *et al.* Diagnostic accuracy of the Xpert MTB/Rif Ultra for tuberculosis adenitis. *BMC Infect Dis* **20**, (2020).
153. Al-Ateah, S. M., Al-Dowaidi, M. M. & El-Khizzi, N. A. Evaluation of direct detection of Mycobacterium tuberculosis complex in respiratory and non-respiratory clinical specimens using the Cepheid Gene Xpert® system. *Saudi Med J* **33**, (2012).
154. Ajbani, K. *et al.* Utility of pyrosequencing for rapid detection of tubercular meningitis (TBM) and associated susceptibility directly from CSF specimens. *Tuberculosis* **111**, (2018).
155. Ablanedo-Terrazas, Y., Alvarado-De La Barrera, C., Hernández-Juan, R., Ruiz-Cruz, M. & Reyes-Terán, G. Xpert MTB/RIF for diagnosis of tuberculous cervical lymphadenitis in HIV-infected patients. *Laryngoscope* **124**, (2014).
156. Thwaites, G. E., van Toorn, R. & Schoeman, J. Tuberculous meningitis: More questions, still too few answers. *The Lancet Neurology* vol. 12 Preprint at https://doi.org/10.1016/S1474-4422(13)70168-6 (2013).
157. 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**, (2017).
158. Woods, G. L. Molecular techniques in mycobacterial detection. *Arch Pathol Lab Med* **125**, (2001).
159. Pai, M., Flores, L. L., Hubbard, A., Riley, L. W. & Colford, J. M. Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: A systematic review and meta-analysis. *BMC Infectious Diseases* vol. 4 Preprint at https://doi.org/10.1186/1471-2334-4-6 (2004).
160. Theron, G. *et al.* Determinants of PCR performance (Xpert MTB/RIF), including bacterial load and inhibition, for TB diagnosis using specimens from different body compartments. *Sci Rep* **4**, (2014).
161. Wright, C. A. *et al.* Fine-needle aspiration biopsy: a first-line diagnostic procedure in paediatric tuberculosis suspects with peripheral lymphadenopathy? *The International journal of tuberculosis and lung disease* **13**, 1373–1379 (2009).
162. Denkinger, C. M. *et al.* Xpert MTB/RIF assay for the diagnosis of extrapulmonary tuberculosis: A systematic review and meta-analysis. *European Respiratory Journal* **44**, (2014).
163. Maynard-Smith, L., Larke, N., Peters, J. A. & Lawn, S. D. Diagnostic accuracy of the Xpert MTB/RIF assay for extrapulmonary and pulmonary tuberculosis when testing non-respiratory samples: A systematic review. *BMC Infect Dis* **14**, (2014).
164. Penz, E. *et al.* Diagnostic accuracy of the Xpert® MTB/RIF assay for extrapulmonary tuberculosis: A meta-analysis. *International Journal of Tuberculosis and Lung Disease* **19**, (2015).
165. Schiller, I. *et al.* Bias due to composite reference standards in diagnostic accuracy studies. *Stat Med* **35**, (2016).

166. Fontanilla, J. M., Barnes, A. & von Reyn, C. F. Current diagnosis and management of peripheral tuberculous Lymphadenitis. *Clinical Infectious Diseases* vol. 53 Preprint at https://doi.org/10.1093/cid/cir454 (2011).
167. Leeflang, M. M. G., Rutjes, A. W. S., Reitsma, J. B., Hooft, L. & Bossuyt, P. M. M. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ. Canadian Medical Association Journal* **185**, (2013).
168. World Health Organization, (WHO). Technical report on critical concentrations for TB drug susceptibility testing of medicines used in the treatment of drug-resistant TB. *Who* (2018).
169. Telenti, A. *et al.* Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. *The Lancet* **341**, (1993).
170. 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 Health* **7**, (2019).
171. Haraka, F. *et al.* Impact of diagnostic test Xpert MTB/RIF® on health outcomes for tuberculosis. *Cochrane Database of Systematic Reviews* **2018**, (2018).
172. Schumacher, S. G. & Denkinger, C. M. The impact of Xpert MTB/RIF—do we have a final answer? *The Lancet Global Health* vol. 7 Preprint at https://doi.org/10.1016/S2214-109X(18)30493-5 (2019).
173. McInnes, M. D. F. *et al.* Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies The PRISMA-DTA Statement. *JAMA - Journal of the American Medical Association* **319**, (2018).
174. Shah, M. *et al.* Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database of Systematic Reviews* vol. 2016 Preprint at https://doi.org/10.1002/14651858.CD011420.pub2 (2016).
175. Bjerrum, S. *et al.* Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV. *Cochrane Database of Systematic Reviews* vol. 2019 Preprint at https://doi.org/10.1002/14651858.CD011420.pub3 (2019).
176. Drain, P. K. *et al.* Guidance for Studies Evaluating the Accuracy of Biomarker-Based Nonsputum Tests to Diagnose Tuberculosis. *Journal of Infectious Diseases* **220**, (2019).
177. Bossuyt, P. M. *et al.* STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Clinical Chemistry* vol. 61 Preprint at https://doi.org/10.1373/clinchem.2015.246280 (2015).
178. Zhang, M., Xue, M. & He, J. qing. Diagnostic accuracy of the new Xpert MTB/RIF Ultra for tuberculosis disease: A preliminary systematic review and meta-analysis. *International Journal of Infectious Diseases* **90**, (2020).
179. UNITAID. Tuberculosis Diagnostics Technology Landscape. *WHO press* (2017).
180. Dicks, K. v. & Stout, J. E. Molecular diagnostics for mycobacterium tuberculosis infection. *Annu Rev Med* **70**, (2019).
181. Organization, W. H. *WHO treatment guidelines for isoniazid-resistant tuberculosis: supplement to the WHO treatment guidelines for drug-resistant tuberculosis*. (2018).
182. Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R. & Takwoingi, Y. Cochrane handbook for systematic reviews of diagnostic test accuracy. Preprint at (2010).
183. Miotto, P., Cabibbe, A. M., Borroni, E., Degano, M. & Cirilloa, D. M. Role of disputed mutations in the rpoB gene in interpretation of automated liquid MGIT culture results for rifampin susceptibility testing of mycobacterium tuberculosis. *J Clin Microbiol* **56**, (2018).
184. Schumacher, S. G. *et al.* Guidance for studies evaluating the accuracy of sputum-based tests to diagnose tuberculosis. *J Infect Dis* **220**, S99–S107 (2019).

185. Berhanu, R. H. *et al.* Performance of Xpert MTB/RIF, Xpert Ultra, and Abbott RealTi *m* e MTB for Diagnosis of Pulmonary Tuberculosis in a High-HIV-Burden Setting. *J Clin Microbiol* **56**, (2018).
186. Scott, L. *et al.* Performance of the abbott realtime MTB and MTB RIF/INH assays in a setting of high tuberculosis and HIV coinfection in South Africa. *J Clin Microbiol* **55**, (2017).
187. Wang, S. F. *et al.* The Abbott RealTime MTB assay and the Cepheid GeneXpert assay show comparable performance for the detection of Mycobacterium tuberculosis in sputum specimens. *International Journal of Infectious Diseases* **45**, (2016).
188. Hofmann-Thiel, S., Molodtsov, N., Antonenka, U. & Hoffmann, H. Evaluation of the Abbott RealTime MTB and RealTime MTB INH/RIF assays for direct detection of mycobacterium tuberculosis complex and resistance markers in respiratory and extrapulmonary specimens. *J Clin Microbiol* **54**, (2016).
189. Hofmann-Thiel, S. *et al.* Capacity of Abbott RealTime MTB RIF/INH to detect rifampicinand isoniazid-resistant tuberculosis. *International Journal of Tuberculosis and Lung Disease* **23**, (2019).
190. Kostera, J. *et al.* Analytical and clinical performance characteristics of the Abbott RealTime MTB RIF/INH Resistance, an assay for the detection of rifampicin and isoniazid resistant Mycobacterium tuberculosis in pulmonary specimens. *Tuberculosis* **101**, (2016).
191. Kostera, J., Leckie, G., Abravaya, K. & Wang, H. Performance of the Abbott RealTime MTB RIF/INH resistance assay when used to test Mycobacterium tuberculosis specimens from Bangladesh. *Infect Drug Resist* **11**, (2018).
192. Tam, K. K. G. *et al.* Direct detection of Mycobacterium tuberculosis and drug resistance in respiratory specimen using Abbott Realtime MTB detection and RIF/INH resistance assay. *Diagn Microbiol Infect Dis* **89**, (2017).
193. Eigner, U., Veldenzer, A. & Holfelder, M. Evaluation of the fluorotype MTB assay for the rapid and reliable detection of mycobacterium tuberculosis in respiratory tract specimens. *Clin Lab* **59**, (2013).
194. Erkose Genc, G. *et al.* Evaluation of FluoroType MTB for direct detection of Mycobacterium tuberculosis complex and GenoType MTBDRplus for determining rifampicin and isoniazid resistance. *Biotechnology and Biotechnological Equipment* **32**, (2018).
195. Hofmann-Thiel, S. & Hoffmann, H. Evaluation of Fluorotype MTB for detection of Mycobacterium tuberculosis complex DNA in clinical specimens from a low-incidence country. *BMC Infect Dis* **14**, 59–59 (2014).
196. Obasanya, J. *et al.* Fluorotype MTB system for the detection of pulmonary tuberculosis. *ERS Monograph* vol. 3 Preprint at https://doi.org/10.1183/23120541.00113-2016 (2017).
197. de Vos, M. *et al.* Diagnostic accuracy and utility of FluoroType MTBDR, a new molecular assay for multidrug-resistant tuberculosis. *J Clin Microbiol* **56**, (2018).
198. Haasis, C. *et al.* Validation of the FluoroType ® MTBDR assay using respiratory and lymph node samples. *Tuberculosis* **113**, (2018).
199. Hillemann, D., Haasis, C., Andres, S., Behn, T. & Kranzer, K. Validation of the FluoroType MTBDR assay for detection of rifampin and isoniazid resistance in mycobacterium tuberculosis complex isolates. *J Clin Microbiol* **56**, (2018).
200. Theron, G. *et al.* Xpert MTB/RIF Results in Patients With Previous Tuberculosis: Can We Distinguish True From False Positive Results? *Clinical Infectious Diseases* **62**, (2016).

201. Theron, G. *et al.* Xpert MTB/RIF Results in Patients With Previous Tuberculosis: Can We Distinguish True From False Positive Results? *Clinical Infectious Diseases* **62**, 995–1001 (2016).
202. Manson, A. L. *et al.* Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. *Nat Genet* **49**, 395–402 (2017).
203. Stop TB Partnership. UNHLM Global TB targets. (2018).
204. World Health Organization (WHO). *Operational handbook on tuberculosis Module 3 Diagnosis Updated 2021*. *Module 3: Diagnosis Rapid diagnotsics for tuberculosis diagnosis* (2021).