WHO consolidated guidelines on tuberculosis

Module 3: Diagnosis

World Health Organization

WHO consolidated guidelines on tuberculosis

Module 3: Diagnosis

World Health Organization

WHO consolidated guidelines on tuberculosis. Module 3: diagnosis

ISBN 978-92-4-010798-4 (electronic version)
ISBN 978-92-4-010799-1 (print version)

© World Health Organization 2025

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (http://www.wipo.int/amc/en/mediation/rules/).

Suggested citation. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Geneva: World Health Organization; 2025. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at https://iris.who.int/.

Sales, rights and licensing. To purchase WHO publications, see https://www.who.int/publications/book-orders. To submit requests for commercial use and queries on rights and licensing, see https://www.who.int/copyright.

Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

Technical editing by Cadman Editing and design by Inis Communication

Contents

Acknowledgements v

Abbreviations and acronyms vii

Definitions ix

Executive summary x

1. Introduction 1

1.1. Background 1

1.2. WHO TB diagnostic class determination and product prequalification 2

1.2.1 Pathway A 2

1.2.2 Pathway B 3

1.3. Testing classes and products 3

1.3.1 Initial tests for TB diagnosis with drug resistance detection 4

1.3.2 Initial tests for TB diagnosis without drug resistance detection 5

1.3.3 Follow-on tests for detection of TB drug resistance 5

1.3.4 Tests for TB infection 6

1.4. Scope of the document 7

1.5. Target audience 7

1.6. Scope of the document 7

1.7. Target audience 7

2. Recommendations for diagnosis of TB disease 9

2.1. Initial diagnostic tests for diagnosis of TB with drug-resistance detection 9

2.1.1 LC-aNAATs for detection of TB and resistance to rifampicin 9

2.1.2 Moderate complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid 24

2.2. Initial diagnostic tests for diagnosis of TB without drug-resistance detection 37

2.2.1 Low-Complexity manual NAATs for detection of TB 37

2.3. Concurrent use of initial diagnostic tests for diagnosis of TB in People living with HIV and children 48

2.3.1 Concurrent use of tests in people living with HIV 50
2.3.2 Concurrent use of tests in children without HIV or with unknown HIV status 57
2.3.3 Concurrent use of tests in children with HIV 66
2.4. Follow-on diagnostic tests for detection of additional drug-resistance after TB confirmation 71
2.4.1 Low complexity automated NAATs for detection of resistance to isoniazid and second-line anti-TB agents 71
2.4.2 First-line LPAs 83
2.4.3 Second-line LPAs 87
2.4.4 High complexity reverse hybridization-based NAATs for detection of pyrazinamide resistance 92
2.4.5 Targeted next-generation sequencing 100
2.5. References 119

3. Recommendations for diagnosis of TB infection 121

3.1. Mycobacterium tuberculosis antigen-based skin tests for the diagnosis of TB infection 121
3.2. TB skin tests and interferon gamma release assays for the diagnosis of TB infection 139
3.3. TB skin tests and interferon gamma release assays for the diagnosis of TB disease 142
3.4. References 147

Annex 1. Guideline development methods 149

Annex 2. Conflict of interest assessment for Guideline Development Group and External Review Group members 153

Annex 3. GDG members expertise, region, gender 163

Web Annexes 166

Web Annex A. Acknowledgements, List of studies included in systematic review, GRADE profiles, Evidence to decision tables
https://doi.org/10.2471/B09340

Web Annex B. Systematic reviews
https://doi.org/10.2471/B09341

Acknowledgements

The recommendations and remarks in this policy guideline on tuberculosis (TB) are the result of the collaborative effort of professionals from a range of specialties. The World Health Organization (WHO) is grateful for their time and support. There were separate Guideline Development Groups (GDGs) for each of the guidelines that have been included in these consolidated guidelines. The acknowledgements provided immediately below are specific to WHO guidelines that are new in this edition. Acknowledgements for prior guidelines are summarized in Web Annex A.

The production and writing of this document – WHO consolidated guidelines on tuberculosis. Module 3: Diagnosis – was coordinated by Alexei Korobitsyn and Patricia Hall-Eidson, with the support of Carl-Michael Nathanson, under the guidance of Matteo Zignol and the overall direction of Tereza Kasaeva, Director of the WHO Global Programme on Tuberculosis & Lung Health (WHO/GTB).

Low complexity nucleic acid amplification testing for detection of TB and resistance to rifampicin

Concurrent testing

WHO steering committee

The preparation of this section of the guidelines was overseen by Alexei Korobitsyn with input from Annabel Baddeley, Annemieke Brands, Dennis Falzon, Tereza Kasaeva, Cecily Miller, Carl-Michael Nathanson, Jasmine Solangon, Maria de los Angeles Vargas, Sabine Vercuijl, Kerri Viney, and Matteo Zignol (all from the WHO Global TB Programme); Ajay Rangaraj and Elena Vovc (from the WHO Global HIV Programme); Soudeh Ehsani (WHO Regional Office for Europe); and Amos Fadare (WHO Country Office for Nigeria).

Guideline development group

Kobto Koura (Co-Chair), International Union Against Tuberculosis and Lung Disease, Paris, France; Jeremiah Chakaya Muhwa, Respiratory Society of Kenya, Nairobi, Kenya; Chamreun Sok Choub, Patients Advocate, Pnom Penh, Cambodia; Daniela Cirillo, San Raffaele Scientific Institute, Milan, Italy; Keertan Dheda, University of Cape Town, Cape Town, South Africa; Katherine Fielding, London School of Hygiene and Tropical Medicine, London, United Kingdom of Great Britain and Northern Ireland; Rumina Hasan, Aga Khan University, Karachi, Pakistan; Sirinapha Jittimanee, NTP Thailand Faculty of Nursing, Chulalongkorn University, Bangkok, Thailand; Katharina Kranzer, London School of Hygiene and Tropical Medicine, London, United Kingdom of Great Britain and Northern Ireland; Andrei Mariandyshev, Northern State Medical University, Ministry of Health of Russia, Arkhangelsk, Russian Federation; Norbert Ndjeka,

National TB Programme of South Africa, Pretoria, South Africa; Shaheed Vally Omar, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa; Hojoon Sohn, Seoul National University, Seoul, Republic of Korea, Thomas Shinnick, Consultant, Atlanta, USA; Sabira Tahseen, National Tuberculosis Control Programme, Ministry of National Health Services Regulations and Coordination, Government of Pakistan, Islamabad, Pakistan; Timothy Walker, University of Oxford, Ho Chi Minh City, Viet Nam; Xichao Ou, National Tuberculosis Control and Prevention Center, Chinese CDC; Beijing, China.

External review group

Martina Casenghi (Elizabeth Glaser Pediatric AIDS Foundation, Washington, USA), Jamilya Ismailova (National TB Program, Dushanbe, Tajikistan), Blessy Kumar (Global coalition of TB advocates, Delhi, India), Fernanda Dockhorn da Costa Johansen (National TB Program, Rio de Janeiro, Brazil), Petra de Haas (KNCV, The Hague, Netherlands), and Platon Eliseev (National Center of Phtisiopulmonology, Moscow, Russian Federation).

Methodologist

Lawrence Mbuagbaw, McMaster University, Hamilton, Canada

Systematic review team

Review of diagnostic accuracy: Johanna Ahsberg (University of Southern Denmark, Odense, Denmark), Stephanie Bjerrum (University of Southern Denmark, Odense, Denmark), David Horne (University of Washington, Seattle, USA), Leeberk Raja Inbaraj (ICMR-National Institute for Research in Tuberculosis, Chennai, India), Nazir Ismail (Wits University, Johannesburg, South Africa), Alexander Kay (Baylor College of Medicine, Houston, USA), Mikashmi Kohli (FIND, Geneva Switzerland), Maia Madison (Baylor College of Medicine, Houston, USA), Laura Olbrich (LMU Klinikum, Munich, Germany), Katie Scandrett (University of Birmingham, Birmingham, United Kingdom of Great Britain and Northern Ireland), Maunank Shah (John Hopkins University, Elkridge, USA), Yemisi Takwoingi (University of Birmingham, Birmingham, United Kingdom of Great Britain and Northern Ireland)

Review of research into costs and cost effectiveness: Suvesh Shrestha and Alice Zwerling (University of Ottawa, Canada).

Review of and research into user perspectives, feasibility, and acceptability: Nora Engel (University of Maastricht, Maastricht, Netherlands) and Eleanor Ochodo (Kenya Medical Research Institute, Kisumu, Kenya).

Observers

Grania Bridgen Global Fund to Fight AIDS, TB and Malaria, Geneva, Switzlerland; Smiljka de Lussigny, Unitaid, Geneva, Switzerland; Anisa Ghadrshenas, Unitaid, Geneva, Switzerland; Brian Kaiser, Global Drug Facility, Stop TB Partnership, Geneva, Switzerland; Melanie Kitongo, Global Fund to Fight AIDS, TB and Malaria, Geneva, Switzerland; Adam Penn-Nicholson, FIND, Geneva Switzerland; Morten Ruhwald, FIND, Geneva, Switzerland.

Funding

Funding from the Gates Foundation is gratefully acknowledged. The views of the funding agencies have not influenced the development and content of these guidelines.

Abbreviations and acronyms

AIDS	acquired immunodeficiency syndrome
AlereLAM	Alere Determine TB LAM Ag
aNAAT	automated nucleic acid amplification test
BAL	bronchoalveolar lavage
BCG	bacille Calmette–Guérin
BD	Becton Dickinson
cfu	colony forming units
CI	confidence interval
CrI	credible interval
CRS	composite reference standard
CSF	cerebrospinal fluid
DALY	disability-adjusted life year
DIAMA	Diagnostics for Multidrug Resistant Tuberculosis in Africa
DNA	deoxyribonucleic acid
DR-TB	drug-resistant tuberculosis
DST	drug susceptibility testing
ERG	External external review group
FIND	Foundation for Innovative New Diagnostics
FL-LPA	first-line line probe assay
GDG	guideline development group
Global Fund	Global Fund to Fight AIDS, Tuberculosis and Malaria
GRADE	Grading of Recommendations Assessment, Development and EvaluationHIV human immunodeficiency virus
ICER	incremental cost–effectiveness ratio
IGRA	interferon-gamma release assay
IVD	in vitro diagnostic
LAM	lipoarabinomannan
LAMP	loop-mediated isothermal amplification
LC-aNAAT	low-complexity automated nucleic acid amplification test
LC-mNAAT	low-complexity manual nucleic acid amplification test
LC-NAAT	low-complexity nucleic acid amplification test
LF-LAM	lateral flow urine lipoarabinomannan assay
LMIC	low- and middle-income countries
LPA	line probe assay
LSHTM	London School of Hygiene & Tropical Medicine
MC-aNAAT	moderate-complexity automated nucleic acid amplification test
MDR	multidrug-resistant

MDR/RR-TB	multidrug-resistant tuberculosis or rifampicin-resistant tuberculosis
MDR-TB	multidrug-resistant tuberculosis
MRS	microbiological reference standard
Mtb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
mWRD	molecular WHO-recommended rapid diagnostic test
NAAT	nucleic acid amplification test
NAT	nucleic acid test
NGS	next-generation sequencing
NTP	national tuberculosis programme
PCR	polymerase chain reaction
PI	prediction interval
PICO	population, intervention, comparator and outcome
PQ	prequalification
PZA	pyrazinamide
QES	quality evidence synthesis
QUADAS	quality assessment of diagnostic accuracy studies
RIF	rifampicin
RNA	ribonucleic acid
RRDR	rifampicin-resistance determining region
RR-TB	rifampicin-resistant tuberculosis
SL-LPA	second-line line probe assay
SRL	supranational TB reference laboratory
SSM	sputum smear microscopy
STARD	Standards for Reporting Diagnostic Accuracy Studies
TB	tuberculosis
TBST	*Mtb* antigen-based skin test
TPT	tuberculosis preventive treatment
TST	tuberculin skin test
UN	United Nations
United Kingdom	United Kingdom of Great Britain and Northern Ireland
UR	uncertainty range
USA	United States of America
USAID	United States Agency for International Development
UV	ultraviolet
WGS	whole genome sequencing
WHO	World Health Organization
WHO/GTB	Global Programme on Tuberculosis & Lung Health of the World Health Organization
WRD	WHO-recommended rapid diagnostic test
WTP	willingness-to-pay
XDR	extensively drug-resistant
XDR-TB	extensively drug-resistant tuberculosis

Definitions

Advanced HIV disease: for adults, adolescents, and children aged 5 years or more, "advanced HIV disease" is defined as a CD4 cell count of less than 200 cells/mm^3 or a WHO clinical stage 3 or 4 event at presentation for care. All children living with HIV aged under 5 years should be considered as having advanced disease at presentation.

Age groups: the following definitions for adults and children are used in these guidelines for the purpose of implementing recommendations (countries may have other definitions under their national regulations):

- an adult is a person aged 10 years and older;
- a child is a person aged under 10 years.

Grading of Recommendations Assessment, Development and Evaluation (GRADE): a system for rating quality of evidence and strength of recommendations; the GRADE approach is explicit, comprehensive, transparent and pragmatic, and is increasingly being adopted by organizations worldwide.

HIV serious illness: HIV serious illness is defined based on any of the following symptoms: respiratory rate of ≥ 30/minute, temperature ≥ 39 °C, heart rate ≥ 120/minute, or unable to walk unaided.

Inpatient health care setting: a health care facility where patients are admitted and assigned a bed while undergoing diagnosis and receiving treatment and care, for at least one overnight stay.

Outpatient health care setting: a health care facility where patients are undergoing diagnosis and receiving treatment and care but are not admitted for an overnight stay (e.g. an ambulatory clinic or a dispensary).

Executive summary

It is estimated that about a quarter of the world's population is infected with *Mycobacterium tuberculosis* – the bacterium that causes tuberculosis (TB) disease. Testing for TB infection can identify individuals who would benefit the most from TB preventive treatment (TPT). However, despite the availability of preventive measures and disease treatment, TB remains a leading cause of death due to a single infectious agent. TB has probably replaced coronavirus disease (COVID-19) as the leading cause of death worldwide for the first time since the start of the global pandemic *(1)*.

In recognition of the need to end TB globally, the United Nations (UN) held the world's first high-level meeting on TB in 2018. The political declaration from the meeting included commitments by Member States to achieving four new global targets *(2)*, which were subsequently renewed at the second UN high-level meeting in 2023. The commitments included two that relied on diagnosis of TB infection and disease: providing TPT to at least 45 million people between 2024 and 2027, and reaching 90% of the estimated number of people who develop TB with quality-assured diagnosis and treatment from 2023 to 2027 *(2)*. These commitments align with the World Health Organization's (WHO's) End TB Strategy, which calls for the detection of individuals living with TB infection who are at higher risk of progression to active TB so that they can receive TPT, as well as the early diagnosis of TB and drug-resistant TB (DR-TB) through universal drug susceptibility testing (DST). These global commitments and plans highlight the critical role of TB testing for the rapid and accurate detection of TB infection, disease and drug resistance *(3)*.

To support countries in their efforts to strengthen detection of TB infection, disease and drug resistance, the WHO Global TB Programme issues evidence-based policy guidance on TB testing strategies and technologies; this guidance is routinely updated. Since the most recent consolidated guidelines on TB diagnosis were issued in 2024:

- new evidence has become available on the use of WHO-recommended rapid diagnostic tests (WRDs) for the initial detection of TB and resistance to rifampicin among populations that are at increased risk of TB-related morbidity and mortality (e.g. people living with HIV and children);
- a systematic assessment of evidence on molecular WRDs (mWRDs) previously recommended as individual products was completed to determine the placement of mWRDs within existing or new classes of TB diagnostic technologies; and
- a call from countries was received to combine the policy guidance on TB infection, disease and drug-resistance testing into these consolidated guidelines on TB diagnosis, to streamline implementation of national testing programmes.

In response, this document is being issued as the fourth edition of the consolidated guidelines on TB diagnosis. When compared with the third edition (issued in 2024), this guideline is the first to combine the WHO policy guidance on diagnosis of TB infection, disease and drug resistance into a single reference document; also, it establishes two new classes of TB diagnostic

technologies (for the initial detection of TB and resistance to rifampicin), and outlines new recommendations on concurrent testing of respiratory and non-respiratory samples among adults and adolescents with HIV, children with HIV, and children without HIV or with unknown HIV status. The main changes from the previous WHO guidelines are summarized in **Box A**.

The set of 21 new and existing recommendations for diagnosis of TB infection, disease and DR-TB are presented in **Table A**. These recommendations supersede those presented in previous editions of the guidelines and are supported by updated operational guidance that is published as the fourth edition of the *WHO operational handbook on tuberculosis. Module 3: diagnosis*. The operational handbook includes further details on the individual tests that are recommended for use; the selection, introduction and implementation of tests for TB infection, diagnosis and drug resistance; and updated diagnostic algorithms that reflect the updates contained within these guidelines.

Box A. Main changes to the guidance in this update

- Two new classes of TB diagnostic tests for the initial detection of TB and resistance to rifampicin were established; these classes differ in the level of procedure and test result automation, and include tests that were previously recommended as standalone products. The new low-complexity automated nucleic acid amplification test (LC-aNAAT) class includes the Xpert® MTB/RIF and Xpert MTB/RIF Ultra assays, and the Truenat® MTB Plus and MTB-RIF Dx assays. The low-complexity manual nucleic acid amplification test (LC-mNAAT) class includes the Loopamp™ MTBC Detection Kit (TB LAMP) (Eiken Chemical). These new class-based recommendations supersede previous product-specific recommendations.
- Concurrent testing of respiratory and non-respiratory samples for the initial detection of TB and resistance to rifampicin is newly recommended for adults and adolescents living with HIV, children living with HIV, and children without HIV or with unknown HIV status.
- Existing guidelines on tests for TB infection were added, to consolidate policy guidance on testing for TB diagnosis, drug resistance and infection.
- A description of TB diagnostic test determination and the pathways for TB diagnostic product prequalification by WHO was added to the Background section.
- TB diagnostic test class description tables were revised to align with the class-determination criteria presented in the Background section.
- The four prior web annexes covering systematic review and guideline development group (GDG) evidence to inform policy updates were consolidated into two web annexes. Web Annex A includes the systematic reviews, Grading of Recommendations Assessment, Development and Evaluation (GRADE) tables and evidence to decision (EtD) tables, and Web Annex B includes the evidence synthesis and analysis findings. Both web annexes now present content by TB diagnostic class.

Table A. Recommendations in the WHO consolidated guidelines on tuberculosis. Module 3: diagnosis, fourth edition

NEW

1. For adults and adolescents with signs or symptoms of TB or who screened positive[1] for pulmonary TB, low-complexity **automated** NAATs should be used on respiratory samples as initial diagnostic tests for TB, rather than smear microscopy or culture.

   (Strong recommendation, high certainty of evidence)

NEW

2. For people with bacteriologically confirmed TB[2], low-complexity **automated** NAATs should be used on respiratory samples as initial tests for detection of resistance to rifampicin, rather than culture-based DST.

   (Strong recommendation, high certainty of evidence)

NEW

3. For people with signs and symptoms of TB meningitis, low-complexity **automated** NAATs on cerebral spinal fluid should be used for the initial diagnosis of TB meningitis, rather than smear microscopy or culture.

   (Strong recommendation, moderate certainty of evidence)

NEW

4. For people with signs and symptoms of extrapulmonary TB, low-complexity **automated** NAATs on lymph node tissue aspirate, pleural tissue, pleural fluid, synovial fluid, peritoneal fluid or pericardial fluid should be used for the initial diagnosis of TB, rather than smear microscopy or culture.

   (Strong recommendation, low certainty of evidence for synovial fluid and pericardial fluid; very low certainty of evidence for lymph node tissue aspirate, pleural tissue, pleural fluid and peritoneal fluid)

5. For people with signs and symptoms of pulmonary TB, moderate-complexity automated NAATs may be used on respiratory samples for the detection of pulmonary TB, and of rifampicin and isoniazid resistance, rather than culture and phenotypic DST.

   (Conditional recommendation, moderate certainty of evidence)

NEW

6. For adults and adolescents with signs or symptoms or who screen positive for pulmonary TB, low-complexity **manual** NAATs should be used on respiratory samples as initial diagnostic tests for TB, rather than smear microscopy or culture.

   (Strong recommendation, high certainty of evidence)

---

[1] Having a positive result of a test, examination or other procedure used to distinguish people with a high likelihood of having TB disease from people who are highly unlikely to have TB. At present, the following tests are WHO-recommended as the screening tests: chest radiography (chest X-ray; CXR) with or without computer-aided detection (CAD), C-reactive protein (CRP) in people living with HIV, and molecular WHO-recommended rapid diagnostic test for TB (mWRD) (https://www.who.int/publications/i/item/9789240022676).

[2] A bacteriologically confirmed TB case is one from whom a biological specimen is positive by smear microscopy, culture or WRD (such as Xpert MTB/RIF). All such cases should be notified, regardless ofwhether TB treatment has started (https://www.who.int/publications/i/item/9789241505345).

NEW

7. For adults and adolescents with HIV who have signs or symptoms of TB, screen positive for TB, are seriously ill or have advanced HIV disease, **concurrent testing** using low-complexity automated NAATs on respiratory samples and LF-LAM on urine should be used as the initial diagnostic strategy for diagnosing TB, rather than low-complexity automated NAATs on respiratory samples alone.

   (Strong recommendation, low certainty of evidence)

NEW

8. For children who are HIV-negative or have an unknown HIV status, who have signs or symptoms or screen positive for pulmonary TB, **concurrent testing** using low-complexity automated NAATs on respiratory and stool samples should be used as the initial diagnostic strategy for diagnosing TB, rather than low-complexity automated NAATs on respiratory or stool samples alone.

   (Strong recommendation, low certainty of evidence)

NEW

9. For children with HIV who have signs or symptoms or screen positive for pulmonary TB, **concurrent testing** using low-complexity automated NAATs on respiratory and stool samples and LF-LAM on urine may be used as the initial diagnostic strategy for diagnosing TB, rather than low-complexity automated NAATs on respiratory or stool samples alone.

   (Conditional recommendation, low certainty of evidence)

10. For people with bacteriologically confirmed pulmonary TB, low-complexity automated NAATs may be used on sputum for the initial detection of resistance to isoniazid and fluoroquinolones, rather than culture-based phenotypic DST.

    (Conditional recommendation, moderate certainty of evidence)

11. For people with bacteriologically confirmed pulmonary TB and resistance to rifampicin, low-complexity automated NAATs may be used on sputum for the initial detection of resistance to ethionamide, rather than DNA sequencing of the *inhA* promoter.

    (Conditional recommendation, very low certainty of evidence)

12. For people with bacteriologically confirmed pulmonary TB and resistance to rifampicin, low-complexity automated NAATs may be used on sputum for the initial detection of resistance to amikacin, rather than culture-based phenotypic DST.

    (Conditional recommendation, low certainty of evidence)

13. For people with a sputum smear-positive specimen or a cultured isolate of MTBC, commercial molecular LPAs may be used as the initial test instead of phenotypic culture-based DST to detect resistance to rifampicin and isoniazid.

    (Conditional recommendation, moderate certainty of evidence)

14. For people with confirmed MDR/RR-TB, SL-LPA may be used as the initial test, instead of phenotypic culture-based DST, to detect resistance to fluoroquinolones.

    (Conditional recommendation, moderate certainty of evidence for test accuracy)

15. For people with confirmed MDR/RR-TB, SL-LPA may be used as the initial test, instead of phenotypic culture-based DST, to detect resistance to the SLIDs.

(Conditional recommendation, low certainty of evidence for test accuracy)

16. For people with bacteriologically confirmed TB, high-complexity reverse hybridization-based NAATs may be used on *Mtb* culture isolates for detection of pyrazinamide resistance rather than culture-based phenotypic DST.

(Conditional recommendation, very low certainty of evidence)

17. For people with bacteriologically confirmed pulmonary TB disease, targeted next-generation sequencing technologies may be used on respiratory samples to diagnose resistance to rifampicin, isoniazid, fluoroquinolones, pyrazinamide and ethambutol, rather than culture-based phenotypic DST.

(Conditional recommendation, certainty of evidence moderate [isoniazid and pyrazinamide] and low [rifampicin, fluoroquinolones and ethambutol])

18. For people with bacteriologically confirmed rifampicin-resistant pulmonary TB disease, targeted next-generation sequencing technologies may be used on respiratory samples to diagnose resistance to isoniazid, fluoroquinolones, bedaquiline, linezolid, clofazimine, pyrazinamide, ethambutol, amikacin and streptomycin, rather than culture-based phenotypic DST.

(Conditional recommendation, certainty of evidence high [isoniazid, fluoroquinolones and pyrazinamide], moderate [ethambutol], low [bedaquiline, linezolid, clofazimine and streptomycin] and very low [amikacin])

19. *Mycobacterium tuberculosis* antigen-based skin tests may be used to test for TB infection.

(Conditional recommendation, very low certainty of evidence for test accuracy)

20. Either a tuberculin skin test or an interferon-gamma release assay can be used to test for TB infection.

(Strong recommendation, very low certainty of evidence for test accuracy)

21. Interferon-gamma release assays (IGRAs) and tuberculin skin tests (TSTs) should not be used in low- and middle-income countries for the diagnosis of pulmonary or extrapulmonary TB or for the diagnostic work-up of adults (including people living with HIV) with suspected active TB.

(Strong recommendation)

DNA: deoxyribonucleic acid; DST: drug susceptibility testing; HIV: human immunodeficiency virus; LF-LAM: lateral flow urine lipoarabinomannan assay; LPA: line probe assay; MDR/RR-TB: multidrug-resistant or rifampicin-resistant TB; *Mtb*: *Mycobacterium tuberculosis*; MTBC: *Mycobacterium tuberculosis* complex; NAAT: nucleic acid amplification test; NGS: next-generation sequencing; SL-LPA: second-line line probe assay; SLID: second-line injectable drug; TB: tuberculosis; WHO: World Health Organization.

1. Introduction

1.1. Background

It is estimated that about a quarter of the world's population is infected with *Mycobacterium tuberculosis* (*Mtb*) – the bacterium that causes tuberculosis (TB) disease. Testing for TB infection can identify individuals who would benefit the most from TB preventive treatment (TPT). Without TPT, it is estimated that about 5–10% of people who are infected will develop TB disease over the course of their lives, usually within 5 years of the initial infection *(1)*.

Despite the availability of preventive measures and disease treatment, TB remains a leading cause of death due to a single infectious agent, and has probably replaced coronavirus disease (COVID-19) as the leading cause of death worldwide for the first time since the start of the global pandemic *(1)*. In 2023, it is estimated that 10.8 million people fell ill with TB, but only 8.2 million were diagnosed. In addition, resistance to the antibiotics that are used to treat TB remains a challenge, with an estimated 400 000 people (95% uncertainty interval [UI]: 370 000–450 000) having developed either rifampicin-resistant TB (RR-TB), or TB resistant to both rifampicin and isoniazid, defined as multidrug-resistant TB (MDR-TB).

In 2018, the United Nations (UN) held the world's first high-level meeting on TB. The political declaration from the meeting included commitments by Member States to achieve four new global targets *(2)*. These commitments were subsequently renewed at the second UN high-level meeting in 2023; they included provision of TPT to at least 45 million people between 2024 and 2027, and reaching 90% of the estimated number of people who develop TB with quality-assured diagnosis and treatment from 2023 to 2027 *(3)*. In addition, the World Health Organization's (WHO's) End TB Strategy calls for the detection of individuals living with TB infection who are at higher risk of progression to active TB so that they can receive TPT, as well as the early diagnosis of TB and universal drug susceptibility testing (DST). These global commitments and plans highlight the critical role of TB testing for the rapid and accurate detection of TB infection, disease and drug resistance *(4)*.

Over recent years, the WHO Global TB Programme (WHO/GTB) has issued evidence-based policy guidance on diagnostic testing, to support countries in their efforts to detect TB infection, disease and drug resistance. When novel diagnostic tools are developed and evidence on their use and impact becomes available, WHO/GTB commissions systematic reviews and convenes guideline development groups (GDGs) to inform guideline updates. Since 2021, these updates have been issued in module-based consolidated guidelines. Until 2024, policy recommendations for testing for TB infection, TB disease and drug resistance were presented separately (in the consolidated guidelines on prevention and diagnosis, respectively).

This document is the fourth edition of WHO policy guidelines on TB diagnosis. Compared with the third edition, issued in 2024, this document:

- is the first to combine guidance on diagnosis of TB infection, disease and drug resistance into a single reference document;
- establishes two new classes of TB diagnostic technologies (for the initial detection of TB and resistance to rifampicin), which include tests previously recommended for use as individual products; and
- outlines new recommendations on concurrent testing of respiratory and non-respiratory samples among adults and adolescents with HIV, children with HIV, and children without HIV or with unknown HIV status.

1.2. WHO TB diagnostic class determination and product prequalification

Over the past 16 years, WHO has endorsed a range of diagnostic technologies (**Table 1.1.1**). The WHO assessment process for TB diagnostics has recently evolved to focus on evaluating classes of TB diagnostic technologies rather than specific products. Class determination is managed by WHO/GTB for new diagnostic testing technologies, and it includes an evaluation of the following characteristics:

- purpose of use (i.e. detection of TB or drug-specific resistance);
- principle of action;
- infrastructure and human resource requirements;
- complexity of the testing procedure and associated instrumentation;
- reporting method (automated versus manual); and
- intended setting of use (e.g. reference or peripheral low-complexity, near point-of-care).

These characteristics are compared between the new technology and each of the existing classes already recommended by WHO. When characteristics differ from existing classes, the new technology will undergo an evidence review as "first-in-class" via Pathway A (described below). When characteristics match those of an existing class, the new technology will undergo a "within-class" assessment (Pathway B below).

1.2.1 Pathway A

New technologies will require a Pathway A review if they differ from technologies in existing classes in terms of the characteristics listed above. Evidence synthesis and review and development of recommendations will be conducted through the established WHO/GTB guideline development process using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology. If recommended by a GDG after evidence review, technologies will be referred for WHO prequalification assessment (as available). If a prequalification assessment procedure is not available, the WHO/GTB recommendation will stand until the prequalification procedure becomes available and is successfully completed.

1.2.2 Pathway B

Technologies will require a Pathway B review if they share characteristics with an existing class and are therefore not first-in-class. Review of these within-class technologies depends on availability of a prequalification assessment procedure for the class:

- If a prequalification assessment procedure is available, manufacturers may proceed directly with assessment.
- If a prequalification assessment procedure is not yet available, an evidence review will be conducted through a WHO/GTB evidence assessment process, facilitated by the Technical Advisory Group on TB Diagnostics and Laboratory Strengthening. If recommended by WHO/GTB, the technology will be added to the relevant class in the latest policy guidance. The recommendation will stand until the prequalification assessment procedure becomes available and is successfully completed.

1.3. Testing classes and products

As highlighted above, all technologies with a WHO/GTB recommendation are expected to undergo prequalification assessment, as available. Successful assessment will be required to maintain a WHO/GTB recommendation. The current set of TB diagnostic testing classes and included products are listed in **Table 1.1.1**, and the two new classes are discussed below.

Table 1.1.1. Classes and products of TB tests for detection of TB, drug-resistant TB and TB infection included in the current guidelines

Technology class	Included products
Initial tests for TB diagnosis with drug-resistance detection	
NEW: Low-complexity automated nucleic acid amplification tests (NAATs) for detection of TB and resistance to rifampicin	Xpert® MTB/RIF and Xpert MTB/RIF Ultra (Cepheid) Truenat® MTB Plus and Truenat MTB-RIF Dx (Molbio)
Moderate-complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid	Abbott RealTi*m*e® MTB and Abbott RealTi*m*e MTB RIF/INH (Abbott) BD MAX™ MDR-TB (Becton Dickinson) cobas® MTB and cobas MTB-RIF/INH (Roche) FluoroType® MTB and FluoroType MTBDR (Hain Lifescience/Bruker)
Initial tests for TB diagnosis without drug-resistance detection	
NEW: Low-complexity manual NAATs for detection of TB	Loopamp™ MTBC Detection Kit (TB LAMP) (Eiken Chemical)
Antigen detection in a lateral flow format (biomarker-based detection) (LF-LAM) for detection of TB	Determine™ TB LAM Ag (Alere/Abbott)

Technology class	Included products
Follow-on tests for detection of TB drug resistance	
Low-complexity automated NAATs for detection of resistance to isoniazid and second-line anti-TB agents	Xpert® MTB/XDR (Cepheid)
Line probe assays (LPAs) for detection of TB drug resistance	GenoType® MTBDR*plus* v1 and v2; and GenoType MTBDRsl (Hain Lifescience/Bruker) Genoscholar™ NTM+MDRTB II and Genoscholar PZA-TB II (Nipro)
Targeted next-generation sequencing (NGS) tests for detection of TB drug resistance	Deeplex® Myc-TB (GenoScreen/Illumina) AmPORE-TB® (Oxford Nanopore Technologies) TBseq® (Shengting Medical Technology Company)
Tests for TB infection	
Mycobacterium tuberculosis antigen-based skin tests (TBSTs)	Diaskintest® (Generium) Siiltibcy™ (Serum Institute of India) C-TST (Anhui Zhifei Longcom)
Interferon-gamma release assays (IGRAs)	T-SPOT.TB (T-Spot) (Revvity) TB-IGRA (Wantai BioPharm) QuantiFERON-TB Gold Plus (QFT-Plus) (QIAGEN) STANDARD E TB-Feron ELISA (SD BIOSENSOR)[3] LIAISON QFT-Plus CLIA (Diasorin)[3]
Tuberculin skin tests	Tuberculin purified protein derivative (PPD) products

NAAT: nucleic acid amplification test; TB: tuberculosis.

1.3.1 Initial tests for TB diagnosis with drug resistance detection

Low-complexity automated nucleic acid amplification tests (NAATs) for detection of TB and resistance to rifampicin

The low-complexity automated nucleic acid amplification tests (LC-aNAATs) include tools such as Xpert® MTB/RIF Ultra (Cepheid) and Truenat® MTB Plus with MTB-RIF Dx (Molbio). These tests provide largely automated solutions suitable for decentralized laboratories, and are currently the most widely used tests for the initial detection of TB and resistance to rifampicin. The testing instruments use software and hardware (computers) to report results, and they require well-established laboratory networks and trained personnel.

[3] For WHO statement and evidence assessment on new IGRAs see WHO operational handbook on tuberculosis. Module 3: diagnosis.

Moderate-complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid

The moderate-complexity automated NAATs (MC-aNAATs) are faster and less complex to perform than phenotypic culture-based DST and line probe assays (LPAs), and are largely automated after the sample preparation step. They may be used as an initial test for simultaneous detection of TB and resistance to rifampicin and isoniazid. This type of NAAT offers the potential for rapid provision of accurate results and for testing efficiency where high volumes of tests are required daily. Hence, they are suited to areas with a high population density and rapid sample referral systems.

1.3.2 Initial tests for TB diagnosis without drug resistance detection

Low-complexity manual NAATs

The low-complexity manual NAAT (LC-mNAAT), loop-mediated isothermal amplification (LAMP), is based on DNA amplification at a single temperature range; this contrasts with the polymerase chain reaction (PCR), which requires a thermocycler. Detection of amplified product is done visually, using an ultraviolet (UV) lamp, directly in the reaction tubes. The method requires only basic equipment and can be implemented at the lowest levels of the laboratory network. However, detection of mutations in resistance-associated genes is not available with the currently recommended technology.

Antigen detection in a lateral flow format (biomarker-based detection)

The currently available lateral flow urine lipoarabinomannan assay (LF-LAM) has suboptimal sensitivity and specificity; thus, it is not suitable as a diagnostic test for TB in all populations. However, in contrast to traditional diagnostic methods, the urine LF-LAM assay demonstrates improved sensitivity for the diagnosis of TB among individuals coinfected with HIV.

1.3.3 Follow-on tests for detection of TB drug resistance

Low-complexity automated NAATs for the detection of resistance to isoniazid and second-line anti-TB agents

The LC-aNAATs are recommended for use as a reflex test in specimens determined to be positive for *Mtb* complex (MTBC); these tests offer rapid DST in intermediate and peripheral laboratories. The first product in this class simultaneously detects resistance to isoniazid, fluoroquinolones, ethionamide and amikacin. Results are available in under 90 minutes; this is faster than with the current standard of care, which includes LPAs and culture-based phenotypic DST.

Line probe assays

LPAs are a family of DNA strip-based tests that can detect the MTBC DNA and determine its drug-resistance profile. The tests do this through the pattern of binding of amplicons (DNA amplification products) to probes that target specific parts of the MTBC genome; that is, common resistance-associated mutations to anti-TB drugs or the corresponding wild-type DNA sequence *(5)*. LPAs are technically more complex to perform than the Xpert MTB/RIF assay; however, they can detect resistance to a broader range of first-line and second-line

agents, and they provide mutation-specific data for common variants. Testing platforms have been designed for a reference laboratory setting and are most applicable to high TB burden countries. Results can be obtained in 5 hours *(5)*.

Targeted next-generation sequencing tests

Tests based on targeted next-generation sequencing (NGS) are used for follow-on detection of resistance to a broad range of anti-TB drugs after the initial detection of TB or of rifampicin resistance. This class of tests is based on technology that combines amplification of selected genes with NGS to detect resistance to many drugs with a single test. Because targeted NGS can interrogate entire genes to identify specific mutations associated with resistance, the accuracy may be better than that of existing WHO-recommended rapid diagnostic tests (WRDs). In addition, new tests based on targeted NGS can detect resistance to new and repurposed drugs that are not currently included in any other molecular assays. Hence, this class of tests offers great potential to provide comprehensive resistance detection matched to modern treatment regimens.

1.3.4 Tests for TB infection

Mtb antigen-based skin tests

Mtb antigen-based skin tests (TBSTs) are used for the indirect detection of TB infection. TBSTs rely on intradermal injection of *Mtb*-specific antigens; the antigens elicit a localized skin reaction in infected individuals that is detected by measurement of a local induration 48–72 hours after administration. Although these tests continue to rely on patient injection and return visits for result interpretation, they are more specific than the WHO-recommended tuberculin skin tests (TSTs).

Interferon-gamma release assays

Interferon-gamma release assays (IGRAs) are in vitro blood-based tests that are used to indirectly test for TB infection. They do this by measuring either the amount of interferon-gamma that is released by lymphocytes in whole blood after exposure to *Mtb*-specific antigens or the number of T-lymphocytes within the whole blood that produce interferon-gamma. IGRA testing requires days to perform owing to the blood incubation steps and it can be challenging to perform among patients for whom phlebotomy can be difficult (e.g. children); however, this is the only type of test for TB infection in which the results are not affected by prior bacille Calmette–Guérin (BCG) vaccination for TB. Hence, IGRAs are a promising alternative for detection of TB infection in settings with high rates of BCG vaccination.

Tuberculin skin tests

TSTs were the first class of tests to be recommended for detection of TB infection; they rely on intradermal injection of a mix of antigens to *Mtb*, non-tuberculous mycobacteria and the BCG vaccine formulation, followed by detection of a localized skin induration-based response after 48–72 hours. As with the TBSTs, these tests can facilitate TB infection testing in children and other patients for whom phlebotomy is challenging, but they may also produce false positive results in people infected with mycobacteria other than TB and in those who are BCG vaccinated.

Regulatory approval from national regulatory authorities or other relevant bodies is required before implementation of new diagnostic tests.

1.4. Scope of the document

This document provides background, justification and recommendations for novel diagnostic tools for detecting *MTBC,* the presence or absence of mutations in target genes proven to be associated with anti-TB drug resistance, and TB infection.

1.5. Target audience

The target audience for these guidelines includes laboratory managers, clinicians and other health care staff, HIV and TB programme managers, policy-makers, technical agencies, donors and implementing partners supporting the use of TB diagnostic tests in resource-limited settings.

The document may also be of use to individuals responsible for programme planning, budgeting, mobilizing resources and implementing training activities for the programmatic management of drug-resistant TB (DR-TB).

1.6. Scope of the document

This document provides background, justification and recommendations on novel diagnostic tools for detecting MTBC, the presence or absence of mutations in target genes proven to be associated with anti-TB drug resistance, and TB infection.

1.7. Target audience

The target audience for these guidelines includes laboratory managers, clinicians and other health care staff, HIV and TB programme managers, policymakers, technical agencies, donors and implementing partners supporting the use of TB diagnostics in resource-limited settings.

Individuals responsible for programme planning, budgeting, mobilizing resources and implementing training activities for the programmatic management of DR-TB may also find this document useful.

2. Recommendations for diagnosis of TB disease

2.1. Initial diagnostic tests for diagnosis of TB with drug-resistance detection

2.1.1 LC-aNAATs for detection of TB and resistance to rifampicin

NEW

Rapid detection of TB and rifampicin resistance is a critical global priority. Over a decade ago, the first recommendation on molecular testing for the diagnosis of TB and detection of resistance significantly transformed the TB diagnostic landscape. These technologies have proven highly accurate compared with smear microscopy, and they can detect rifampicin resistance rapidly. They do not require highly skilled individuals or designated molecular laboratory infrastructure for testing. In addition, they are largely automated after sample loading, up to the final report generation. These features make this class of low-complexity automated tests appealing for use in low- and middle-income countries (LMIC).

Uptake of these technologies has been slowed by barriers related to costs, the supply chain, equipment maintenance and technical support. The lack of a healthy competitive environment has also been a contributory factor. The WHO Prequalification (PQ) programme for TB in vitro diagnostics (IVDs) has opened a pathway to allow more products to come to market and ensure quality. The current guidelines facilitated this process with the introduction of class-based recommendations for low-complexity NAATs. WHO PQ assessment progress for all low-complexity NAATs is reported on the WHO PQ website.[4]

[4] In Vitro Diagnostics Under Assessment | WHO – Prequalification of Medical Products (IVDs, Medicines, Vaccines and Immunization Devices, Vector Control).

Diagnostic class description

The features shown in **Table 2.1.1.1** define the class of LC-aNAATs.:

Table 2.1.1.1 Class criteria for LC-aNAATs

Purpose		Detection of TB and rifampicin resistance
Principle of action		Nucleic acid amplification testing
Complexity	Reagents	Most reagents are enclosed in a disposable sealed container to which a clinical specimen is added. The disposable sealed container does not have special storage requirements
	Skills	Basic technical skills (e.g. basic pipetting, precision not critical)
	Pipetting	Either no, or only one, pipetting step in the process
	Testing procedure	• May require an initial manual specimen treatment step before transferring the specimen into the disposable sealed container for automated processing • Automated DNA extraction • Automated real-time PCR • Results generation
Type of test result reporting		Automated
Setting of use		Basic laboratory (no special infrastructure needed)

DNA: deoxyribonucleic acid; LC-aNAAT: low-complexity automated nucleic acid amplification test; PCR: polymerase chain reaction; TB: tuberculosis.

The products for which eligible data met the class-based performance criteria for LC-aNAATs were:

- Xpert MTB/RIF Ultra (Cepheid, Sunnyvale, United States of America [USA]) – for pulmonary TB, extrapulmonary TB and resistance to rifampicin; and
- Truenat MTB Plus and Truenat MTB-RIF Dx (Molbio, Goa, India) – for pulmonary TB and resistance to rifampicin.

Data on Truenat MTB Plus and MTB-RIF Dx were more limited than those for Xpert Ultra.

Regulatory approval from national regulatory authorities or other relevant bodies is required before implementation of these diagnostic tests. Extrapolation to other brand-specific tests cannot be made, and any new in-class technologies, or new indications for technologies currently included in the class, will need to be evaluated by WHO PQ and WHO/GTB, respectively.

The publication *WHO operational handbook on tuberculosis. Module 3: Diagnosis* describes the tests included in this class.

Recommendations

1. **For adults and adolescents with signs or symptoms of TB or who screened positive for pulmonary TB, low-complexity automated NAATs should be used on respiratory samples as initial diagnostic tests for TB rather than smear microscopy or culture.**

   (Strong recommendation, high certainty of evidence)

Remarks

- For adults, respiratory samples include sputum (expectorated or induced), tracheal aspirate or bronchoalveolar lavage (BAL).
- The term "person screened positive" refers to a person in whom a screening test has yielded a positive result.[5]
- Children and specifically children living with HIV are discussed in the section on the concurrent use of initial TB diagnostic tests in children.
- Adults and adolescents living with HIV are discussed in the section on the concurrent use of initial TB diagnostic tests in people living with HIV.
- The products for which eligible data met the class-based performance criteria for LC-aNAATs for this recommendation were Xpert MTB/RIF Ultra (Cepheid, Sunnyvale, United States of America [USA]) and Truenat MTB Plus (Molbio, Goa, India). Data on Truenat MTB Plus and MTB-RIF Dx were more limited than those for Xpert Ultra.

2. **For people with bacteriologically confirmed TB, low-complexity automated NAATs should be used on respiratory samples as initial tests for detection of resistance to rifampicin rather than culture-based DST.**

   (Strong recommendation, high certainty of evidence)

Remarks

- This recommendation applies to all people living with HIV.
- The recommendation was extrapolated to children based on the generalization of data from adults and limited data from children. For children, respiratory samples include sputum, BAL, induced sputum, nasopharyngeal aspirate and gastric aspirate.
- The recommendation was extrapolated to people with extrapulmonary TB based on the generalization of data from adults with pulmonary TB.
- The products for which eligible data met the class-based performance criteria for LC-aNAATs for this recommendation were Xpert MTB/RIF Ultra (Cepheid, Sunnyvale, United States of America [USA]) and Truenat MTB-RIF Dx (Molbio, Goa, India). Data on MTB-RIF Dx were more limited than those for Xpert Ultra.

[5] Having a positive result of a test, examination or other procedure used to distinguish people with a high likelihood of having TB disease from people who are highly unlikely to have TB. At present, the following tests are WHO-recommended as the screening tests: chest radiography (chest X-ray; CXR) with or without computer-aided detection (CAD), C-reactive protein (CRP) in people living with HIV, and molecular WHO-recommended rapid diagnostic test for TB (mWRD) (https://www.who.int/publications/i/item/9789240022676).

3. **For people with signs and symptoms of TB meningitis, low-complexity automated NAATs on cerebral spinal fluid should be used for the initial diagnosis of TB meningitis rather than smear microscopy or culture.**

(Strong recommendation, high certainty of evidence)

Remarks

- This recommendation applies to all people with signs and symptoms of TB meningitis, including people living with HIV and children.
- Where possible, culture may be performed in addition to automated NAAT testing, to maximize the opportunity for diagnosis and detection of DR-TB.
- The product for which eligible data met the class-based performance criteria for LC-aNAATs for this recommendation was Xpert MTB/RIF Ultra (Cepheid, Sunnyvale, United States of America [USA]). Data on Truenat MTB Plus and MTB-RIF Dx were limited and variable and thus were insufficient for evaluation.

4. **For people with signs and symptoms of extrapulmonary TB, low-complexity automated NAATs on lymph node tissue aspirate, pleural tissue, pleural fluid, synovial fluid, peritoneal fluid or pericardial fluid should be used for the initial diagnosis of TB rather than smear microscopy or culture.**

(Strong recommendation, high certainty of evidence)

Remarks

- This recommendation applies to all people with signs and symptoms of the respective form of extrapulmonary TB, including people living with HIV and children.
- Data on the performance of LC-aNAATs when used with urine and blood samples were limited or inconsistent.
- Where possible, culture may be performed in addition to automated NAAT testing, to maximize the opportunity for diagnosis and detection of DR-TB.
- The product for which eligible data met the class-based performance criteria for LC-aNAATs for this recommendation was Xpert MTB/RIF Ultra (Cepheid, Sunnyvale, United States of America [USA]). Data on Truenat MTB Plus and MTB-RIF Dx were limited and variable and thus were insufficient for evaluation.

Justification and evidence

WHO/GTB initiated an update of the previous guidelines and commissioned a systematic review on the use of LC-aNAATs (Xpert Ultra, Truenat MTB Plus and Truenat MTB-RIF Dx assays) for the diagnosis of TB and resistance to rifampicin in people with signs and symptoms of TB, or who screened positive for TB. The data on the performance of LC-aNAATs alone in these populations, compared with smear microscopy and culture, are presented in Web Annexes B.1–B.4. Recommendations on concurrent testing for children and people living with HIV supersede the use of LC-aNAATs alone in these populations (see **Section 2.3** of this document).

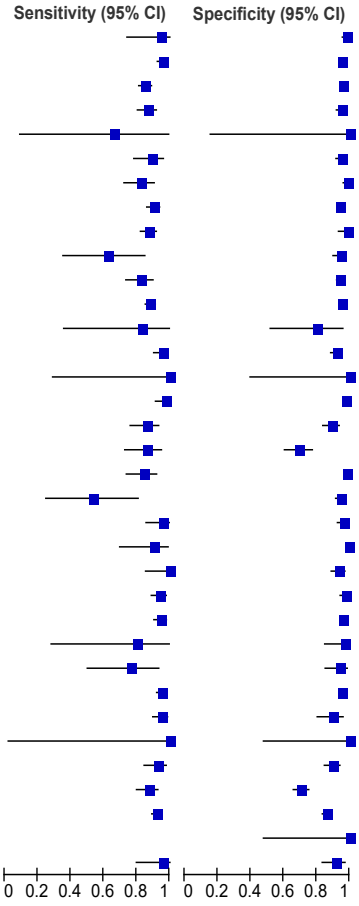
Detection of pulmonary TB

Should LC-aNAATs on respiratory samples be used to diagnose pulmonary TB in adults and adolescents with signs and symptoms or who screened positive for pulmonary TB, against a microbiological reference standard?

Thirty-five studies (14 845 participants) assessed diagnostic accuracy using sputum specimens and comparing with a microbiological reference standard (MRS); however, one of those studies had no people with TB (Zar 2019) and so sensitivity was not estimable. The sensitivities in the remaining 34 studies (14 840 participants) included in the meta-analysis were between 54% and 100%, and the specificities were between 71% and 100% (**Fig. 2.1.1**). The summary sensitivity was 90.4% (95% confidence interval [CI]: 88.0–92.4), and the summary specificity was 94.9% (95% CI: 93.0–96.3). The certainty of evidence for sensitivity and specificity was graded as "high". For more details, see **Web Annex B.1**.

Fig. 2.1.1. Forest plot of LC-aNAAT sensitivity and specificity for detection of pulmonary TB in sputum samples and MRS[a]

Study	TP	FP	FN	TN	Brand	Sensitivity (95% CI)	Specificity (95% CI)
Jose 2024	18	3	1	181	Truenat MTB Plus	0.95 [0.74, 1.00]	0.98 [0.95, 1.00]
Ngangue 2022	224	35	10	676	Truenat MTB Plus	0.96 [0.92, 0.98]	0.95 [0.93, 0.97]
Penn-Nicholson 2021	295	51	51	1144	Truenat MTB Plus	0.85 [0.81, 0.89]	0.96 [0.94, 0.97]
Theron 2024	131	11	20	222	Truenat MTB Plus	0.87 [0.80, 0.92]	0.95 [0.92, 0.98]
Aguilera-Alonso 2022	2	0	1	2	Xpert Ultra	0.67 [0.09, 0.99]	1.00 [0.16, 1.00]
Berhanu 2018	50	8	6	173	Xpert Ultra	0.89 [0.78, 0.96]	0.96 [0.91, 0.98]
Boyles 2020	62	1	13	131	Xpert Ultra	0.83 [0.72, 0.90]	0.99 [0.96, 1.00]
Calderwood 2023	223	37	24	612	Xpert Ultra	0.90 [0.86, 0.94]	0.94 [0.92, 0.96]
Chakravorty 2017	175	1	25	76	Xpert Ultra	0.88 [0.82, 0.92]	0.99 [0.93, 1.00]
Chien 2020	10	8	6	141	Xpert Ultra	0.63 [0.35, 0.85]	0.95 [0.90, 0.98]
Chilukutu 2022	76	39	16	608	Xpert Ultra	0.83 [0.73, 0.90]	0.94 [0.92, 0.96]
Dorman 2018	408	43	54	934	Xpert Ultra	0.88 [0.85, 0.91]	0.96 [0.94, 0.97]
Jaganath 2021	5	3	1	12	Xpert Ultra	0.83 [0.36, 1.00]	0.80 [0.52, 0.96]
Li 2023	96	21	4	247	Xpert Ultra	0.96 [0.90, 0.99]	0.92 [0.88, 0.95]
Liu 2021	3	0	0	4	Xpert Ultra	1.00 [0.29, 1.00]	1.00 [0.40, 1.00]
MacLean 2023	77	13	2	498	Xpert Ultra	0.97 [0.91, 1.00]	0.97 [0.96, 0.99]
Mishra 2020a	62	18	10	149	Xpert Ultra	0.86 [0.76, 0.93]	0.89 [0.84, 0.93]
Mishra 2020b	38	38	6	86	Xpert Ultra	0.86 [0.73, 0.95]	0.69 [0.60, 0.77]
Moodley 2022	59	13	11	706	Xpert Ultra	0.84 [0.74, 0.92]	0.98 [0.97, 0.99]
Mukoka 2023	7	17	6	289	Xpert Ultra	0.54 [0.25, 0.81]	0.94 [0.91, 0.97]
Opota 2019	45	5	2	144	Xpert Ultra	0.96 [0.85, 0.99]	0.97 [0.92, 0.99]
Park 2019	19	1	2	260	Xpert Ultra	0.90 [0.70, 0.99]	1.00 [0.98, 1.00]
Pereira 2020	23	10	0	147	Xpert Ultra	1.00 [0.85, 1.00]	0.94 [0.89, 0.97]
Piersimoni 2019	116	3	7	140	Xpert Ultra	0.94 [0.89, 0.98]	0.98 [0.94, 1.00]
Saavedra 2021	149	52	8	1210	Xpert Ultra	0.95 [0.90, 0.98]	0.96 [0.95, 0.97]
Sabi 2018	4	1	1	33	Xpert Ultra	0.80 [0.28, 0.99]	0.97 [0.85, 1.00]
Sabi 2022	13	4	4	61	Xpert Ultra	0.76 [0.50, 0.93]	0.94 [0.85, 0.98]
Sava 2023	270	20	14	426	Xpert Ultra	0.95 [0.92, 0.97]	0.96 [0.93, 0.97]
Sessolo 2023	103	7	5	62	Xpert Ultra	0.95 [0.90, 0.98]	0.90 [0.80, 0.96]
Ssengooba 2020	1	0	0	5	Xpert Ultra	1.00 [0.03, 1.00]	1.00 [0.48, 1.00]
Ssengooba 2024	66	18	5	160	Xpert Ultra	0.93 [0.84, 0.98]	0.90 [0.84, 0.94]
Wang 2019	102	112	15	269	Xpert Ultra	0.87 [0.80, 0.93]	0.71 [0.66, 0.75]
Wang 2022	322	102	27	627	Xpert Ultra	0.92 [0.89, 0.95]	0.86 [0.83, 0.88]
Zar 2019	0	0	0	5	Xpert Ultra	Not estimable	1.00 [0.48, 1.00]
Zhang 2021	24	6	1	68	Xpert Ultra	0.96 [0.80, 1.00]	0.92 [0.83, 0.97]



CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by assay and author.

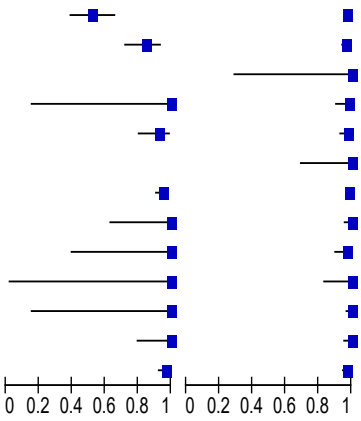
Detection of rifampicin resistance

Should LC-aNAATs on respiratory samples be used to diagnose rifampicin resistance in adults and adolescents with signs and symptoms or who screened positive for pulmonary TB, against an MRS?

Of the 13 studies (2553 participants) that evaluated sputum specimens, sensitivity for detecting rifampicin resistance was not estimable for two studies (**Fig. 2.1.2**). The sensitivities in the remaining 11 studies (2540 participants) included in the meta-analysis were between 53% and 100%, and the specificities were between 97% and 100%. The summary sensitivity was 95.1% (95% CI: 83.1–98.7), and the summary specificity was 98.1% (95% CI: 97.0–98.7). Only two of the 11 included studies assessed Truenat MTB-RIF Dx; one of them, a study from a single country, had a sensitivity outside of confidence interval limits (53%). Nevertheless, overall, the certainty of evidence for both sensitivity and specificity was considered high.

Fig. 2.1.2. Forest plot of LC-aNAAT sensitivity and specificity for detection of rifampicin resistance in respiratory specimens and MRS[a]

Study	TP	FP	FN	TN	Brand	Sensitivity (95% CI)	Specificity (95% CI)
Gomathi 2020	31	17	28	558	Truenat MTB-RIF Dx	0.53 [0.39, 0.66]	0.97 [0.95, 0.98]
Penn-Nicholson 2021	44	9	8	271	Truenat MTB-RIF Dx	0.85 [0.72, 0.93]	0.97 [0.94, 0.99]
Aguilera-Alonso 2022	0	0	0	3	Xpert Ultra	Not estimable	1.00 [0.29, 1.00]
Berhanu 2018	2	1	0	54	Xpert Ultra	1.00 [0.16, 1.00]	0.98 [0.90, 1.00]
Chakravorty 2017	38	2	3	96	Xpert Ultra	0.93 [0.80, 0.98]	0.98 [0.93, 1.00]
Chien 2020	0	0	0	10	Xpert Ultra	Not estimable	1.00 [0.69, 1.00]
Dorman 2018	166	6	9	370	Xpert Ultra	0.95 [0.90, 0.98]	0.98 [0.97, 0.99]
Li 2023	8	0	0	79	Xpert Ultra	1.00 [0.63, 1.00]	1.00 [0.95, 1.00]
MacLean 2023	4	2	0	66	Xpert Ultra	1.00 [0.40, 1.00]	0.97 [0.90, 1.00]
Park 2019	1	0	0	20	Xpert Ultra	1.00 [0.03, 1.00]	1.00 [0.83, 1.00]
Piersimoni 2019	2	0	0	105	Xpert Ultra	1.00 [0.16, 1.00]	1.00 [0.97, 1.00]
Wang 2019	16	0	0	74	Xpert Ultra	1.00 [0.79, 1.00]	1.00 [0.95, 1.00]
Wang 2022	123	10	4	313	Xpert Ultra	0.97 [0.92, 0.99]	0.97 [0.94, 0.99]



CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TN: true negative; TP: true positive.

[a] Studies are sorted by assay and author.

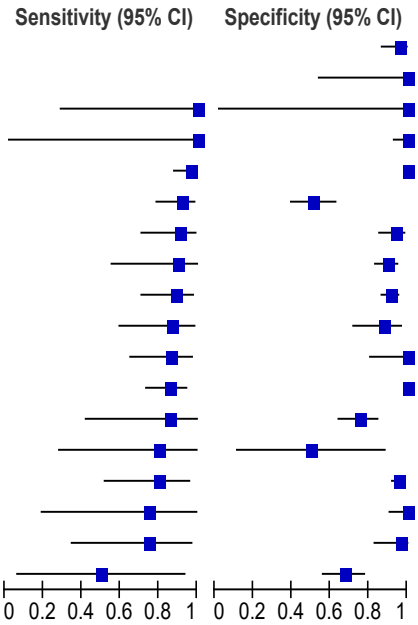
Detection of TB meningitis

Should LC-aNAATs on cerebrospinal fluid (CSF) be used to diagnose TB meningitis in adults with signs and symptoms of TB meningitis, against an MRS?

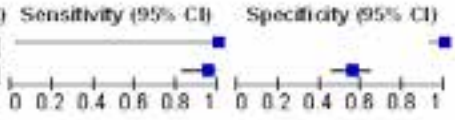
LC-aNAAT summary sensitivity and specificity were 88.2% (95% CI: 83.7–91.6) and 96.0% (95% CI: 86.8–98.9), respectively, based on 16 Xpert Ultra studies (1684 participants); the certainty of evidence was high for sensitivity and moderate for specificity (**Fig. 2.1.3**). Only data on Xpert Ultra were included in the evaluation to answer this population, intervention, comparator and outcome (PICO) question. Of note, trace results from Xpert Ultra were considered positive and formed a significant proportion of positive results (16–63%). Data on Truenat were limited and variable and thus were not included. For more details, see **Web Annex B.3**.

Fig. 2.1.3. Forest plot of LC-aNAAT sensitivity and specificity for detection of TB meningitis in cerebrospinal fluid and MRS[a]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mekkaoui 2021	0	2	0	49	Not estimable	0.96 [0.87, 1.00]
Osei 2019	0	0	0	6	Not estimable	1.00 [0.54, 1.00]
Perez-Risco 2018	3	0	0	1	1.00 [0.29, 1.00]	1.00 [0.03, 1.00]
Spencer-Gomes 2021	1	0	0	49	1.00 [0.03, 1.00]	1.00 [0.93, 1.00]
Sharma 2020	54	0	2	188	0.96 [0.88, 1.00]	1.00 [0.98, 1.00]
Sharma 2021	35	38	3	40	0.92 [0.79, 0.98]	0.51 [0.40, 0.63]
Donovan 2020	20	4	2	62	0.91 [0.71, 0.99]	0.94 [0.85, 0.98]
Bahr 2017	9	12	1	107	0.90 [0.55, 1.00]	0.90 [0.83, 0.95]
Cresswell 2020	24	15	3	162	0.89 [0.71, 0.98]	0.92 [0.86, 0.95]
Quinn 2021	13	4	2	29	0.87 [0.60, 0.98]	0.88 [0.72, 0.97]
Wang 2019	19	0	3	17	0.86 [0.65, 0.97]	1.00 [0.80, 1.00]
Yadav 2023	43	0	7	250	0.86 [0.73, 0.94]	1.00 [0.99, 1.00]
Huang 2021	6	19	1	58	0.86 [0.42, 1.00]	0.75 [0.64, 0.84]
Chin 2019	4	3	1	3	0.80 [0.28, 0.99]	0.50 [0.12, 0.88]
Penata 2021	12	8	3	181	0.80 [0.52, 0.96]	0.96 [0.92, 0.98]
Ninan 2022	3	0	1	37	0.75 [0.19, 0.99]	1.00 [0.91, 1.00]
Slail 2023	6	1	2	29	0.75 [0.35, 0.97]	0.97 [0.83, 1.00]
Shao 2020	2	26	2	54	0.50 [0.07, 0.93]	0.68 [0.56, 0.78]



Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Jose 2024	1	0	0	45	1.00 [0.03, 1.00]	1.00 [0.92, 1.00]
Sharma 2021	36	49	2	61	0.95 [0.82, 0.99]	0.55 [0.46, 0.65]



CI: confidence interval; CSF: cerebrospinal fluid; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

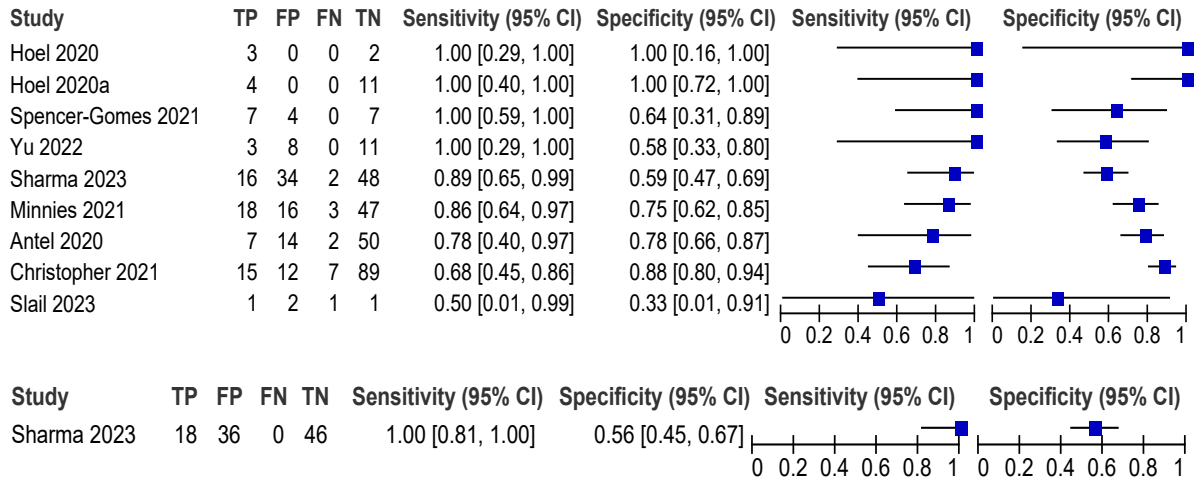
[a] Studies are sorted by decreasing sensitivity.

Detection of extrapulmonary TB

Should LC-aNAATs on lymph node fluid be used to diagnose lymph node TB in adults and adolescents with signs and symptoms of lymph node TB, against an MRS?

LC-aNAAT summary sensitivity and specificity from nine Xpert Ultra studies (445 participants) to diagnose lymph node TB in lymph node fluid in adults and adolescents with signs and symptoms of lymph node TB (**Fig. 2.1.4**) were 85.3% (95% CI: 73.4–92.4) and 74.1% (95% CI: 63.5–82.5), respectively. The certainty of evidence was low for sensitivity and very low for specificity. Only data on Xpert Ultra were included in the evaluation to answer this PICO question. The diagnostic accuracy of LC-aNAATs against a composite reference standard (CRS) that comprised the MRS plus patients who received clinical diagnoses (but were bacteriologically unconfirmed) was also considered. The use of the CRS markedly increased specificity to 97.4% (95% CI: 82.2–99.7) but decreased sensitivity to 71.3% (95% CI: 64.3–77.4), highlighting the known challenges with culture-based confirmation of TB with this sample type (see **Web Annex B.3**). Data on Truenat were limited and thus were not included.

Fig. 2.1.4. LC-aNAAT sensitivity and specificity for detection of lymph node TB in lymph node aspirate and MRS[a]



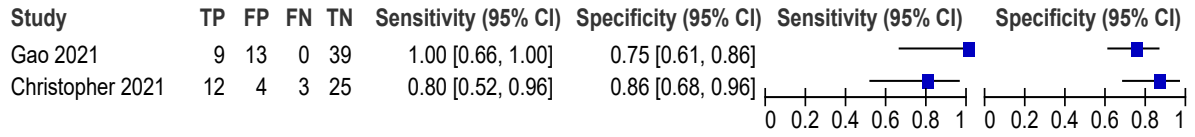
CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; LN: lymph node; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by decreasing sensitivity.

Should LC-aNAATs on pleural tissue be used to diagnose pleural TB in adults and adolescents with signs and symptoms of pleural TB, against an MRS?

From two Xpert Ultra studies (105 participants), LC-aNAAT sensitivities were 80% and 100%, and specificities were 75% and 86% (**Fig. 2.1.5**); the certainty of evidence was low for sensitivity and very low for specificity. Only data on Xpert Ultra were included in the evaluation to answer this PICO question, as data on Truenat were not available. Given known challenges with culture-based confirmation of TB using this sample, the data using the CRS were also considered. The use of the CRS increased specificity of the LC-aNAAT on pleural tissue to 94–97%, but it decreased sensitivity to 54–81%[6] (see **Web Annex B.3**).

Fig. 2.1.5. LC-aNAAT sensitivity and specificity for detection of pleural TB in pleural tissue and MRS[a]



CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

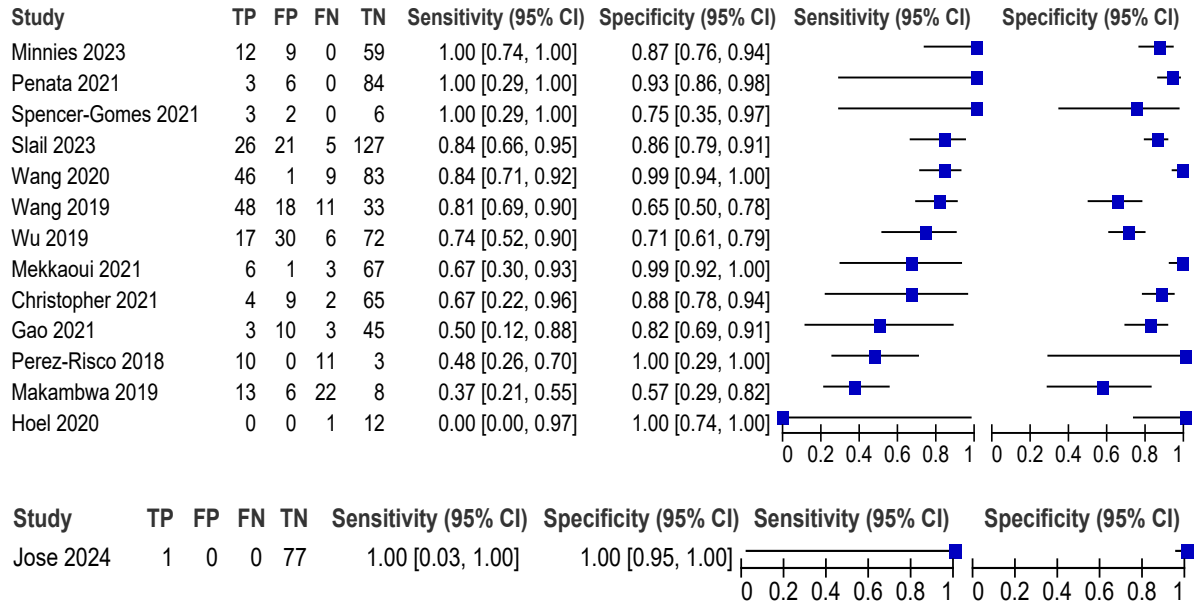
[a] Studies are sorted by decreasing sensitivity.

[6] Data were not pooled due to the limited number of studies.

Should LC-aNAATs on pleural fluid be used to diagnose pleural TB in adults and adolescents with signs and symptoms of pleural TB, against an MRS?

LC-aNAAT summary sensitivity and specificity were 74.0% (95% CI: 60.8–83.9) and 88.1% (95% CI: 78.8–93.6), respectively, from 13 Xpert Ultra studies (1041 participants) (**Fig. 2.1.6**). The certainty of evidence was low for sensitivity and very low for specificity. Only one study (Jose 2024) provided accuracy estimates for pleural fluid for Truenat MTB Plus (88 participants), with sensitivity of 100% (95% CI: 0.03–100) and specificity of 100% (95% CI: 0.95–100). Similar to lymph node fluid and pleural tissue, the data using the CRS were also considered for this sample type. The use of the CRS increased specificity of LC-aNAATs on pleural fluid to 99.2% (95% CI: 95.2%–99.9%) but decreased sensitivity to 71.3% (95% CI: 64.3%–77.4%) (see **Web Annex B.3**). Only data on Xpert Ultra were included in the evaluation to answer this PICO question. Data on Truenat were limited.

Fig. 2.1.6. LC-aNAAT sensitivity and specificity for detection of pleural TB in pleural fluid and MRS[a]



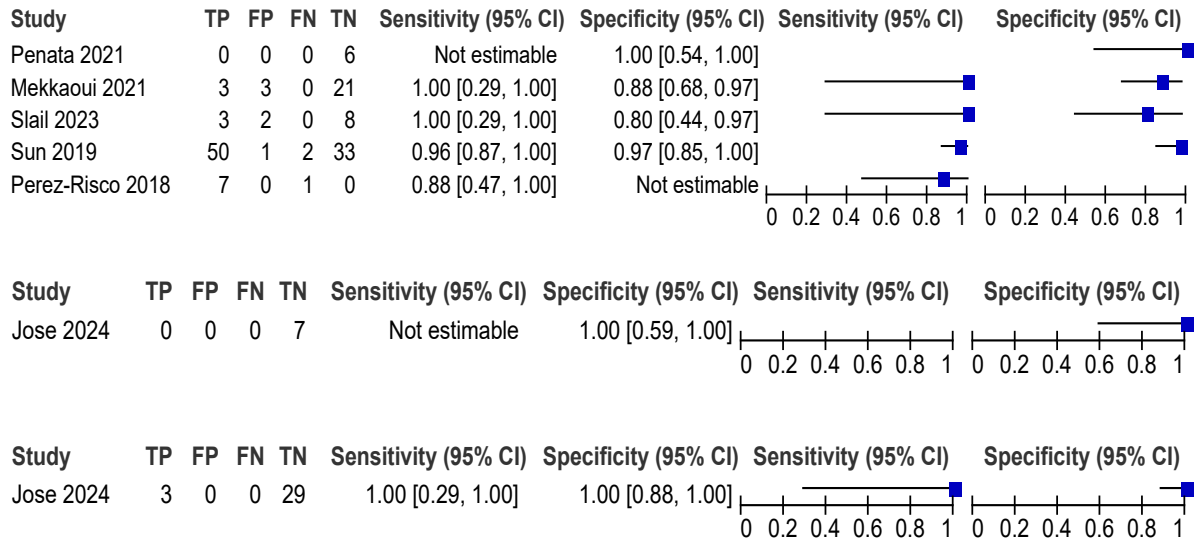
CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by decreasing sensitivity.

Should LC-aNAATs on synovial fluid be used to diagnose bone or joint TB in adults and adolescents with signs and symptoms of bone or joint TB, against an MRS?

LC-aNAAT summary sensitivity and specificity were 96.6% (95% CI: 87.2–99.1) and 91.1% (95% CI: 80.8–96.2), respectively, from three Xpert Ultra studies (126 participants) (**Fig. 2.1.7**); the certainty of evidence was low. Similar to other extrapulmonary TB sample types, the data using the CRS were also considered. The use of the CRS increased specificity of the LC-aNAAT on synovial fluid to 97.0% (95% CI: 85.0–100.0), whereas the impact on sensitivity was minimal (96%) and largely involved tightening of the confidence interval (95% CI: 91–99%). Only data on Xpert Ultra were included in the evaluation to answer this PICO question. Data on Truenat were limited.

Fig. 2.1.7. LC-aNAAT sensitivity and specificity for detection of bone or joint TB in synovial fluid or tissue and MRS[a]



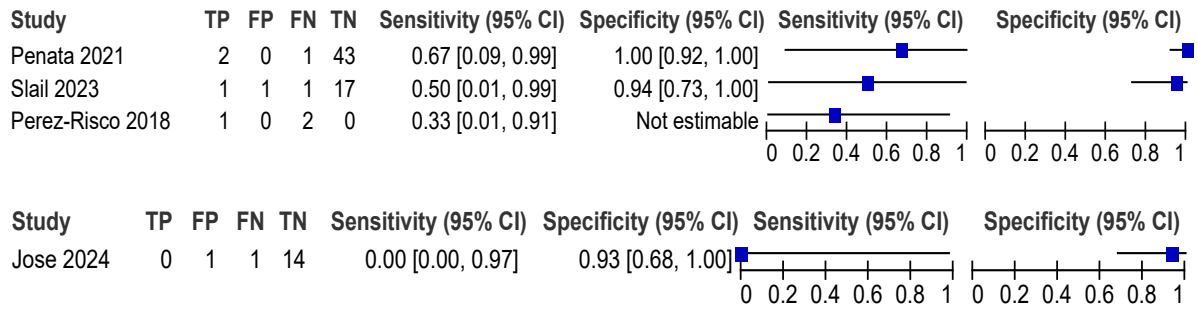
CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by decreasing sensitivity.

Should LC-aNAATs on peritoneal fluid be used to diagnose peritoneal TB in adults and adolescents with signs and symptoms of peritoneal TB, against an MRS?

The sensitivities of the LC-aNAATs ranged from 33% to 67%, and the specificities from 94% to 100%, from three Xpert Ultra studies (69 participants); the certainty of evidence was very low for sensitivity and low for specificity (**Fig. 2.1.8**). Only data on Xpert Ultra were included in the evaluation to answer this PICO question. Data on Truenat were limited.

Fig. 2.1.8. LC-aNAAT sensitivity and specificity for detection of peritoneal TB in peritoneal fluid and MRS[a]



CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

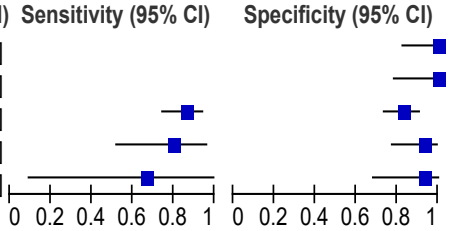
[a] Studies are sorted by decreasing sensitivity.

Should LC-aNAATs on pericardial fluid be used to diagnose pericardial TB in adults and adolescents with signs and symptoms of pericardial TB, against an MRS?

LC-aNAAT summary sensitivity and specificity were 84.0% (95% CI: 73.9–90.7) and 86.6% (95% CI: 79.5–91.5), respectively, from three Xpert Ultra studies (202 participants); certainty of evidence was low for both sensitivity and specificity (**Fig. 2.1.9**).

Fig. 2.1.9. LC-aNAAT sensitivity and specificity for detection of pericardial TB in pericardial fluid and MRS[a]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mekkaoui 2021	0	0	0	19	Not estimable	1.00 [0.82, 1.00]
Penata 2021	0	0	0	15	Not estimable	1.00 [0.78, 1.00]
Minnies 2023	49	14	8	69	0.86 [0.74, 0.94]	0.83 [0.73, 0.90]
Alomatu 2023	12	2	3	27	0.80 [0.52, 0.96]	0.93 [0.77, 0.99]
Slail 2023	2	1	1	14	0.67 [0.09, 0.99]	0.93 [0.68, 1.00]



CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

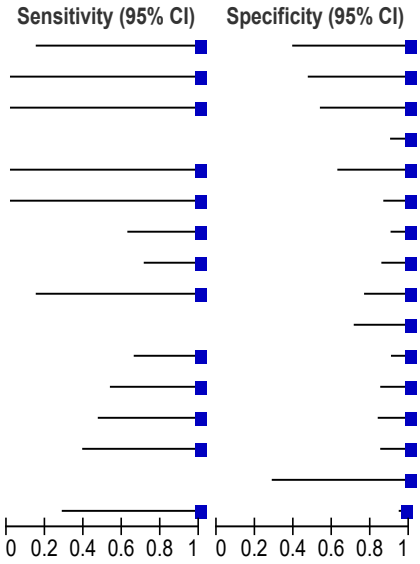
[a] Studies are sorted by decreasing sensitivity.

Should LC-aNAAT on extrapulmonary specimens be used to diagnose rifampicin resistance in adults and adolescents with presumed extrapulmonary TB?

LC-aNAAT summary sensitivity and specificity were 100.0% (95% CI: 93.4–100.0) and 99.4% (95% CI: 92.1–100.0), respectively, from 13 Xpert Ultra studies (446 participants) (**Fig. 2.1.10**); certainty of evidence was high for both sensitivity and specificity.

Fig. 2.1.10. LC-aNAAT sensitivity and specificity for detection of pericardial TB in pericardial fluid and MRS[a]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hoel 2020a	2	0	0	4	1.00 [0.16, 1.00]	1.00 [0.40, 1.00]
Huang 2021	1	0	0	5	1.00 [0.03, 1.00]	1.00 [0.48, 1.00]
Huerga 2023	1	0	0	6	1.00 [0.03, 1.00]	1.00 [0.54, 1.00]
Mekkaoui 2021	0	0	0	36	Not estimable	1.00 [0.90, 1.00]
Minnies 2021	1	0	0	8	1.00 [0.03, 1.00]	1.00 [0.63, 1.00]
Penata 2021	1	0	0	26	1.00 [0.03, 1.00]	1.00 [0.87, 1.00]
Sharma 2020	8	0	0	37	1.00 [0.63, 1.00]	1.00 [0.91, 1.00]
Sharma 2021	11	0	0	24	1.00 [0.72, 1.00]	1.00 [0.86, 1.00]
Sharma 2023	2	0	0	14	1.00 [0.16, 1.00]	1.00 [0.77, 1.00]
Spencer-Gomes 2021	0	0	0	11	Not estimable	1.00 [0.72, 1.00]
Sun 2019	9	0	0	38	1.00 [0.66, 1.00]	1.00 [0.91, 1.00]
Wang 2019	6	0	0	23	1.00 [0.54, 1.00]	1.00 [0.85, 1.00]
Wang 2020	5	0	0	21	1.00 [0.48, 1.00]	1.00 [0.84, 1.00]
Wu 2019	4	0	0	23	1.00 [0.40, 1.00]	1.00 [0.85, 1.00]
Chin 2019	0	0	0	3	Not estimable	1.00 [0.29, 1.00]
Slail 2023	3	3	0	160	1.00 [0.29, 1.00]	0.98 [0.95, 1.00]



CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by sensitivity.

Cost–effectiveness analysis

This section deals with the following additional question:

What are the comparative costs, affordability and cost–effectiveness of implementation of LC-aNAATs?

WHO commissioned a systematic review to identify, evaluate and summarise the evidence on cost, affordability and cost-effectiveness of LC-aNAATs, among other technologies.

A total of 1534 studies were identified in the original search; after removing duplicates, 736 potentially relevant studies were screened. Of these, 107 were assigned for full-text review and were evaluated against the inclusion and exclusion criteria, and 29 studies were included in the final systematic review. Of the 29 included studies, 22 (76%) assessed Xpert MTB/RIF, six (21%) assessed Xpert Ultra, and one study (3%) evaluated Truenat tests (both Truenat MTB Plus and Truenat MTB-RIF Dx). Ten of the studies evaluating Xpert MTB/RIF were cost–effectiveness analyses, and 12 were cost analyses. All the included Xpert Ultra studies and the study evaluating Truenat were cost–effectiveness analyses.

The cost-effectiveness analysis on Xpert was considered because of similarity between two technologies and scarcity of the data on Ultra.

The studies included in the review were diverse, were conducted across various settings and covered all income levels. This broad spectrum of research provided a comprehensive view of the economic evidence on LC-aNAATs, with a focus on adults. Only three cost–effectiveness analyses included children. There were various comparator tests, including smear and culture. Most of the studies included sputum specimens. Six of the 10 cost–effectiveness analyses on Xpert MTB/RIF presented results in natural units (additional people with TB detected), and the other four presented utility outcomes (quality-adjusted life years and disability-adjusted life years [DALYs]).

The findings from the cost–effectiveness analyses showed that Xpert MTB/RIF was generally cost effective across the included studies when compared with smear or culture, except for one study from Thailand, where TB LAMP was the dominant strategy. In contrast, there was more heterogeneity in the methodology used in the cost–effectiveness studies for Xpert Ultra, and the findings showed that, for the Xpert Ultra versus sputum smear microscopy (SSM), an incremental cost–effectiveness ratio (ICER) ranged from US$ 72.72 to US$ 160.23 per DALY averted. In the only study on Truenat, it was found to be cost effective for children in India compared with Xpert MTB/RIF, with an ICER of US$ 94.72 per DALY averted.

In general LC-aNAATs are likely to be cost effective across various settings when compared with SSM and culture.

More details on the economic evaluation of LC-aNAATs are available in **Web Annex B.9.**

User perspective

This section deals with the following question:

Are there implications for user preferences and values, patient equity, accessibility, feasibility and human rights from the implementation of Xpert MTB/RIF and Xpert Ultra?

This review included 49 qualitative studies, of which 17 were identified in the updated search (since 2022). All studies about LC-aNAATs for detection of TB and DR-TB were conducted in high TB burden settings in Africa, Asia and Eastern Europe. Two studies provided user perspectives on Xpert Ultra and the rest on Xpert MTB/RIF or other rapid molecular tests. The studies about Xpert Ultra were conducted in Africa and Eastern Europe and focused on all people with presumptive TB, DR-TB and extrapulmonary TB.

Although standard Xpert MTB/RIF has been superseded by Xpert Ultra and other rapid NAATs, qualitative evidence for the latter NAATs is limited. Whereas LC-NAATs are generally valued for their accuracy, ease of use and potential to reduce time to diagnosis, the most recent generation of NAATs, such as Xpert Ultra, are valued for their greater accuracy in hard-to-diagnose patients, ease of implementation on existing GeneXpert platforms and ease of integration with rapid testing for other diseases. Challenges limiting the realization of these values for more recent NAATs are similar to those with Xpert MTB/RIF – that is, weak infrastructure, fragmented systems, heavy workloads, and limited availability of NAATs and their supplies. We recommend that qualitative studies be conducted to ascertain perspectives on concurrent use of NAATs.

There was high confidence in the evidence contributing to the findings of this review. More details on the qualitative evaluation of LC-aNAATs are available in **Web Annex B.10.**

User preferences and values

Findings from Xpert MTB/RIF studies showed that providers valued its utility in making a diagnosis of drug resistance in people living with HIV, accuracy and resulting confidence in the test, rapid turnaround times, low costs of diagnostic testing for patients, and improved patient–provider relationships. Providers also valued the diversity of sample types that can be analysed by the test. Laboratory personnel valued its ease of use, and they reported increased staff satisfaction compared with sputum microscopy. People with TB valued receiving an accurate diagnosis, avoiding diagnostic delays and having low costs associated with diagnostic testing. Compared with Xpert MTB/RIF, providers valued Xpert Ultra's capacity for improving TB case detection among hard-to-diagnose patients (those with extrapulmonary TB, paediatric TB or coinfection with HIV) and detecting more people with TB.

Compared with Xpert MTB/RIF, providers valued Xpert Ultra for its ease of implementation and integration with testing for other diseases (made possible by its having been built on existing Xpert platforms). Acceptability of Xpert Ultra among providers seemed high, but there was uncertainty about its accuracy, potentially leading to reduced trust and litigation in the event of a false diagnosis.

Patient equity

The limited availability of Xpert Ultra in health facilities and the high costs incurred by patients and health facilities for its use were reported as concerns in terms of equity.

Acceptability

There were challenges to using Xpert MTB/RIF in the health care system. These challenges included underuse of the test and delays in the diagnostic pathway because of poor sample quality, insufficient resources and maintenance of the testing platforms, lack of functional data connectivity systems or record systems, inefficient patient flows, unavailability of updated clinical guidelines, and poor ownership of and accountability for the tests by health facilities. Overreliance on test results, rather than clinical judgement, and a lack of data-driven implementation processes were reported.

Access to the test may be limited owing to lack of sustainable funding, restrictions by donors, poor referral systems, dependence on outreach workers, unavailability of community TB diagnostic facilities and too many eligibility restrictions.

Feasibility

As with Xpert MTB/RIF, implementation of Xpert Ultra could be hindered by infrastructural problems, such as power outages, staff shortages, limited availability of transportation for sputum samples and limited availability of Xpert testing platforms in health facilities.

Implementation considerations

- Diagnostic products in the low-complexity classes of tests should be prequalified by WHO or approved by another regulator before clinical use.
- Diagnostic test manufacturers, laboratory and programme managers, and policy-makers should be educated on the WHO PQ process for TB IVDs (https://extranet.who.int/prequal/).
- Ensuring sufficient volume and specimen quality is important to obtain accurate results.
- Safe waste disposal of used test consumables needs to be planned in advance to minimize environmental risk.
- Trace positive results on respiratory samples may present false-positive results for TB disease (*M. tb.* non-viable but DNA detected) in those that are HIV negative or not at risk for HIV, and those with a prior history of TB and an end of treatment within the last 5 years.
- For tests that do not have integrated rifampicin-resistance detection as an all-in-one test, reflex testing for resistance should be performed at the same time for all TB-positive patients to support universal access to DST for rifampicin, at a minimum, and to reduce the risk of loss to follow-up.
- In settings with a very low prevalence of rifampicin resistance[7], i.e. less than 2%, a positive test result for rifampicin resistance may represent a false positive result, and indicate a need for further testing with an alternative method or, at a minimum, repeat testing.

[7] The 2% prevalence was used as the lowest one in evidence synthesis and analysis to inform GDG meeting. At this prevalence level the number of false-positive results amounted to 19 out of 1000 eligible patients tested and equalized number of true-positive results.

- If rifampicin resistance is detected, further resistance testing for fluoroquinolones and bedaquiline is essential to guide selection of a shorter multidrug-resistant TB or rifampicin-resistant TB (MDR/RR-TB) treatment regimen.
- Use of a higher volume of CSF (≥6 mL) with concentration, where possible, is encouraged to increase the sensitivity of LC-aNAATs.

Monitoring and evaluation

- Track unsuccessful and indeterminate test result rates for currently recommended products and new products to be introduced in this class.
- Monitor the proportion of trace results from paucibacillary samples (e.g. CSF), including those that are culture-positive or culture-negative.
- Undertake surveillance to monitor the frequency of mutations (e.g. *I491F* mutation) outside of a *rpoB* rifampicin-resistance determining region (RRDR) over time.
- Monitor the proportion of people with bacteriologically confirmed TB without a rifampicin-resistance result or further recommended drug susceptibility reflex testing over time.

Research priorities

- Review the field performance of the current technologies used in routine practice (programmatic settings).
- Conduct operational research to ensure that tests are used optimally in terms of both clinical efficiency and cost efficiency in intended settings.
- Evaluate the impact of LC-aNAAT testing on patient-important outcomes (cure, mortality, time to diagnosis and time to start of treatment).
- Evaluate the strengths, weaknesses and cost differences of different LC-aNAAT products to inform country selection.
- Evaluate the different classes of tests, including LC-aNAATs, to determine which classes or testing strategies yield superior diagnostic accuracy, cost–effectiveness and impact on equity and acceptability.
- Evaluate the impact on incremental accuracy and case detection and the cost–effectiveness of alternative sample types that are easier to collect.
- Evaluate the individual product performance with different paediatric and extrapulmonary TB sample types.
- Develop new tools that are rapid, affordable, feasible and acceptable to children and their parents.
- Optimize or develop tests or simple pre-step sample handling procedures for extrapulmonary TB.
- Identify an improved reference standard that accurately defines TB disease in children, paucibacillary specimens, and people who cannot produce sputum, because the sensitivity of all available diagnostics is suboptimal.
- Develop and apply standardized methods for cost–effectiveness and economic studies, to limit variability.

2.1.2 Moderate complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid

Rapid detection of TB and rifampicin resistance is increasingly available as new technologies are developed and adopted by countries. However, what has also emerged is the relatively high burden of isoniazid-resistant, rifampicin-susceptible TB that is often undiagnosed. Globally, isoniazid-resistant, rifampicin-susceptible TB is estimated to occur in 13.1% (95% CI: 9.9–16.9%) of new cases and 17.4% (95% CI: 0.5–54.0%) of previously treated cases (*1*).

A new class of technologies has come to market with the potential to address this gap. Several manufacturers have developed moderate complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid on high throughput platforms for use in laboratories. The tests belonging to this class are faster and less complex to perform than phenotypic culture-based drug susceptibility testing (DST) and line probe assays (LPA). They have the advantage of being largely automated following the sample preparation step. Moderate complexity automated NAATs may be used as an initial test for detection of TB and resistance to both first-line TB drugs simultaneously (rifampicin and isoniazid). They offer the potential for the rapid provision of accurate results (important to patients) and for testing efficiency where high volumes of tests are required daily (important to programmes). Hence, these technologies are suited to areas with a high population density and rapid sample referral systems.

Table 2.1.2.1 Class criteria for MC-aNAATs

Purpose		Detection of TB and resistance to rifampicin and isoniazid
Principle of action		Nucleic acid amplification testing
Complexity	Reagents	Reagents are available within standardized kits and may have temperature requirements for storage. The sample is added automatically or manually to a disposable sealed container for testing.
	Skills	Moderate technical skills (i.e., multiple sample or reagent handling steps, precision pipetting may be required, molecular workflows may be required)
	Pipetting	One or more non-precision or precision pipetting steps required by the procedure.
	Testing Procedure	May require multiple specimen treatment steps before transferring the specimen into a sealed test container for automated processing. Automated or manual DNA extraction Automated real-time PCR Results generation
Type of test result reporting		Automated
Setting of use		Laboratory (special infrastructure may be required)

Recommendation

5. In people with signs and symptoms of pulmonary TB, moderate complexity automated NAATs may be used on respiratory samples for the detection of pulmonary TB, and of rifampicin and isoniazid rsesistance, rather than culture and phenotypic DST.

(Conditional recommendation, moderate certainty of evidence for diagnostic accuracy)

There are several subgroups to be considered for this recommendation:

- The recommendation is based on evidence of diagnostic accuracy in respiratory samples of adults with signs and symptoms of pulmonary TB.
- The recommendation applies to people living with HIV (studies included a varying proportion of such individuals); performance on smear-negative samples was reviewed but was only available for TB detection, not for rifampicin and isoniazid resistance, and data stratified by HIV status were not available.
- The recommendation applies to adolescents and children based on the generalization of data from adults; an increased rate of indeterminate results may be found with paucibacillary TB disease in children.
- The review did not consider extrapolation of the finding for use in people with extrapulmonary TB and testing on non-sputum samples because data on diagnostic accuracy of technologies in the class for non-sputum samples were limited.

Justification and evidence

The WHO Global TB Programme initiated an update of the current guidelines and commissioned a systematic review on the use of moderate complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid in people with signs and symptoms of TB.

Three PICO questions were designed to form the basis for the evidence search, retrieval and analysis:

1. Should moderate complexity automated NAATs be used on respiratory samples in people with signs and symptoms of pulmonary TB for detection of pulmonary TB, as compared with culture?

2. Should moderate complexity automated NAATs be used on respiratory samples in people with signs and symptoms of pulmonary TB for detection of resistance to rifampicin, as compared with culture-based phenotypic DST?

3. Should moderate complexity automated NAATs be used on respiratory samples in people with signs and symptoms of pulmonary TB for detection of resistance to isoniazid, as compared with culture-based phenotypic DST?

A comprehensive search of the following databases (PubMed, Embase, BIOSIS, Web of Science, LILACS and Cochrane) for relevant citations was performed. The search was restricted to the period January 2009 to July 2020. Reference lists from included studies were also searched. No language restriction was applied. Because there were few studies for the selected index tests,

the diagnostic companies were contacted for reports of their internal validation data. Studies were also included from the WHO public call for submission of data. Mycobacterial culture was used as the reference standard for evaluation of *Mtb* detection. Resistance detection was compared with a phenotypic DST reference standard and a composite reference standard (that combines phenotypic and genotypic DST results) in studies where both had been performed.

Bivariate random-effects meta-analyses were performed using Stata software, to obtain pooled sensitivity and specificity estimates with 95% CIs for rifampicin resistance, isoniazid resistance and *Mtb* detection. Where only a limited number of studies were available, descriptive analyses were conducted.

For meta-analysis, studies were first meta-analysed separately for each test. Studies from all the tests were then used to obtain a pooled estimate for all technologies.

To decide whether pooling of all the tests would give meaningful estimates, various criteria for pooling were developed and agreed upon by the GDG panel before they were applied. Data were also evaluated and visualized using head-to-head comparisons of the tests with Xpert® MTB/RIF or any other WHO-recommended test.

Data for all the index platforms were only pooled to answer PICO questions if they met the preconditions given in **Table 2.1.2.2** and fulfilled either Condition 1 or Condition 2.

Table 2.1.2.2 Criteria for pooling studies on moderate complexity automated NAATs

Parameters	Sensitivity	Specificity
Preconditions	n ≥50 culture-positive TB	n ≥100 culture-negative TB
Condition 1 (pool based on clinical grounds)	The pooled estimate of one test lies within ±5% of the overall pooled estimate	The pooled estimate of one test lies within ±2% of the overall pooled estimate
Condition 2 (pool based on statistical grounds)	The pooled estimate for one test lies within 95% CI of the overall pooled estimate AND The pooled estimate for one test lies within ±10% of the overall pooled estimate	The pooled estimate for one test lies within 95% CI of the overall pooled estimate AND The pooled estimate for one test lies within ±5% of the overall pooled estimate

CI: confidence interval; n: number; NAAT: nucleic acid amplification test; TB: tuberculosis.

Data synthesis was structured around the three preset PICO questions, as outlined below. Three web annexes[8] give additional information, as follows:

- details of studies included in the current analysis (**Web Annex 1.3:** *Moderate complexity automated NAATs*;
- a summary of the results and details of the evidence quality assessment (**Web Annex 2.3:** *Moderate complexity automated NAATs*); and

[8] A complete list of web annexes is provided at pp 172–173.

- a summary of the GDG panel judgements (**Web Annex 3.3:** *Moderate complexity automated NAATs*).

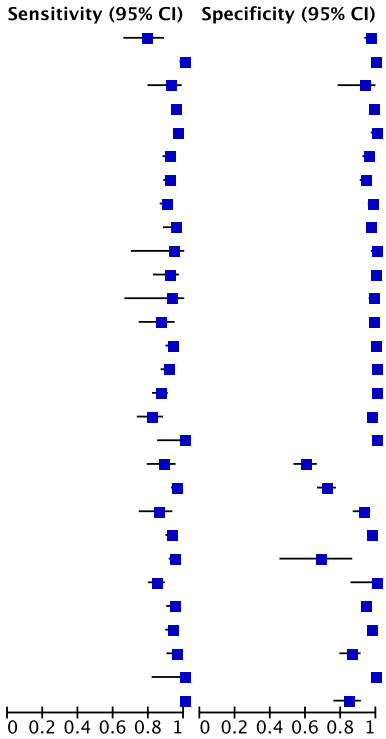
PICO 1: Should moderate complexity automated NAATs be used on respiratory samples in people with signs and symptoms of pulmonary TB for detection of pulmonary TB, as compared with culture?

A total of 29 studies with 13 852 specimens provided data for evaluating TB detection from the five index tests (**Fig. 2.1.2.1**). Of these 29 studies, 12 were conducted on the Abbott RealTi*m*e MTB test, six on FluoroType MTB, four on FluoroType MTBDR, five on BD MAX and two on the cobas MTB test. The reference standard for each of these studies for TB detection was mycobacterial culture.

Of the 29 studies, 16 (55%) had high or unclear risk of bias because they tested specimens before inclusion in the study, used convenience sampling or did not report the method of participant selection. Thus, the evidence was downgraded one level for risk of bias. Overall, the certainty of the evidence was moderate for sensitivity and high for specificity.

Fig. 2.1.2.1 Forest plot of included studies for TB detection with culture as the reference standard

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Berhanu 2018	44	6	12	175	0.79 [0.66, 0.88]	0.97 [0.93, 0.99]
Chen 2015	126	2	0	392	1.00 [0.97, 1.00]	0.99 [0.98, 1.00]
Ciesielczuk 2020	36	2	3	28	0.92 [0.79, 0.98]	0.93 [0.78, 0.99]
Cobas 2020	391	6	21	326	0.95 [0.92, 0.97]	0.98 [0.96, 0.99]
de Vos 2018	322	0	13	105	0.96 [0.93, 0.98]	1.00 [0.97, 1.00]
de Vos 2020	231	9	21	196	0.92 [0.88, 0.95]	0.96 [0.92, 0.98]
de Vos 2020–FT	294	17	27	259	0.92 [0.88, 0.94]	0.94 [0.90, 0.96]
de Vos 2020–FTMTBDR	291	7	33	270	0.90 [0.86, 0.93]	0.97 [0.95, 0.99]
Eigner 2013	77	21	4	559	0.95 [0.88, 0.99]	0.96 [0.95, 0.98]
Genc 2018	15	0	1	108	0.94 [0.70, 1.00]	1.00 [0.97, 1.00]
Haasis 2018	64	1	6	271	0.91 [0.82, 0.97]	1.00 [0.98, 1.00]
Hinic 2017	13	3	1	181	0.93 [0.66, 1.00]	0.98 [0.95, 1.00]
Hofmann–Thiel 2014	45	17	7	902	0.87 [0.74, 0.94]	0.98 [0.97, 0.99]
Hofmann–Thiel 2016	201	2	14	391	0.93 [0.89, 0.96]	0.99 [0.98, 1.00]
Hofmann–Thiel 2016–FT	196	0	19	393	0.91 [0.87, 0.95]	1.00 [0.99, 1.00]
Hofmann–Thiel 2020	213	0	33	257	0.87 [0.82, 0.91]	1.00 [0.99, 1.00]
Howlett 2020	96	8	22	299	0.81 [0.73, 0.88]	0.97 [0.95, 0.99]
Luukinen 2020	22	0	0	131	1.00 [0.85, 1.00]	1.00 [0.97, 1.00]
Obasanya 2017	62	91	8	135	0.89 [0.79, 0.95]	0.60 [0.53, 0.66]
Paradis 2018	274	81	13	207	0.95 [0.92, 0.98]	0.72 [0.66, 0.77]
Scott 2017	53	10	9	121	0.85 [0.74, 0.93]	0.92 [0.86, 0.96]
Shah 2019	262	17	20	593	0.93 [0.89, 0.96]	0.97 [0.96, 0.98]
Shaheed 2020	231	7	13	15	0.95 [0.91, 0.97]	0.68 [0.45, 0.86]
Shaheed 2020–FTMTBDR	209	0	38	23	0.85 [0.80, 0.89]	1.00 [0.85, 1.00]
Tam 2017	152	27	9	422	0.94 [0.90, 0.97]	0.94 [0.91, 0.96]
Tang 2015	198	11	14	348	0.93 [0.89, 0.96]	0.97 [0.95, 0.98]
Vinuesa 2016	106	21	5	125	0.95 [0.90, 0.99]	0.86 [0.79, 0.91]
Vinuesa 2018	18	7	0	1384	1.00 [0.81, 1.00]	0.99 [0.99, 1.00]
Wang 2016	159	15	0	81	1.00 [0.98, 1.00]	0.84 [0.76, 0.91]



CI: confidence interval; FN: false negative; FP: false positive; TB: tuberculosis; TN: true negative; TP: true positive.

The overall sensitivity in these 29 studies ranged from 79% to 100%, and the specificity from 60% to 100%. **The pooled sensitivity was 93.0% (95% CI: 90.9–94.7%) and the pooled specificity was 97.7% (95% CI: 95.6–98.8%).**

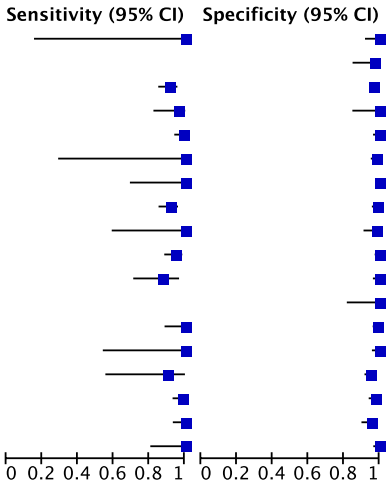
PICO 2: Should moderate complexity automated NAATs be used on respiratory samples in people with signs and symptoms of pulmonary TB for detection of resistance to rifampicin, as compared with culture-based phenotypic DST?

A total of 18 studies with 2874 specimens provided data for resistance testing of rifampicin using moderate complexity automated NAATs (**Fig. 2.1.2.2**). Of these 18 studies, nine were conducted on the Abbott RealTi*m*e RIF/INH test, three on FluoroType MTBDR, four on BD MAX and two on the cobas RIF/INH test. The reference standard for each of these studies for resistance detection was phenotypic DST, using a composite reference standard with both phenotypic DST and sequencing results.

Eight (44%) of the 18 studies had high or unclear risk of bias because they did not report participant selection or tested specimens before inclusion in the study.

Fig. 2.1.2.2 Forest plot of included studies for rifampicin resistance detection with phenotypic DST as the reference standard

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Berhanu 2018	2	0	0	42	1.00 [0.16, 1.00]	1.00 [0.92, 1.00]
Ciesielczuk 2020	0	1	0	33	Not estimable	0.97 [0.85, 1.00]
de Vos 2020	115	14	11	367	0.91 [0.85, 0.96]	0.96 [0.94, 0.98]
Haasis 2018	28	0	1	22	0.97 [0.82, 1.00]	1.00 [0.85, 1.00]
Hillemann 2018	87	0	1	92	0.99 [0.94, 1.00]	1.00 [0.96, 1.00]
Hofmann–Thiel 2020	3	3	0	163	1.00 [0.29, 1.00]	0.98 [0.95, 1.00]
Hofmann–Thiel 2016	10	0	0	164	1.00 [0.69, 1.00]	1.00 [0.98, 1.00]
Hofmann–Thiel 2018	109	3	10	185	0.92 [0.85, 0.96]	0.98 [0.95, 1.00]
Howlett 2020	7	1	0	57	1.00 [0.59, 1.00]	0.98 [0.91, 1.00]
Kostera 2016	91	0	5	120	0.95 [0.88, 0.98]	1.00 [0.97, 1.00]
Kostera 2018	28	0	4	90	0.88 [0.71, 0.96]	1.00 [0.96, 1.00]
Luukinen 2020	0	0	0	18	Not estimable	1.00 [0.81, 1.00]
Paradis 2018	30	3	0	206	1.00 [0.88, 1.00]	0.99 [0.96, 1.00]
Scott 2020	6	0	0	79	1.00 [0.54, 1.00]	1.00 [0.95, 1.00]
Shah 2019	9	11	1	211	0.90 [0.55, 1.00]	0.95 [0.91, 0.98]
Shaheed 2020	75	3	1	132	0.99 [0.93, 1.00]	0.98 [0.94, 1.00]
Shaheed 2020–FTMTBDR	51	5	0	104	1.00 [0.93, 1.00]	0.95 [0.90, 0.98]
Tam 2017	17	0	0	95	1.00 [0.80, 1.00]	1.00 [0.96, 1.00]



CI: confidence interval; DST: drug susceptibility testing; FN: false negative; FP: false positive; TB: tuberculosis; TN: true negative; TP: true positive.

The overall sensitivity for rifampicin resistance in these 18 studies ranged from 88% to 100% and the specificity from 98% to 100%. **The pooled sensitivity was 96.7% (95% CI: 93.1–98.4%) and the pooled specificity was 98.9% (95% CI: 97.5–99.5%).**

In determining rifampicin resistance, the results from genetic sequencing (genotypic DST) were obtained where possible, and a composite reference standard was developed that combined the results from phenotypic and genotypic DST. For rifampicin resistance detection, the diagnostic test accuracy of moderate complexity automated NAATs was similar for phenotypic DST and the composite reference standard.

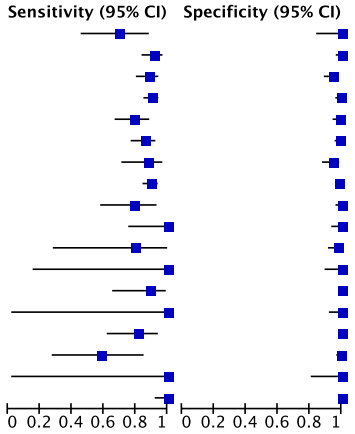
PICO 3: Should moderate complexity automated NAATs be used on respiratory samples in people with signs and symptoms of pulmonary TB for detection of resistance to isoniazid, as compared with culture-based phenotypic DST?

A total of 18 studies with 1758 specimens provided data for resistance testing of isoniazid using moderate complexity automated NAATs (**Fig. 2.1.2.3**). Of these 18 studies, nine were conducted on the Abbott RealTi*m*e RIF/INH test, three on FluoroType MTBDR, four on BD MAX and two on the cobas MTB-RIF/INH test. The reference standard for each of these studies for resistance detection was phenotypic DST, and a composite reference standard with both phenotypic DST and sequencing results.

Eight (44%) of the 18 studies had high or unclear risk of bias, because participant selection was not reported or prior testing was done on the included specimens.

Fig. 2.1.2.3 Forest plot of included studies for isoniazid resistance detection with phenotypic DST as the reference standard

Study	TP	FP	FN	TN	%RIF resistant	Sensitivity (95% CI)	Specificity (95% CI)
Haasis 2018	14	0	6	21	56.85	0.70 [0.46, 0.88]	1.00 [0.84, 1.00]
Hillemann 2018	77	0	7	92	48.89	0.92 [0.84, 0.97]	1.00 [0.96, 1.00]
Kostera 2016	83	7	11	116	44.44	0.88 [0.80, 0.94]	0.94 [0.89, 0.98]
Hofmann-Thiel 2018	163	1	18	129	38.76	0.90 [0.85, 0.94]	0.99 [0.96, 1.00]
Shaheed 2020-FTMTBDR	49	1	13	91	36.02	0.79 [0.67, 0.88]	0.99 [0.94, 1.00]
Shaheed 2020	78	1	13	116	31.88	0.86 [0.77, 0.92]	0.99 [0.95, 1.00]
Kostera 2018	28	5	4	85	26.23	0.88 [0.71, 0.96]	0.94 [0.88, 0.98]
de Vos 2020	163	5	19	323	24.85	0.90 [0.84, 0.94]	0.98 [0.96, 1.00]
Tam 2017	19	0	5	88	15.18	0.79 [0.58, 0.93]	1.00 [0.96, 1.00]
Howlett 2020	13	0	0	53	10.77	1.00 [0.75, 1.00]	1.00 [0.93, 1.00]
Scott 2020	4	2	1	78	7.06	0.80 [0.28, 0.99]	0.97 [0.91, 1.00]
Ciesielczuk 2020	2	0	0	32	5.88	1.00 [0.16, 1.00]	1.00 [0.89, 1.00]
Hofmann-Thiel 2016	16	0	2	160	5.75	0.89 [0.65, 0.99]	1.00 [0.98, 1.00]
Berhanu 2018	1	0	0	43	4.55	1.00 [0.03, 1.00]	1.00 [0.92, 1.00]
Shah 2019	22	0	5	205	4.31	0.81 [0.62, 0.94]	1.00 [0.98, 1.00]
Hofmann-Thiel 2020	7	1	5	152	1.79	0.58 [0.28, 0.85]	0.99 [0.96, 1.00]
Luukinen 2020	1	0	0	17	0.0	1.00 [0.03, 1.00]	1.00 [0.80, 1.00]
Paradis 2018	43	0	0	199		1.00 [0.92, 1.00]	1.00 [0.98, 1.00]



CI: confidence interval; DST: drug susceptibility testing; FN: false negative; FP: false positive; RIF: rifampicin; TB: tuberculosis; TN: true negative; TP: true positive.

The overall sensitivity for isoniazid resistance in these 18 studies ranged from 58% to 100% and the specificity from 94% to 100%. **The pooled sensitivity was 86.4% (95% CI: 82.1–89.8%) and the pooled specificity was 99.8% (95% CI: 98.3–99.8%).**

In determining isoniazid resistance, the results from genetic sequencing (genotypic DST) were obtained where possible, and a composite reference standard was developed that combined the results from phenotypic and genotypic DST. For detecting isoniazid resistance, the diagnostic test accuracy of phenotypic DST was similar to that of the composite reference standard.

Cost–effectiveness analysis

This section answers the following additional question:

What is the comparative cost, affordability and cost–effectiveness of implementation of moderate complexity automated NAATs?

A systematic review was conducted, focusing on economic evaluations of moderate complexity automated NAATs. Four online databases (Embase, Medline, Web of Science and Scopus) were searched for new studies published from 1 January 2010 through 17 September 2020. The citations of all eligible articles, guidelines and reviews were reviewed for additional studies. Experts and test manufacturers were also contacted to identify any additional unpublished studies.

The objective of the review was to summarize current economic evidence and further understand the costs, cost–effectiveness and affordability of moderate complexity automated NAATs.

Several commercially available tests were included as eligible tests in the moderate complexity automated NAATs category; however, no published studies were identified assessing the costs or cost–effectiveness of any of those tests. One unpublished study comparing available data on two technologies from moderate complexity automated NAATs class was identified, and the data from that study are described below.

Unpublished data from FIND was provided through direct communication. This costing-only study used time and motion studies combined with a bottom-up, ingredients-based approach to estimate the unit test cost for the two selected technologies.[9] Time and motion studies were conducted at a reference-level laboratory in South Africa. Several important simplifying assumptions were made that may limit the generalizability of the results; for example, 50% of laboratory operations dedicated to TB, a minimum daily throughput of 24 samples or the equivalent of one BD MAX run (24 tests/run), equipment costs fixed at US$ 100 000 for both platforms, a 5% annual maintenance cost, and the standard 3% discount rate and 10 years expected useful life years.

Additional literature searches conducted to look for economic data using similar platforms from non-TB disease areas identified three additional studies from HIV and hepatitis C virus (HCV) with limited cost data: one (*5*) using Abbott RealTi*m*e HIV and two on HCV (6,7). Data were limited to cost per unit test kit and are not transferrable to test kit costs for the tests being considered in this review.

How large are the resource requirements (costs)?

Available unit test costs for two moderate complexity automated NAATs ranged from US$ 18.52 (US$ 13.79–40.70) and US$ 15.37 (US$ 9.61–37.40), with one study reporting cheaper per-test kit costs and higher operational costs associated with laboratory processing time. Equipment costs were strong drivers of cost variation and will vary across laboratory networks and operations. If equipment can be optimally placed or multiplexed to ensure high testing volume, the per-test cost can be minimized.

In one-way sensitivity analyses, annual testing volumes varied from fewer than 5000 tests/year to more than 25 000 tests/year. Per-test cost was highly sensitive to testing volume when fewer than 5000 tests were conducted per year; however, unit test costs begin to stabilize between 5000 and 10 000 tests/year, and above 10 000 tests/year, unit cost estimate was robust. When equipment can be multiplexed and used at capacity, per-test cost can be minimized.

[9] Data courtesy of H Sohn and W Stevens at FIND (unpublished).

What is the certainty of the evidence of resource requirements (costs)?

Available per-test cost data were unpublished but did include overheads, equipment, building, staff and consumable costs; however, complete quality assessment of the study was not possible. Test cost will vary according to testing volume and laboratory operations. There is limited evidence to assess the important variability across sites, countries and implementation approaches.

Does the cost–effectiveness of the intervention favour the intervention or the comparison?

No studies were identified that assessed cost–effectiveness for any of the moderate complexity automated NAATs, and extrapolation was not appropriate given differences in standard of care, care cascades and associated costs, operational conditions, testing volume and diagnostic accuracy. Implementation considerations (e.g. test placement, laboratory network and ability of the programme to initiate treatment quickly) are all likely to affect unit test cost and cost–effectiveness. Economic modelling is needed across various settings to understand the range of cost–effectiveness profiles of moderate complexity automated NAATs, and how they are likely to vary under different operational criteria.

Additional details on economic evidence synthesis and analysis are provided in **Web Annex B.12:** *Systematic literature review of economic evidence for NAATs to detect TB and DR-TB in adults and children*.

User perspective

This section answers the following questions about **key informants' views and perspectives on the use of moderate complexity automated NAATs:**

- Is there important uncertainty about or variability in how much end-users value the main outcomes?
- What would be the impact on health equity?
- Is the intervention acceptable to key stakeholders?
- Is the intervention feasible to implement?

User perspectives on the value, feasibility, usability and acceptability of diagnostic technologies are important in the implementation of such technologies. If the perspectives of laboratory personnel, clinicians, patients and TB programme personnel are not considered, the technologies risk being inaccessible to and underused by those for whom they are intended.

To address questions related to user perspective, two activities were undertaken:

- A systematic review of evidence on user perspectives and experiences with NAATs for detection of TB and TB drug resistance (moderate and low complexity automated assays, and high complexity hybridization-based assays) was undertaken from July to November 2020.
- A total of 14 semi-structured interviews with clinicians, programme officers, laboratory staff and patient advocates were conducted in India, Moldova and South Africa from October to November 2020.

The findings from these activities are discussed below.

Systematic review

A total of 27 studies were identified that met inclusion criteria, of which 21 were sampled for inclusion in the analysis. All of the sampled studies were published between 2012 and 2020. Of the 21 included studies, 18 were located in high TB burden countries: six in India, four in South Africa, two each in Kenya and Uganda, and one each in Brazil, Cambodia, Myanmar and Viet Nam. One study covered projects in nine countries (Bangladesh, Cambodia, Democratic Republic of the Congo, Kenya, Malawi, Moldova, Mozambique, Nepal and Pakistan). In addition, there was one study located in Eswatini, one in Mongolia and one in Nepal. All studies focused on Xpert MTB/RIF, except for one that focused on Xpert MTB/RIF Ultra (Xpert Ultra).

A summary of the core characteristics of studies included in this review is presented in a study characteristics table in **Web Annex B.13:** *User perspectives on NAATs to detect TB and DR-TB: results from qualitative evidence synthesis: systematic review*.

Interviews

The aim of the interviews was to understand participants' experiences of using the various technologies (i.e. NAATs for detection of TB and TB drug resistance) and their general TB diagnostic experiences. The three countries – India, Moldova and South Africa – were selected based on them being on WHO's list of 30 high MDR-TB burden countries *(1)* and that index tests have been used to some extent in research contexts within these countries. Due to the short time frame, participants were purposively sampled and approached based on convenience through personal contacts and colleagues.

An overview of the participants is given in **Table 2.1.2.3** To mask the identity of study participants they were coded by their country (Moldova [M], India [I] or South Africa [S]), their profession (clinician or medical doctor [M], patient advocate/representative [R], laboratory personnel [L] or programme officers [P]) and a number.

Table 2.1.2.3 Overview of participants for the end-users' interviews

	Moldova	India	South Africa
Clinician or medical doctor	1	1	1
Patient advocate/representative	1	1	1
Laboratory personnel	2[a]	5[a]	2
Programme officers	2[a]	2	1

[a] Participants were interviewed as a group.

Interviews were conducted using Zoom, Skype or phone. Topics discussed included:

- current approach to diagnosing TB, MDR-TB and extensively drug-resistant TB (XDR-TB), including specific challenges;
- experiences with using molecular TB diagnostics and the index tests specifically, including details on steps taken in the diagnostic process;
- experiences with determining eligibility and treatment initiation, and challenges and benefits of using the index tests;

- overall usefulness of the index tests;
- the feasibility of implementing the index tests;
- the potential impact of the index tests on health equity; and
- how the potential impact of the index tests relates to current policy context.

Several important limitations of this approach were noted. Only a few participants were interviewed per country. Owing to the use of Zoom, Skype or phone for interviews, it was not possible to triangulate interview data with other evidence commonly collected through ethnographic approaches (e.g. multiple interviews and informal conversations at the same facility, observations or site visits). In addition, only some of the participants had personal experience with one or all of the index tests, and those participants who did have experience with the tests had used them in research settings rather than for routine practice.

More details on these interviews are given in **Web Annex B.14:** *User perspectives on nucleic acid amplification tests for tuberculosis and tuberculosis drug resistance: Interviews study.*

Findings of the review and interviews

The main findings of the systematic review and interviews are given below. Where information is from the review, a level of confidence in the quality evidence synthesis (QES) is given; where it is from interviews, this is indicated with 'Interviews'.

Is there important uncertainty about or variability in how much end-users value the main outcomes?

- Patients in high burden TB settings value:
  - getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me");
  - avoiding diagnostic delays because they exacerbate existing financial hardships and emotional and physical suffering, and make patients feel guilty for infecting others (especially children);
  - having accessible facilities; and
  - reducing diagnosis-associated costs (e.g. travel, missing work) as important outcomes of the diagnostic.
    QES: moderate confidence
- Moderate complexity automated NAATs meet several preferences and values of clinicians and laboratory staff, in that they:
  - are faster than culture-based phenotypic DST (similar to LPA or cartridge-based tests);
  - have the advantage of being automated (unlike LPA);
  - provide additional clinically relevant drug-resistance information such as high versus low resistance (unlike the current Xpert MTB/RIF cartridge).
    Interviews

What would be the impact on health equity?

- Various factors – for example, lengthy diagnostic delays, underuse of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions – hamper access to prompt and accurate testing and treatment, particularly for vulnerable groups.
  QES: high confidence

- Staff and managers voiced concerns about:
  - sustainability of funding and maintenance;
  - complex conflicts of interest between donors and implementers; and
  - the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics.

  QES: high confidence
- Access to clear and comprehensible information for TB patients on what TB diagnostics are available to them and how to interpret results is a vital component of equity, and lack of such access represents an important barrier for patients.

  Interviews
- New treatment options need to be matched with new diagnostics. It is important to improve access to treatment based on new diagnostics and to improve access to diagnostics for new treatment options.

  Interviews
- The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment:
  - between countries (i.e. between those that can and cannot quickly keep up with the rapidly changing TB diagnostic environment); and
  - within countries (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies).

  Interviews
- The identified challenges with the use of NAATs for detection of TB and DR-TB, and accumulated delays, risk compromising the added value as identified by the users, ultimately leading to underuse. The challenges also hamper access to prompt and accurate testing and treatment, particularly for vulnerable groups.

  QES: high confidence

Is the intervention acceptable to key stakeholders?

- Patients can be reluctant to test for TB or MDR-TB because of:
  - stigma related to MDR-TB or having interrupted treatment in the past;
  - fears of side-effects;
  - failure to recognize symptoms;
  - inability to produce sputum; and
  - cost, distance and travel concerns related to (repeat) clinic visits.

  QES: high confidence
- Health workers can be reluctant to test for TB or MDR-TB because of:
  - TB-associated stigma and consequences for their patients;
  - fear of acquiring TB;
  - fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified;
  - fear of side-effects of drugs in children; and
  - community awareness of disease manifestations in children.

  QES: high confidence

- In relation to the acceptability of moderate complexity automated NAATs:
  - the automation of this class of technologies, which recognizes the high workload of laboratory staff, improves their acceptability;
  - in terms of the physical size of the platform and how it fits into the laboratory space and workflow, a smaller footprint may be more acceptable; and
  - the number of samples run on the system is acceptable provided that the platform is placed within a laboratory that receives a sufficient sample load to run the system.

    Interviews

Is the intervention feasible to implement?

- The feasibility of all diagnostic technologies is challenged if there is an accumulation of diagnostic delays or underuse (or both) at every step in the process, mainly because of health system factors such as:
  - non-adherence to testing algorithms, testing for TB or MDR-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor or delayed sample transport and sample quality, poor or delayed communication of results, delays in scheduling follow-up visits and recalling patients, and inconsistent recording of results;
  - lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners and sputum containers; dusty environment; and delayed or absent local repair option);
  - inefficient or unclear workflows and patient flows (e.g. inefficient organizational processes, poor links between providers, and unclear follow-up mechanisms or information on where patients need to go); and
  - lack of data-driven and inclusive national implementation processes.

    QES: high confidence
- The feasibility of moderate complexity automated NAATs is also challenged by:
  - how or whether the platform fits into the physical space of the laboratory (considering bench size and weight of the platform) and sample workflow;
  - a poorly functioning sample transport system that affects the quality of samples; and
  - the need to ensure that clinicians and laboratory staff have time to communicate effectively regarding diagnostic results if the platform is centralized, while also ensuring that the laboratory location is central enough to receive adequate numbers of samples to make the machine worth running.

    Interviews
- Implementation of new diagnostics must be accompanied by training for clinicians to help them interpret results from new molecular tests and understand how this information is translated into prompt and proper patient management. In the past, with the introduction of Xpert MTB/RIF, this has been a challenge.

  QES: high confidence and *interviews*
- Introduction of new diagnostics must be accompanied by guidelines and algorithms that support clinicians and laboratories in communicating with each other, such that they can discuss discordant results and interpret laboratory results in the context of drug availability, patient history and patient progress on a current drug regimen.

  Interviews

Implementation considerations

Factors to consider when implementing moderate complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid are as follows:

- local epidemiological data on resistance prevalence should guide local testing algorithms, whereas pretest probability is important for the clinical interpretation of test results;
- the cost of a test varies depending on parameters such as the number of samples in a batch and the staff time required; therefore, a local costing exercise should be performed;
- low, moderate and high complexity tests have successive increase in technical competency needs (qualifications and skills) and staff time, which affects planning and budgeting;
- availability and timeliness of local support services and maintenance should be considered when selecting a provider;
- laboratory accreditation and compliance with a robust quality management system (including appropriate quality control) are essential for sustained service excellence and trust;
- training of both laboratory and clinical staff is needed to ensure effective delivery of services and clinical impact;
- use of connectivity solutions for communication of results is encouraged, to improve efficiency of service delivery and reduce time to treatment initiation;
- moderate complexity automated NAATs may already be used programmatically for other diseases – for example, severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), HIV and antimicrobial resistance (AMR) – which could potentially facilitate implementation of TB testing on shared platforms;
- implementation of moderate complexity automated NAATs requires laboratories with the required infrastructure, space and efficient sample referral systems;
- although these are automated tests, well-trained skilled staff are needed to set up assays and complete maintenance requirements; and
- implementation of these tests should be context specific; thus, it should take into account access issues, especially in remote areas, where less centralized WHO-recommended technologies may be more appropriate.

Research priorities

Research priorities for moderate complexity automated NAATs for detection of TB and resistance to rifampicin and isoniazid are as follows:

- diagnostic accuracy in specific patient populations (e.g. children, people living with HIV, and patients with signs and symptoms of extrapulmonary TB) and in non-sputum samples;
- impact of diagnostic technologies on clinical decision-making and outcomes that are important to patients (e.g. cure, mortality, time to diagnosis and time to start treatment) in all patient populations;
- impact of specific mutations on treatment outcomes among people with DR-TB;
- use, integration and optimization of diagnostic technologies in the overall landscape of testing and care, as well as diagnostic pathways and algorithms;
- economic studies evaluating the costs, cost–effectiveness and cost–benefit of different diagnostic technologies;
- qualitative studies evaluating equity, acceptability, feasibility and end-user values of different diagnostic technologies;

- effect of non-actionable results (indeterminate, non-determinate or invalid) on diagnostic accuracy and outcomes that are important to patients;
- operational research on the advantages and disadvantages of individual technologies within the class of moderate complexity automated NAATs;
- effect of moderate complexity automated NAATs in fostering collaboration and integration between disease programmes; and
- the potential utility of detecting *katG* resistance to identify MDR-TB clones that may be missed because they do not have an RRDR mutation (e.g. the Eswatini MDR-TB clone, which has both the *katG S315T* and the non-RRDR *rpoB I491F* mutation).

2.2. Initial diagnostic tests for diagnosis of TB without drug-resistance detection

A new class of low-complexity manual NAATs (LC-mNAATs) has now emerged for alternative molecular solutions that have improved accuracy when compared with smear microscopy and very basic infrastructure, power and equipment requirements (e.g. heat block). LC-mNAATs can be performed at the microscopy level and are currently cheaper than other molecular tests. Collectively, these characteristics are useful for testing in constrained settings. However, like smear microscopy, this class of tests does not incorporate rifampicin-resistance detection and therefore requires reflex testing with a complementary solution for drug-resistance determination.

2.2.1 Low-Complexity manual NAATs for detection of TB

NEW

Diagnostic class description

The features shown in **Table 2.2.1.1** define the class of LC-mNAATs.

Table 2.2.1.1 Class criteria for LC-mNAATs

Purpose		Detection of TB
Principle of action		Nucleic acid amplification testing
Complexity	Reagents	Reagents are enclosed in multiple disposable sealed containers not requiring special storage requirements
	Skills	Basic technical skills (e.g. basic pipetting, precision not critical)
	Pipetting	Multiple pipetting steps (maximum of 10) from processed sample to result generation
	Testing procedure	At least three distinct steps: • Specimen treatment step before transferring the specimen into the disposable sealed container • DNA extraction • PCR amplification • Results visualization
Type of test result reporting		Automated or manual
Setting of use		Basic laboratory (no special infrastructure needed)

DNA: deoxyribonucleic acid; LC-mNAAT: low-complexity manual nucleic acid amplification test; PCR: polymerase chain reaction; TB: tuberculosis.

The only product for which eligible data met the class-based performance criteria for LC-mNAATs is Loopamp MTBC Detection Kit (TB LAMP) (Eiken Chemical, Tokyo, Japan) for pulmonary TB.

Regulatory approval from national regulatory authorities or other relevant bodies is required before implementation of this diagnostic test. Extrapolation to other brand-specific tests cannot be made, and any new in-class technologies or new indications for the technology currently included in the class will need to be evaluated by WHO PQ and WHO/GTB, respectively.

The publication *WHO operational handbook on tuberculosis. Module 3: Diagnosis* describes the tests included in this class.

Recommendations

6. For adults and adolescents with signs or symptoms or who screen positive for pulmonary TB, low-complexity manual NAATs should be used on respiratory samples as initial diagnostic tests for TB rather than smear microscopy or culture.

(Strong recommendation, high certainty of evidence)

Remarks

- This recommendation applies to all people living with HIV, with the caveat of low to moderate certainty of evidence. However, wherever available, concurrent testing with an LC-aNAAT and LF-LAM is recommended for people living with HIV. For more details, see Section 2.3.1.
- This recommendation was extrapolated to children for use with respiratory samples (including induced sputum and gastric aspirate) based on the generalization of data from adults and very limited data for children, acknowledging the difficulties of collecting sputum specimens from this population. However, wherever available, concurrent testing with an LC-aNAAT on a respiratory and stool samples is recommended for children. For more details, see Section 2.3.2.
- Data on the use of the test with paediatric stool samples were very limited, and there were no data on the use of nasopharyngeal aspirates. The recommendation was, therefore, not extrapolated to these sample types.
- No recommendation was made on test use for extrapulmonary TB due to insufficient data.
- As LC-mNAATs do not provide rifampicin-resistance results, all positive diagnostic tests for TB require follow-up and referral for DST for, at a minimum, rifampicin.

Justification and evidence

WHO/GTB initiated an update of the previous guidelines and commissioned a systematic review on the use of LC-mNAATs (TB LAMP) for the diagnosis of TB in people with signs and symptoms of TB, or who screened positive for TB.

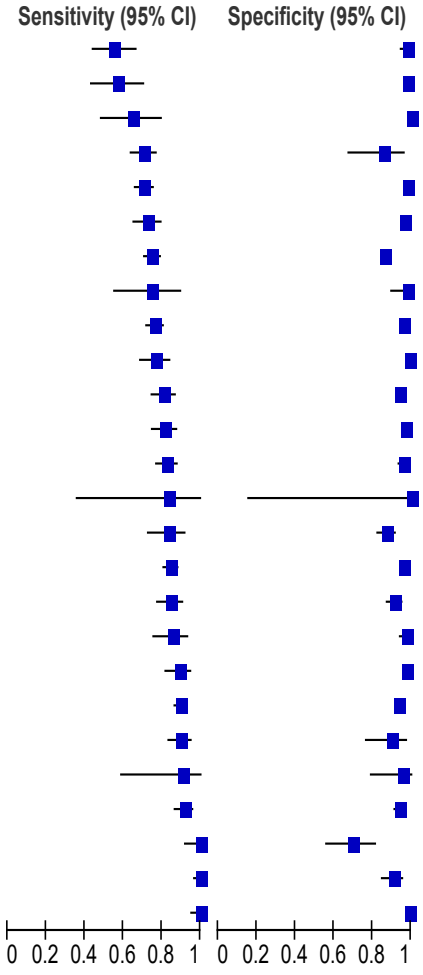
Detection of pulmonary TB

Should LC-mNAATs on respiratory samples be used to diagnose pulmonary TB in adults and adolescents with signs and symptoms or who screened positive for pulmonary TB, against an MRS?

Twenty-six studies (18 297 participants) assessed diagnostic accuracy using sputum specimens and comparing with an MRS. The sensitivities were between 55% and 100%, and the specificities were between 70% and 100% (**Fig. 2.2.1.1**). The summary sensitivity was 84.1% (95% CI: 78.3–88.6), and the summary specificity was 96.1% (95% CI: 94.2–97.4). The certainty of evidence for sensitivity and specificity was high.

Fig. 2.2.1.1 Forest plot of LC-mNAAT sensitivity and specificity for detection of pulmonary TB in respiratory samples and MRS[a]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Nakiyingi 2018	46	3	37	147	0.55 [0.44, 0.66]	0.98 [0.94, 1.00]
Spooner 2022	32	12	24	637	0.57 [0.43, 0.70]	0.98 [0.97, 0.99]
Nliwasa 2016	26	0	14	193	0.65 [0.48, 0.79]	1.00 [0.98, 1.00]
Xin 2023	141	4	59	24	0.70 [0.64, 0.77]	0.86 [0.67, 0.96]
Ou 2014	265	16	110	938	0.71 [0.66, 0.75]	0.98 [0.97, 0.99]
Reddy 2017	119	17	45	514	0.73 [0.65, 0.79]	0.97 [0.95, 0.98]
Ou 2016	322	147	108	942	0.75 [0.71, 0.79]	0.87 [0.84, 0.88]
Gelaw 2017	21	1	7	49	0.75 [0.55, 0.89]	0.98 [0.89, 1.00]
Pham 2018	292	22	91	571	0.76 [0.72, 0.80]	0.96 [0.94, 0.98]
Odeume 2021	99	17	30	2490	0.77 [0.68, 0.84]	0.99 [0.99, 1.00]
Danfack 2024b	141	48	33	736	0.81 [0.74, 0.87]	0.94 [0.92, 0.95]
Cheng 2020	128	9	29	333	0.82 [0.75, 0.87]	0.97 [0.95, 0.99]
Danfack 2018	171	12	36	290	0.83 [0.77, 0.88]	0.96 [0.93, 0.98]
Promsena 2022	5	0	1	2	0.83 [0.36, 1.00]	1.00 [0.16, 1.00]
Kim 2018	56	28	11	193	0.84 [0.73, 0.92]	0.87 [0.82, 0.91]
Gray 2016	331	52	61	1333	0.84 [0.80, 0.88]	0.96 [0.95, 0.97]
Danfack 2023	110	19	20	205	0.85 [0.77, 0.90]	0.92 [0.87, 0.95]
Kaku 2016	60	3	10	136	0.86 [0.75, 0.93]	0.98 [0.94, 1.00]
Wang 2019	91	10	11	389	0.89 [0.82, 0.94]	0.97 [0.95, 0.99]
Ou 2019	376	181	44	2519	0.90 [0.86, 0.92]	0.93 [0.92, 0.94]
Mitarai 2011	107	4	12	36	0.90 [0.83, 0.95]	0.90 [0.76, 0.97]
Yadav 2021	10	1	1	23	0.91 [0.59, 1.00]	0.96 [0.79, 1.00]
N'Guessan 2016	144	18	13	294	0.92 [0.86, 0.96]	0.94 [0.91, 0.97]
Wahid 2020	42	17	0	39	1.00 [0.92, 1.00]	0.70 [0.56, 0.81]
Bojang 2016	97	12	0	119	1.00 [0.96, 1.00]	0.91 [0.85, 0.95]
Yadav 2017	69	3	0	381	1.00 [0.95, 1.00]	0.99 [0.98, 1.00]



CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

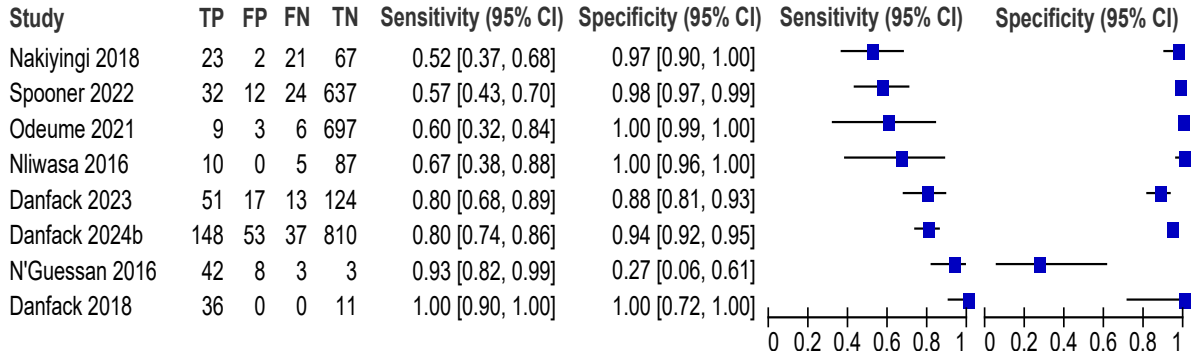
[a] Studies are sorted by increasing sensitivity.

Detection of TB in people living with HIV

Should LC-mNAATs on respiratory samples be used to diagnose pulmonary TB in adult and adolescent living with HIV with signs and symptoms of pulmonary TB, against an MRS?

In the eight studies (2991 participants) included in this meta-analysis, the sensitivities ranged between 52% and 100%, and the specificities between 27% and 100% (**Fig. 2.2.1.2**). The summary sensitivity was 77.1% (95% CI: 60.8–87.9), and the summary specificity was 95.9% (95% CI: 84.9–99.0). The certainty of evidence was low for sensitivity and moderate for specificity.

Fig. 2.2.1.2 Forest plot of LC-mNAAT sensitivity and specificity for detection of pulmonary TB in respiratory samples from people living with HIV and MRS[a]



CI: confidence interval; FN: false negative; FP: false positive; HIV: human immunodeficiency virus; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

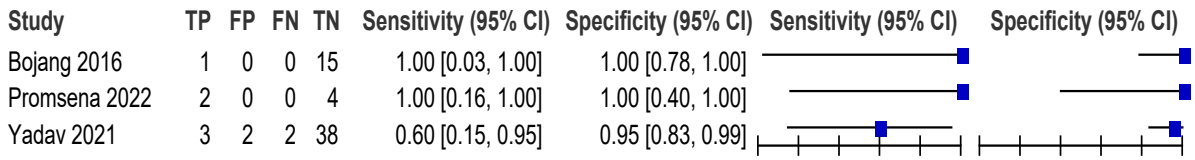
[a] Studies are sorted by increasing sensitivity.

Detection of TB in children

Should LC-mNAATs on respiratory samples be used to diagnose pulmonary TB in children with signs and symptoms of pulmonary TB, against an MRS?

Three studies (62 participants, including eight with pulmonary TB) assessed the accuracy of LC-mNAATs for detecting pulmonary TB using respiratory samples (sputum, BAL and tracheal aspirate) and an MRS (**Fig. 2.2.1.3**). The sensitivities were between 60% and 100%, and the specificities were between 95% and 100%. The certainty of evidence was very low for sensitivity and low for specificity.

Fig. 2.2.1.3 Forest plot of LC-mNAAT sensitivity and specificity for detection of pulmonary TB in respiratory samples and MRS[a]



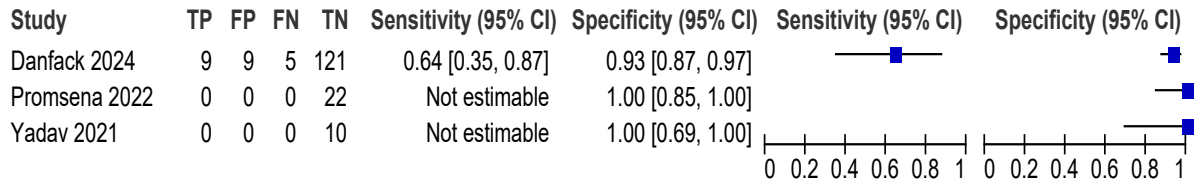
CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by increasing sensitivity.

Should LC-mNAATs on gastric aspirate be used to diagnose pulmonary TB in children with signs and symptoms of pulmonary TB, against an MRS?

Three studies (176 participants, including 14 with pulmonary TB) assessed the accuracy of LC-mNAATs for detecting pulmonary TB using gastric aspirate against a MRS (**Fig. 2.2.1.4**). Sensitivity was not estimable for two studies and was 64% in the third study. The specificities were between 93% and 100%.

Fig. 2.2.1.4 Forest plot of LC-mNAAT sensitivity and specificity for detection of pulmonary TB in gastric aspirate and MRS[a]



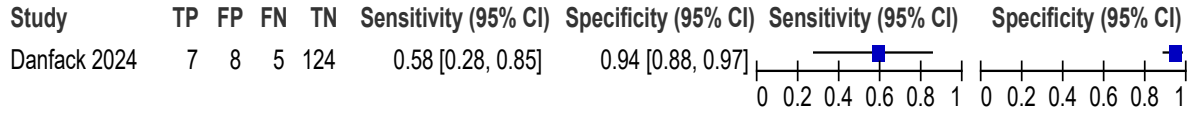
CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by increasing sensitivity.

Should LC-mNAATs on nasopharyngeal aspirate be used to diagnose pulmonary TB in children with signs and symptoms of pulmonary TB, against an MRS?

One study (144 participants including 12 with pulmonary TB) assessed the accuracy of LC-mNAATs for detecting pulmonary TB using nasopharyngeal aspirate against an MRS (**Fig. 2.2.1.5**). The sensitivity was 58% and specificity was 94%. Due to limited data, a recommendation on using LC-mNAATs with nasopharyngeal aspirate for detection of pulmonary TB was not made.

Fig. 2.2.1.5 Forest plot of LC-mNAAT sensitivity and specificity for detection of pulmonary TB in nasopharyngeal aspirate and MRS[a]



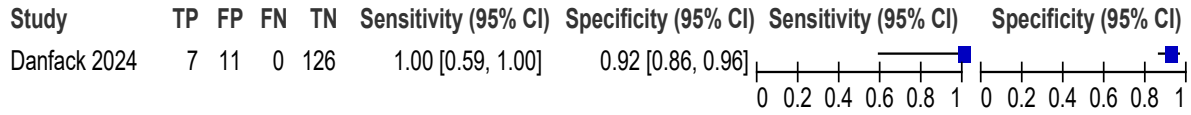
CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted by increasing sensitivity.

Should LC-mNAATs on stool be used to diagnose pulmonary TB in children with signs and symptoms of pulmonary TB, against an MRS?

One study (144 participants, including seven with pulmonary TB) assessed the accuracy of LC-mNAATs for detecting pulmonary TB using stool against a MRS (**Fig. 2.2.1.6**). The sensitivity was 100% and specificity was 92%. The certainty of evidence was very low for sensitivity and moderate for specificity. Due to limited data, a recommendation on using LC-mNAATs with stool for detection of pulmonary TB was not made.

Fig. 2.2.1.6 Forest plot of LC-mNAAT sensitivity and specificity for detection of pulmonary TB in stool and MRS



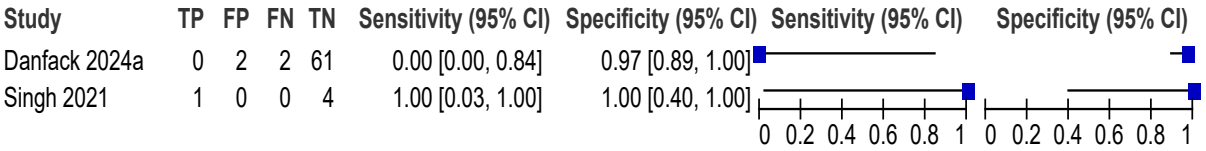
CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Detection of TB meningitis

Should LC-mNAATs on CSF be used to diagnose TB meningitis in adults and adolescents with signs and symptoms of TB meningitis, against an MRS?

Two studies (70 participants, including three with TB meningitis) assessed the accuracy of LC-mNAATs for detecting TB meningitis using CSF and an MRS (**Fig. 2.2.1.7**). Estimated sensitivity and specificity were both 100% in one study, and 0% and 97%, respectively, in the other. The certainty of evidence was very low for sensitivity and low for specificity. Due to limited data, a recommendation on using LC-mNAATs with CSF for detection of TB meningitis was not made.

Fig. 2.2.1.7 Forest plot of LC-mNAAT sensitivity and specificity for detection of TB meningitis in CSF and MRS



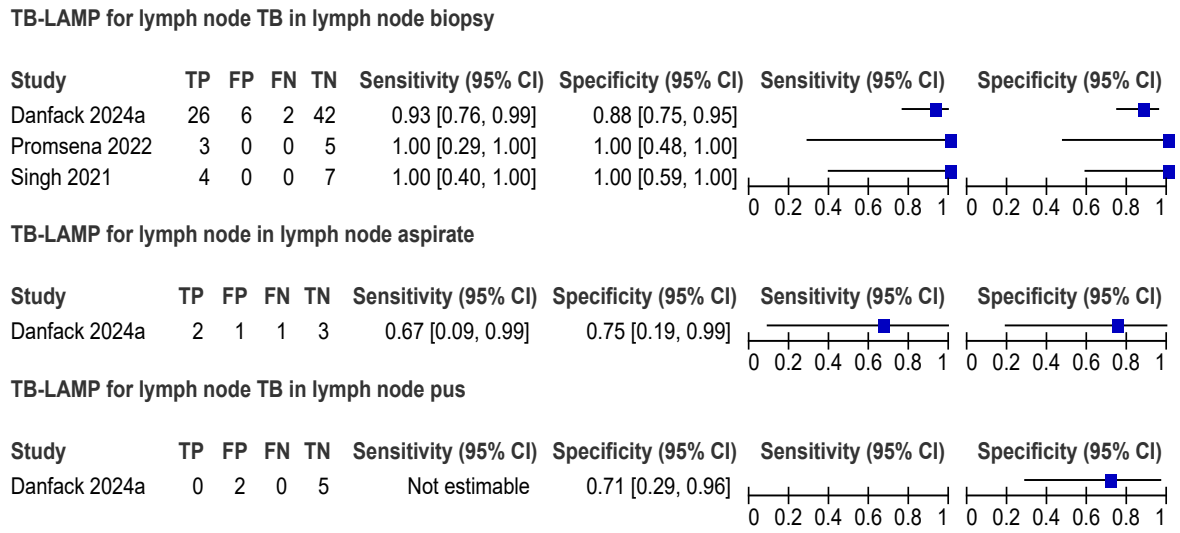
CI: confidence interval; CSF: cerebrospinal fluid; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Detection of extrapulmonary TB

Should LC-mNAATs on lymph node tissue be used to diagnose lymph node TB in adults and adolescents with signs and symptoms of lymph node TB, against an MRS?

Three studies (95 participants, including 35 people with TB) assessed the accuracy of LC-mNAATs for detecting lymph node TB using lymph node tissue from biopsy and an MRS (**Fig. 2.2.1.8**). The estimated sensitivities were between 93% and 100%, and specificities were between 88% and 100%. The summary sensitivity was 94.3% (95% CI: 79.8–98.6), and the summary specificity was 90.0% (95% CI: 79.5–95.4). The certainty of evidence was low for both sensitivity and specificity. Due to limited data, a recommendation on using LC-mNAATs with lymph node tissue for the detection of lymph node TB was not made.

Fig. 2.2.1.8 Forest plot of LC-mNAAT sensitivity and specificity for detection of lymph node TB in lymph node tissue and MRS

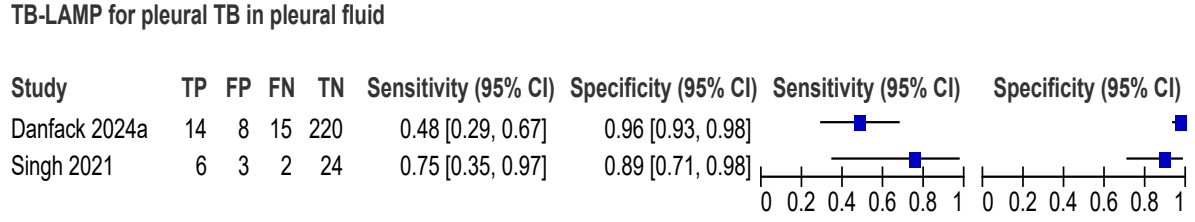


CI: confidence interval; FN; false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; LN: lymph node; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Should LC-mNAATs on pleural fluid be used to diagnose pleural TB in adults and adolescents with signs and symptoms of pleural TB, against an MRS?

Two studies (292 participants, including 37 people with TB) assessed the accuracy of LC-mNAATs for detecting pleural TB using pleural fluid and an MRS (**Fig. 2.2.1.9**). Estimated sensitivities were 48% and 75%, and estimated specificities were 89% and 96%. Due to limited data, a recommendation on using LC-mNAATs with pleural fluid for detection of pleural TB was not made.

Fig. 2.2.1.9 Forest plot of LC-mNAAT sensitivity and specificity for detection of pleural TB in pleural fluid and MRS

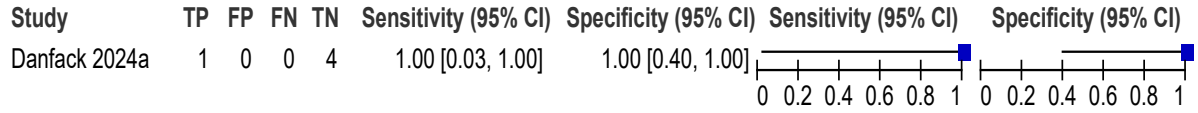


CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Should LC-mNAATs on synovial fluid be used to diagnose bone or joint TB in adults and adolescents with signs and symptoms of bone or joint TB, against an MRS?

One study (five participants, including one case) assessed the accuracy of LC-mNAATs for detecting bone or joint TB using synovial fluid and an MRS (**Fig. 2.2.1.10**). Estimated sensitivity and specificity were both 100%. Due to limited data, a recommendation on using LC-mNAATs with synovial fluid for detection of bone or joint TB was not made.

Fig. 2.2.1.10 Forest plot of LC-mNAAT sensitivity and specificity for detection of bone or joint TB in synovial fluid and MRS

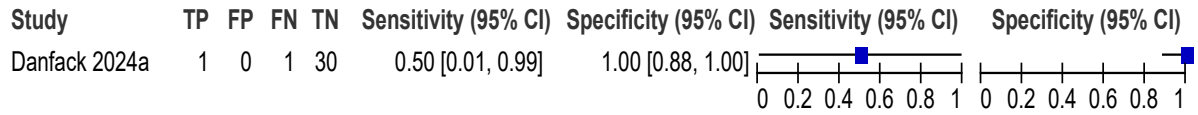


CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Should LC-mNAATs on urine be used to diagnose genitourinary TB in adults and adolescents with signs and symptoms of genitourinary TB, against an MRS?

One study (32 participants, including two people with TB) assessed the accuracy of LC-mNAATs for detecting genitourinary TB using urine and an MRS (**Fig. 2.2.1.11**). Estimated sensitivity and specificity were 50% and 100%, respectively. Due to limited data, a recommendation on using LC-mNAATs with urine for detection of genitourinary TB was not made.

Fig. 2.2.1.11 Forest plot of LC-mNAAT sensitivity and specificity for detection of genitourinary TB in urine and MRS



CI: confidence interval; FN: false negative; FP: false positive; LC-mNAAT: low-complexity manual nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Cost–effectiveness analysis

This section deals with the following additional question:

What are the comparative costs, affordability and cost–effectiveness of implementation of LC-mNAATs?

A systematic review commissioned by WHO aimed to identify, evaluate and summarize the findings of available economic evidence on LC-mNAATs, among other technologies. The systematic review provided an in-depth analysis of the financial implications and cost–effectiveness of implementing TB LAMP in diverse settings. Through a range of economic analyses, including cost–utility, cost–benefit and cost–affordability assessments, this study contributes valuable insights into the potential role of TB LAMP in TB diagnostics.

After removing 638 duplicate studies from those identified in the original search, 1990 unique studies remained. Of these, six studies were included in the final systematic review. Studies that did not involve people with TB, used TB LAMP as a diagnostic intervention or did not contain cost data were excluded. Of the six included studies, one performed a cost–utility analysis, and two performed a cost–affordability analysis. The three other studies estimated the cost of TB LAMP.

All included studies were conducted in LMIC. Specifically, two studies were conducted in Thailand, one in Malawi, one in both Malawi and Viet Nam, and one each in India and Cameroon. The studies were conducted between 2014 and 2021 across various settings, such as outpatient departments at health centres, peripheral laboratories, a laboratory for the development of modified TB LAMP, and prisons and villages involving inmates and refugees. One study used sputum samples from people known to have TB, and another used fine needle aspiration of lymph node samples from HIV-positive patients with TB lymphadenitis. The other four studies used sputum samples from people with presumptive TB.

According to the three costing studies, the cost per test ranged from US$ 1 to US$ 19 (all values in 2024 US dollars). All these studies used in-house techniques and were not using the commercially available TB LAMP test. The reviewed studies found that factors such as batching scenarios and larger test capacity influence the per-test cost, with the cost per test decreasing in specific scenarios. Testing volumes, location and operational parameters can also affect the cost. Notably, the cost–utility analysis positioned TB LAMP favourably in terms of cost–effectiveness compared with other diagnostic algorithms.

The findings of the cost–utility analysis suggested that TB LAMP, followed by DST, is not only effective but also cost-saving when compared with the standard diagnostic approach (i.e. smear, culture and DST). These results provide valuable insights for health care practitioners and policy-makers in terms of optimizing TB diagnostic strategies while considering cost–effectiveness.

One cost–affordability analysis, conducted in peripheral laboratories in Malawi and Viet Nam, highlighted the economic considerations of implementing TB LAMP and Xpert MTB/RIF. The study showed that per-test costs for TB LAMP were lower than those for Xpert MTB/RIF. However, the potential financial burden of widespread implementation underscored the importance of cost–effectiveness assessments in shaping diagnostic strategies. For more details see **Web Annex B.9**.

The reviewed studies had some limitations, such as variations in settings, sample sources and comparators, which may influence the generalizability of findings. Additionally, the cost–affordability analysis underscores the financial implications of nationwide implementation, suggesting the need for careful budgetary planning and allocation. Furthermore, the Global Drug Facility's recent decrease in the price of TB LAMP (new price, US$ 6) may have an impact on the results of economic evaluations, potentially enhancing the cost–effectiveness and affordability of implementing TB LAMP in diverse settings.

These collective findings suggest that TB LAMP holds promise as a cost-effective and efficient diagnostic tool for TB when integrated into broader diagnostic algorithms, particularly in resource-constrained settings.

More details on the economic evaluation of LC-mNAATs are available in **Web Annex B.9**.

User perspective

This section deals with the following question:

Are there implications for user preferences and values, acceptability, feasibility, patient equity and human rights from the implementation of LC-mNAATs?

The findings from the studies that focused on LC-aNAATs are largely applicable to LC-mNAATs, with the caveat of slightly lower sensitivity and the lack of ability to detect resistance to rifampicin.

A systematic review of the qualitative evidence of LC-NAATs (Web Annex B.10) did not identify any studies focused on LC-mNAATs (acknowledging that a few studies did not specify the type of NAAT they were focusing on).

However, selected findings from the interview study, did focus specifically on TB-LAMP:

- In 2018, Nigeria adopted the use of TB LAMP (along with GeneXpert and Truenat). At the time of the interview, there were 199 TB LAMP machines in Nigeria. These are placed both in sites where GeneXpert is available, to decrease workload, and in peripheral laboratories where the infrastructure is insufficient to accommodate GeneXpert. Positive results from TB LAMP are sent to the nearest site with a GeneXpert or Truenat machine for DST.
- The Philippines National TB Programme (NTP) guidelines advise the use of TB LAMP as an alternative primary diagnostic test in settings where access to GeneXpert is limited and that currently rely on sputum transport riders (STRiders). TB LAMP was piloted in 2019 (April to September) and 2020 (October to February 2021) in a rural health unit, polyclinic and private hospital, and for TB mass screening in a rural health unit. The pilot implementation only tested sputum with TB LAMP and did not test for TB in MDR risk groups, children or people living with HIV. According to a laboratory manager, there are about six or seven TB LAMP machines in the Philippines. These are not currently in use but could be if there was support for buying reagents.

According to Nigeria's TB LAMP guidelines, the following criteria should be used to prioritize sites for TB LAMP testing:

- facilities with high workload;
- facilities with or without an existing molecular platform;
- laboratories with adequate space and infrastructure;
- availability of qualified medical laboratory personnel;
- an adequate number of medical laboratory personnel; and
- a storage facility (e.g. refrigerator) for laboratory and administrative supply.

User preferences and values

An interview study on TB LAMP involving sites in Nigeria and the Philippines found the following views of laboratory personnel and programme officers:

- TB LAMP is making laboratory work easier over time through familiarity and because it clears the workbench;
- compared to SSM, TB LAMP is easier to use; and
- in direct comparison with Xpert Ultra, TB LAMP is more hands-on and requires more user steps and time for preparing and processing specimens.

Acceptability

Acceptability of the test seems to be slightly reduced because TB LAMP cannot test for rifampicin resistance and has no multiplexing opportunities.

Feasibility

Summarized findings from the interview study on TB LAMP are as follows. TB LAMP improves access to TB diagnosis for people who would otherwise have been missed, because it can run in laboratories with limited infrastructure, has high throughput, is more accurate than SSM and reduces workload at GeneXpert sites. It allows decentralization of testing and therefore has the potential to reduce catastrophic cost to patients. However, TB LAMP adoption decisions are also driven by donors and investment considerations.

Overall, TB LAMP allows staff to carry out more tests, and faster. Its high throughput contributes to acceptability and utilization. Laboratory staff in Nigeria are given incentives for the number of tests they carry out, making use of TB LAMP even more attractive. These incentives also support swift action when maintenance or repair of the devices is needed.

Programmatic feasibility seems to be less of a concern than with GeneXpert. Compared with implementing GeneXpert, programme officers find TB LAMP more feasible to implement due to its lower requirements for infrastructure, skills level and maintenance. As with all LC-NAATs, staffing and reagent supply issues challenge its use. TB LAMP's impact on the overall turnaround time for DR-TB diagnosis, MDR-TB treatment initiation and loss to follow-up at sites without Xpert Ultra testing depends on the efficiency and robustness of the sample transport or referral system.

More details on the qualitative evaluation of LC-mNAATs are available in **Web Annex B.10.**

Implementation considerations

- Diagnostic products in the low-complexity classes of tests should be prequalified by WHO or approved by another regulator before clinical use.
- Diagnostic test manufacturers, laboratory and programme managers, and policy-makers should be educated on the WHO PQ process for TB IVDs.
- Ensuring sufficient volume and specimen quality is important to obtain accurate results.
- Safe waste disposal of used test consumables needs to be planned in advance to minimize environmental risk.

Monitoring and evaluation

- Track errors and invalid test result rates for currently recommended products and new products to be introduced in this class.
- Monitor the proportion of people with bacteriologically confirmed TB without rifampicin-resistance reflex testing or access to further DST over time.

Research priorities

- Evaluate the performance of this class using alternative sample types for paediatric TB (e.g. gastric and nasopharyngeal aspirates, stool, induced sputum, BAL) and extrapulmonary TB.
- Evaluate the impact of LC-mNAAT testing on patient-important outcomes (cure, mortality, time to diagnosis and time to start of treatment).
- Evaluate the effect of sample concentration approaches (e.g. centrifugation) and volume on the performance of LC-mNAAT technologies, including in extrapulmonary TB sample types.
- Evaluate the impact on incremental accuracy and case detection of alternative sample types that are easier to collect.
- Develop a test in this class that can detect TB drug resistance.
- Review the field performance of the current technologies used in programmatic settings.
- Conduct operational research to ensure that tests are used optimally in intended settings.
- Evaluate the different classes of tests, including LC-mNAATs, to determine which classes or testing strategies yield superior diagnostic accuracy, cost–effectiveness and impact on equity and acceptability.
- Identify an improved reference standard that accurately defines TB disease in children, paucibacillary specimens and people who cannot produce sputum, because the sensitivity of all available diagnostics is suboptimal.
- Assess the budget impact and cost–effectiveness of LC-mNAATs compared with other classes of tests.
- Develop and apply standardized methods for assessment of costs and cost–effectiveness, to improve comparability and scope of economic evidence.

2.3. Concurrent use of initial diagnostic tests for diagnosis of TB in People living with HIV and children

There are significant burdens of tuberculosis in people living with HIV and children, particularly in low- and middle-income countries (LMICs). Persons living with HIV are at substantially higher risk of developing TB disease due to immunosuppression, with TB being a leading cause of death among this population. Children, especially those under five, are at high risk of progression from TB infection to TB disease and rapid disease progression and often present with broad respiratory symptoms, which complicate diagnosis and increase morbidity and mortality if not promptly treated. Addressing TB in these at-risk populations requires concerted efforts that account for their unique clinical presentations and diagnostic needs.

Diagnosing TB in persons living with HIV and children is challenging, particularly because of unspecific clinical presentations and often low and varying numbers of mycobacteria in their samples that lower the sensitivity of existing diagnostic tests. Furthermore, children and people living with HIV with advanced immunosuppression may be unable to provide sputum samples and can have disseminated TB, which is challenging to confirm with laboratory methods. To, in part address this challenge, WHO recommends the use of stool to aid in laboratory confirmation of TB in children, and the use of urine to aid in the confirmation of TB in persons living with HIV. However, even highly sensitive tests for TB diagnosis, such as LC-aNAATs, can miss TB in these

groups. There is therefore a need for improved diagnostic approaches to accurately confirm TB in these higher-risk populations to ensure early and effective treatment.

Tests based on the detection of the lipoarabinomannan (LAM) antigen are biomarker-based tests that may be used on urine at the point of care for TB detection. The currently available urinary LAM assay is rapid (<1 hour to result) but has suboptimal sensitivity and is therefore not suitable as general diagnostic tests for TB. However, unlike traditional diagnostic methods, it demonstrates improved sensitivity for the diagnosis of TB among individuals coinfected with HIV. The estimated sensitivity is even greater in patients with low CD4 cell counts. The lateral flow urine LAM assay (LF-LAM) strip-test – the Abbott/Alere Determine TB LAM Ag (USA), hereafter referred to as LF-LAM – is currently the only commercially available urinary LAM test.

Using concurrent[10] testing of different sample types offers a promising approach that considers the diagnostic testing barriers for HIV-positive adults and adolescents, HIV-positive children, and children without HIV or for whom HIV status is unknown. For instance, testing of sputum and stool during the same visit, when feasible, using LC-aNAATs increases the likelihood of detecting TB in children who may have scant bacilli in respiratory samples alone. Similarly, for persons living with HIV, testing of sputum and urine during the same visit, when sputum can be produced, using LC-aNAATs and LF-LAM increases the likelihood of detecting TB with a rapid point-of-care result while also ensuring detection of rifampicin resistance. This concurrent testing approach builds on the prior recommendation for LF-LAM test use among eligible persons living with HIV, which underscored the need for mWRD testing of available respiratory samples to support universal patient access to resistance testing services.

Implementing a diagnostic approach that includes concurrent sample testing could simplify diagnostic processes, shorten the patient journey, and improve TB detection rates and health outcomes for these at-risk populations. At the same time, the inability to collect one or more specimens at the same initial visit, or lack of one of the two test types should not delay testing of available specimens and tests, but instead trigger specimen collection and testing as soon as possible.

The following three scenarios of recommendations:

- LC-aNAAT on respiratory samples and urine LF-LAM among adults and adolescents living with HIV
- LC-aNAAT on respiratory samples and stool in children
- LC-aNAAT on respiratory samples and stool, as well as urinary LF-LAM among children with HIV

These recommendations should be implemented within recommendations for the comprehensive diagnosis and management of persons living with HIV and children.

[10] Concurrent use of tests: samples are taken simultaneously (when possible), and testing is conducted for both tests. A positive result on either test is a positive result for the combination.

2.3.1 Concurrent use of tests in people living with HIV

NEW

Recommendations

7. For adults and adolescents with HIV who have signs or symptoms of TB, screen positive for TB, are seriously ill or have advanced HIV disease, concurrent testing using low-complexity automated NAATs on respiratory samples and LF-LAM on urine should be used as the initial diagnostic strategy for diagnosing TB rather than low-complexity automated NAATs on respiratory samples alone.

(Strong recommendation, low certainty of evidence)

Remarks

- Serious illness in people living with HIV is defined based on any of the following symptoms: respiratory rate ≥30 breaths per minute, temperature ≥39 °C, heart rate ≥120 beats per minute or unable to walk unaided.
- Advanced HIV disease is defined in people living with HIV who have a CD4 cell count of <200 cells/mm^3 or presenting with a WHO Stage 3/4 AIDS-defining illness.
- This concurrent testing recommendation supersedes prior guidance on using LF-LAM for people living with HIV and the use of a single molecular test for diagnosis of TB in this group.
- This recommendation is strong despite the low certainty of evidence because the findings indicate large desirable effects (i.e. rapid and accurate diagnosis of TB in a highly vulnerable population – people living with HIV – in whom diagnosing TB is often challenging) over small undesirable effects (i.e. negative consequences of this testing strategy).
- The LC-aNAAT products for which eligible data met the class-based performance criteria for this recommendation were Xpert MTB/RIF Ultra and Truenat MTB Plus. Data for performance of Truenat MTB Plus and MTB-RIF Dx were only available for testing among persons living with HIV without concurrent LF-LAM testing.

Justification and evidence

In a 2016 Cochrane systematic review of the diagnostic accuracy of LF-LAM, sensitivity increased by 13% when combining LF-LAM and sputum Xpert MTB/RIF, compared with sputum Xpert alone, while the specificity decreased by 4%. However, results were based on only a few studies, and analyses were restricted to participants able to produce sputum.

Incremental diagnostic accuracy

In 2023, WHO commissioned a series of systematic reviews to evaluate the incremental diagnostic accuracy[11] of concurrent use of either two different tests – LC-aNAAT on respiratory samples and LF-LAM on urine among people living with HIV – or the same test on two samples (LC-aNAAT on respiratory and stool samples) in children, or alternatively LC-aNAAT on respiratory and stool samples along with LF-LAM on urine among children with HIV.

[11] Incremental change in diagnostic accuracy with concurrent testing compared with individual sample testing.

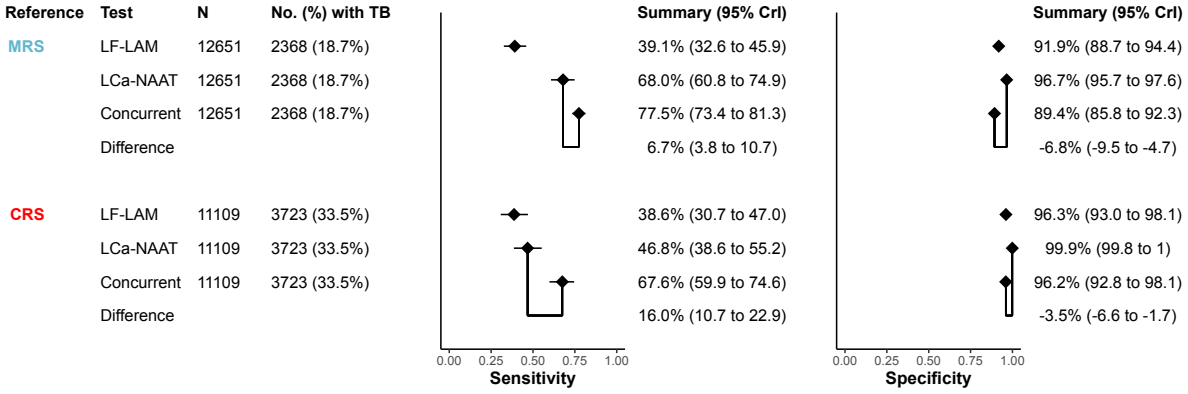
What is the incremental diagnostic accuracy of concurrent use of respiratory LC-aNAATs and LF-LAM on urine for diagnosis of TB disease in adults and adolescents with HIV who present with presumptive TB, compared with any of the tests alone?

Of 31 studies, 27 evaluated diagnostic accuracy against an MRS, and 23 against a CRS, with 20 studies evaluating accuracy against both reference standards.

A total of 27 studies (12 651 participants, including 2368 [18.7%] with TB) compared the accuracy of the concurrent use of LC-aNAAT on a respiratory sample and LF-LAM versus each of the tests alone, using an MRS. The pooled differences in sensitivity and specificity between concurrent testing versus LC-aNAAT alone were 6.7% (95% credible interval [CrI]: 3.8 to 10.7; 95% prediction interval [PI]: 0.6 to 45.9) and –6.8% (95% CrI: –9.5 to –4.7; 95% PI: –32.8 to –6.8), respectively (**Fig. 2.3.1.1**). Certainty of evidence was low for both sensitivity and specificity.

A total of 23 studies (11 109 participants, including 3723 [33.5%] with TB) compared the accuracy of the concurrent use of LC-aNAAT and LF-LAM versus LC-aNAAT alone, using a CRS. The pooled differences in sensitivity and specificity between concurrent testing versus LC-aNAAT alone were 16.0% (95% CrI: 10.7 to 22.9; 95% PI: 2.3 to 60.3) and –3.5% (95% CrI: –6.6 to –1.7; 95% PI: –47.2 to –0.1), respectively (**Fig. 2.3.1.1**). Certainty of evidence was low for sensitivity and very low for specificity.

Fig. 2.3.1.1. Forest plot of pooled differences in sensitivity and specificity (all studies combined) by index test: LF-LAM, LC-aNAAT and their concurrent use[a]



CrI: credible interval; CRS: composite reference standard; LC-aNAAT: low-complexity automated nucleic acid amplification test; LF-LAM: lateral flow urine lipoarabinomannan assay; MRS: microbiological reference standard; TB: tuberculosis.

[a] The diamonds represent the pooled sensitivity and specificity, and the black horizontal line its 95% CrI. The pooled difference in sensitivity and specificity between concurrent testing and LC-aNAAT alone is indicated by a line connecting two diamonds. This pooled difference may not correspond to the difference between the pooled single test accuracy estimates (see **Web Annex B.8**).

In addition to diagnostic accuracy, clinical outcome data on mortality, time to diagnosis and time to treatment were assessed. Data on cure and loss to follow-up were not assessed due to a lack of data. The data from three studies indicated that an intervention including LC-aNAAT on respiratory samples and LF-LAM on urine in adult inpatients with HIV was associated with slightly reduced 8-week mortality (risk ratio: 0.93; 95% CI: 0.74–1.17). The adjusted hazard ratio of time to diagnosis in adult inpatients with HIV was 1.55 (95% CI: 1.29–1.87). This means

that participants in the intervention groups (i.e. those undergoing concurrent LC-aNAAT on respiratory samples and LF-LAM on urine) were 1.55 times more likely to be diagnosed with TB within fewer days (relative reduction of 2 days and 1 day to same-day) than those in the control group. The pooled risk ratio of adult inpatients with HIV diagnosed with TB was 1.56 (95% CI: 1.29–1.88), indicating that the intervention group had 1.56 times the risk of being diagnosed with TB (either microbiologically confirmed or clinically diagnosed) compared with the standard of care, which included LC-aNAAT on sputum alone. The pooled risk ratio of adult inpatients with HIV with a bacteriologically confirmed TB diagnosis was 3.06 (95% CI: 1.82–5.16), indicating that the intervention group had three times the risk of being microbiologically confirmed with TB compared with the standard of care. Finally, the pooled risk ratio of adult inpatients with HIV treated for TB was 1.47 (95% CI: 1.25–1.73), indicating that the intervention group had 1.47 times the likelihood of being treated for TB, compared with the standard of care.

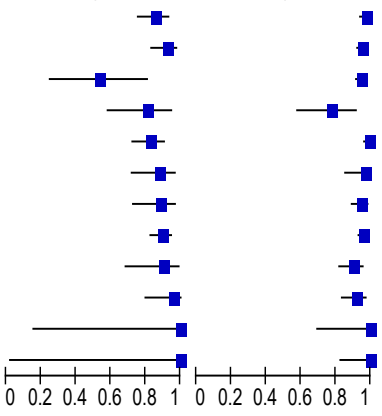
Single sample testing in people living with HIV compared with the MRS

Should LC-aNAATs on respiratory samples be used to diagnose pulmonary TB in PLHIV (adults and adolescents) with signs and symptoms or screened positive for pulmonary TB, against a microbiological reference standard?

Twelve studies (2016 participants) evaluated sputum specimens from people living with HIV (**Fig. 2.3.1.2**). The sensitivities ranged between 54% and 100% and the specificities between 78% and 100%. The summary sensitivity (95% CI) was 87.4% (83.8 to 90.3) and the summary specificity was 95.2% (92.7 to 96.9). The certainty of evidence for both sensitivity and specificity were graded as "High".

Fig. 2.3.1.2 Forest plot of LC-aNAAT sensitivity and specificity for detection of pulmonary TB in PLHIV using a microbiological reference standard

Study	TP	FP	FN	TN	Brand	Sensitivity (95% CI)	Specificity (95% CI)
Theron 2024	60	3	10	130	Truenat MTB Plus	0.86 [0.75, 0.93]	0.98 [0.94, 1.00]
Ngangue 2022	60	14	5	273	Truenat MTB Plus	0.92 [0.83, 0.97]	0.95 [0.92, 0.97]
Mukoka 2023	7	17	6	289	Xpert Ultra	0.54 [0.25, 0.81]	0.94 [0.91, 0.97]
Mishra 2020a	17	6	4	21	Xpert Ultra	0.81 [0.58, 0.95]	0.78 [0.58, 0.91]
Boyles 2020	62	1	13	131	Xpert Ultra	0.83 [0.72, 0.90]	0.99 [0.96, 1.00]
Sessolo 2023	29	1	4	34	Xpert Ultra	0.88 [0.72, 0.97]	0.97 [0.85, 1.00]
Berhanu 2018	30	6	4	107	Xpert Ultra	0.88 [0.73, 0.97]	0.95 [0.89, 0.98]
Dorman 2018	103	14	12	303	Xpert Ultra	0.90 [0.82, 0.94]	0.96 [0.93, 0.98]
Ssengooba 2024	18	9	2	80	Xpert Ultra	0.90 [0.68, 0.99]	0.90 [0.82, 0.95]
Zhang 2021	24	6	1	68	Xpert Ultra	0.96 [0.80, 1.00]	0.92 [0.83, 0.97]
Sabi 2018	2	0	0	10	Xpert Ultra	1.00 [0.16, 1.00]	1.00 [0.69, 1.00]
Park 2019	1	0	0	19	Xpert Ultra	1.00 [0.03, 1.00]	1.00 [0.82, 1.00]



Studies are sorted on the plot by assay and sensitivity (low to high). FN: false negative; FP: false positive; TN: true negative; TP: true positive.

Cost–effectiveness analysis

To date, evidence of cost–effectiveness for concurrent testing is limited. Several studies have assessed Xpert MTB/RIF with LF-LAM for diagnosing TB among people living with HIV. These studies have shown that concurrent testing is likely to increase the life expectancy of people living with HIV and be cost effective compared with using Xpert MTB/RIF in sputum samples

alone. Fekuda et al. evaluated the cost–effectiveness of concurrently using Xpert Ultra and LF-LAM among people living with HIV and concluded that concurrent testing is the preferred cost-effective strategy. Previous cost–effectiveness analyses primarily focused on Xpert MTB/RIF or Xpert Ultra, leaving a gap in evidence regarding the other technologies that may meet the LC-aNAAT class criteria. For details of particular studies see **Web Annex B.9.**

In preparation for the GDG meeting in May 2024, WHO commissioned a study to assess the cost–effectiveness of using LC-aNAATs (including Xpert Ultra, Truenat and other novel LC-aNAATs in the development pipeline) for the detection of TB when used concurrently among people living with HIV and children, including children with HIV, across two different country settings (Malawi and the Philippines). An objective of the study was to assess the cost–effectiveness of concurrent use of LC-aNAAT on respiratory samples and LF-LAM on urine for TB diagnosis and rifampicin-resistance detection among adult people living with HIV with presumptive TB, compared with a single LC-aNAAT on respiratory samples alone.

In the hypothetical model, a cohort of people living with HIV with signs and symptoms of TB progressed through a decision analytical framework. In the intervention arm, TB diagnosis involved the concurrent use of LC-aNAAT on respiratory samples and LF-LAM on urine, whereas the comparator arm exclusively used LC-aNAAT on respiratory specimens. The probability of being able to provide a respiratory sample was considered, and testing was carried out, either on both respiratory and urine samples concurrently or solely on urine. In both intervention and comparator arms, participants not diagnosed through the diagnostic strategy had the opportunity for clinical diagnosis. People with bacteriologically confirmed TB underwent DST for rifampicin and began either drug-susceptible TB or DR-TB treatment, depending on the DST result. All individuals were followed over time, including those with false negative or false positive diagnostic results, to account for unnecessary treatment or additional mortality due to missed diagnoses.

The cost–effectiveness results of concurrent use of LC-aNAAT with LF-LAM among people living with HIV, when used in the emblematic settings of Malawi and the Philippines, are shown in Table 2.3.1.1 In Malawi, the average cost of implementing an LC-aNAAT on a respiratory sample was US$ 276, with a corresponding average DALY of 2.44. When used concurrently with LF-LAM, the average cost rose to US$ 298, while the average DALY decreased to 1.93. The resulting incremental cost per DALY averted was US$ 42, with a 95% uncertainty range (UR) of US$ 18 to US$ 345. Similarly, in the Philippines, LC-aNAAT on a respiratory sample had an average cost of US$ 220, with an average DALY of 2.78, whereas concurrent use with LF-LAM incurred an average cost of US$ 238 and an average DALY of 2.13. The incremental cost per DALY averted was US$ 28 (95% UR: 12–249).

Table 2.3.1.1 Cost–effectiveness analysis of concurrent use of LC-aNAAT and LF-LAM among people living with HIV in Malawi and the Philippines

Country	Diagnostic strategy	Cost, US$	Effectiveness, DALYs	ICER (95% UR), US$
Malawi	LC-aNAAT on respiratory sample	276	2.44	Ref
	LC-aNAAT on respiratory sample and LF-LAM	298	1.93	42 (18–345)
Philippines	LC-aNAAT on respiratory sample	220	2.78	Ref
	LC-aNAAT on respiratory sample and LF-LAM	238	2.13	28 (12–249)

DALY: disability-adjusted life year; HIV: human immunodeficiency virus; ICER: incremental cost–effectiveness ratio; LC-aNAAT: low-complexity automated nucleic acid amplification test; LF-LAM: lateral flow urine lipoarabinomannan assay; UR: uncertainty range.

More information on the cost–effectiveness analysis of concurrent use of tests in people living with HIV is available in **Web Annex B.9.**

User perspective

This section deals with the following question:

Are there implications for user preferences and values, equity, acceptability, feasibility and human rights from the implementation of a concurrent testing approach (LC-aNAATs + LF-LAM)?

The GDG assessed whether concurrent testing of multiple samples would increase the diagnostic accuracy (i.e. the benefit to patients or the programme in terms of finding more people with TB). Three PICO questions concerned the different concurrent sample combinations for specific groups facing challenges from reliance on respiratory samples alone (children and people living with HIV). One question focused on the concurrent use of LC-aNAAT on a respiratory sample and LF-LAM on urine for the diagnosis of TB in people living with HIV.

User preferences and values

As important outcomes of the diagnostic test, people in high TB burden settings value:

- getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me");
- avoiding diagnostic delays, as they exacerbate existing financial hardships and emotional and physical suffering and make people feel guilty for infecting others (especially children);
- having accessible facilities; and
- reducing diagnosis-associated costs (e.g. travel, missing work).

More details on patient-important outcomes are available in **Web Annex B.10.**

Equity

Concurrent specimen testing was not practiced in the interview study countries. However, it was believed to improve access to care by minimizing repeat visits and loss to follow-up. According to the interview study respondents, using non-sputum specimens has the potential to improve access to care, especially with a test that can be performed at all levels of the health care system. Challenges with producing a sufficient quality and quantity of sputum are well documented and can lead to repeat testing or false results.

Acceptability

Based on the results of the interview study, LF-LAM is being used inconsistently for people living with HIV and only for very ill patients who cannot produce sputum. Our results are in accordance with published literature on LF-LAM.

Prior research on user perspectives on LF-LAM showed that it is generally described as acceptable by key stakeholders, due to its fast turnaround time, ease of use (lack of technical expertise required), low or no maintenance and equipment required, and urine being more accessible and less stigmatized than sputum. LF-LAM is deemed particularly acceptable when used in combination with other tests and clinical considerations. As the sensitivity of LF-LAM is especially low where the pretest probability is low, participants commented that it should not be used as a standalone test but should instead be used in combination with other tests, and that the results should be interpreted by a doctor considering the full clinical context, rather than being considered in isolation.

Feasibility

Interview study findings highlighted that the benefits of LF-LAM are crucially dependent on how several feasibility challenges are addressed.

- Hygienic, safe and private sanitary facilities with running water are necessary for LF-LAM implementation at a testing site, but they are not always available, particularly in rural areas. Investments in staffing and sanitary facilities are required.
- Not everybody can spontaneously produce or collect urine samples. This can be the case, for example, when the patient is too ill or septic or has to be catheterized because collecting urine samples from diapers is impossible, or if the hospital has no clean, private space to produce urine.
- Visibility of faint results and result interpretation can be problematic. Comprehensive health care worker training in test interpretation (including mandatory use of the reference reading card, where appropriate) is crucial to ensure accurate result interpretation for clinical action.
- The need for CD4 cell count results to select people for the test is problematic because these are not always immediately available. To facilitate implementation and benefit a wider range of individuals, eliminating the CD4 cell count as an eligibility criterion for people living with HIV should be considered.
- In a hospital setting, bedside testing may violate patient confidentiality.
- Results must be captured in a standardized way that feeds into facility and NTP reporting systems.
- Quality assurance schemes need to be rolled out, and external quality controls need to be made available, to ensure tests and testing processes are quality controlled.

Concurrent testing needs to be framed as a more efficient way of working (i.e. testing two samples concurrently during the same visit, instead of testing one sample during each of two separate visits) that also allows increasing access and reducing costs for patients. According to a laboratory manager, this framing of the benefits outweighing the additional workload, and potentially resulting in reduced work in the long run, will be critical to avoid concurrent testing being perceived as additional work for already overburdened health care workers (see **Web Annex B.10**).

Prior investments made in frontrunner technologies, donor preferences, limited health systems thinking and unnecessary competition between manufacturers all pose challenges to policy adoption and implementation of novel molecular diagnostics. In addition, national in-country health technology and cost-efficacy assessments can delay decisions to implement newer technologies and diagnostic strategies using different samples (see **Web Annex B.10**).

Implementation considerations

- Global and national HIV and TB programmes need to communicate regularly and clearly, indicating responsibilities for concurrent testing for people living with HIV.
- Concurrent testing maximizes diagnostic opportunity and accuracy of case detection, is a more efficient way to address the needs of this population and is preferred even if the testing workload may increase.
- A positive result on either test is sufficient to confirm TB diagnosis.
- Patient loss to follow-up for the second test result should be monitored and prevented. Patients should be provided with information to understand the concurrent testing approach and the need for follow-up.
- The LF-LAM performed in point-of-care settings may be the first positive result and is sufficient to make the initial diagnosis. A respiratory sample is still required for rifampicin-resistance detection, and is also required when the LF-LAM result is negative.
- Where LF-LAM is not available for testing of people living with HIV, efforts should be made to ensure access to testing.
- LF-LAM does not differentiate *Mtb* from other mycobacterial species. However, the LAM antigen detected in a clinical sample in TB endemic areas is most likely attributable to *Mtb*.
- When LF-LAM results are consistently positive, without positive LC-aNAAT results, investigation of the quality of testing and local epidemiology of non-tuberculosis mycobacteria and extrapulmonary TB in the tested population is warranted to understand the difference.
- Interpreting bands on the LF-LAM test strip should be performed using the manufacturer's reading card to minimize incorrect results.
- LF-LAM test strips must be stored according to the manufacturer's instructions (e.g. between 2 and 30 °C) in sealed bags and not used after expiration.
- Infrastructure to collect a urine sample privately should be available. Patients should be instructed how to properly and sanitarily collect a urine sample to minimize contamination and prevent false positive results.
- Trained staff will be required to perform the LF-LAM test at the point of care.
- As with all WHO-recommended TB diagnostics, quality assurance programmes and quality controls for both tests are required.

- LF-LAM is designed to detect mycobacterial LAM antigen in human urine. Other samples (e.g. sputum, serum, plasma, CSF and other body fluids) or pooled urine specimens should not be used.

Monitoring and evaluation

- Monitor simultaneous specimen collection and turnaround time for the test results in a concurrent testing approach.
- Monitor patient access to, and loss to follow-up from, a second test in a concurrent testing approach.
- Monitor patient access to, and loss to follow-up from, follow-on DST among those with a positive LF-LAM result but a negative LC-aNAAT result.
- Monitor trends in the discordance rate between the LF-LAM and LC-aNAAT results. If these differences vary from other local or regional patterns, or if the trends change, further investigation is required and outcomes should be tracked for recurrence over time.

Research priorities

- Conduct more rigorous studies with higher quality reference standards, including multiple specimen types and extrapulmonary samples, to improve confidence in specificity estimates.
- Gather evidence on the impact of concurrent testing on TB treatment initiation and mortality.
- Determine training, competency and quality assessment needs by setting and by cadre of staff (i.e. health care worker, laboratory technician or clinical staff).
- Perform country-specific cost–effectiveness and cost–benefit analyses of the concurrent testing approaches or sequential testing approaches in different programmatic settings.
- Develop and apply standardized methods for assessment of costs and cost–effectiveness, to improve comparability and scope of economic evidence.
- Perform operational research on availability, requirements and best practices for the point-of-care set-up: private specimen collection facility, tabletop space for testing samples, and reporting system (preferably digital) for entry of results, with linkages to existing information management systems (i.e. health and laboratory information management systems).

2.3.2 Concurrent use of tests in children without HIV or with unknown HIV status

NEW

Recommendations

8. For children who are HIV-negative or have an unknown HIV status, who have signs or symptoms or screen positive for pulmonary TB, concurrent testing using low-complexity automated NAATs on respiratory and stool samples should be used as the initial diagnostic strategy for diagnosing TB rather than low-complexity automated NAATs on respiratory or stool samples alone.

(Strong recommendation, low certainty of evidence for test accuracy)

Remarks

- This recommendation prioritizes concurrent testing of two different sample types over the use of a single molecular test for diagnosis of TB in children.
- Use of LC-aNAATs on isolated specimens was also evaluated. The findings supported the use of LC-aNAATs for initial diagnostic testing for TB in children with signs or symptoms or who screen positive for pulmonary TB, using respiratory sample, gastric aspirate, stool or nasopharyngeal aspirate, rather than smear or culture.
- This recommendation is strong despite the low certainty of evidence because the findings indicate large desirable effects (i.e. rapid and accurate diagnosis of TB in a highly vulnerable population – children – in whom diagnosing TB is often challenging) over trivial undesirable effects (i.e. negative consequences of this testing strategy) (for more details, see GRADE evidence to decision [EtD] table, **Web Annex A.4**).
- The product for which eligible data met the LC-aNAAT class-based performance criteria for this recommendation was Xpert MTB/RIF Ultra. The performance of Truenat MTB Plus and MTB-RIF Dx for this recommendation could not be assessed, as data were unavailable.

Justification and evidence

LC-aNAATs on respiratory and stool samples are recommended as the first test for symptomatic children presenting with presumptive TB disease, and are widely used to diagnose TB.

Previous systematic reviews have traditionally assessed diagnostic accuracy of LC-aNAATs on two samples in isolation for the detection of TB in children, but in clinical practice the tests may be used concurrently (i.e. LC-aNAAT on a respiratory sample and a stool sample) and together they increase sensitivity.

Incremental diagnostic accuracy of concurrent testing compared with single sample testing

What is the incremental diagnostic accuracy of concurrent use of LC-aNAATs on respiratory and stool samples for diagnosis of pulmonary TB disease in children who are HIV-negative or have an unknown HIV status, with signs and symptoms or who screened positive for pulmonary TB, compared with use of an LC-aNAAT on one sample type (either respiratory or stool)?

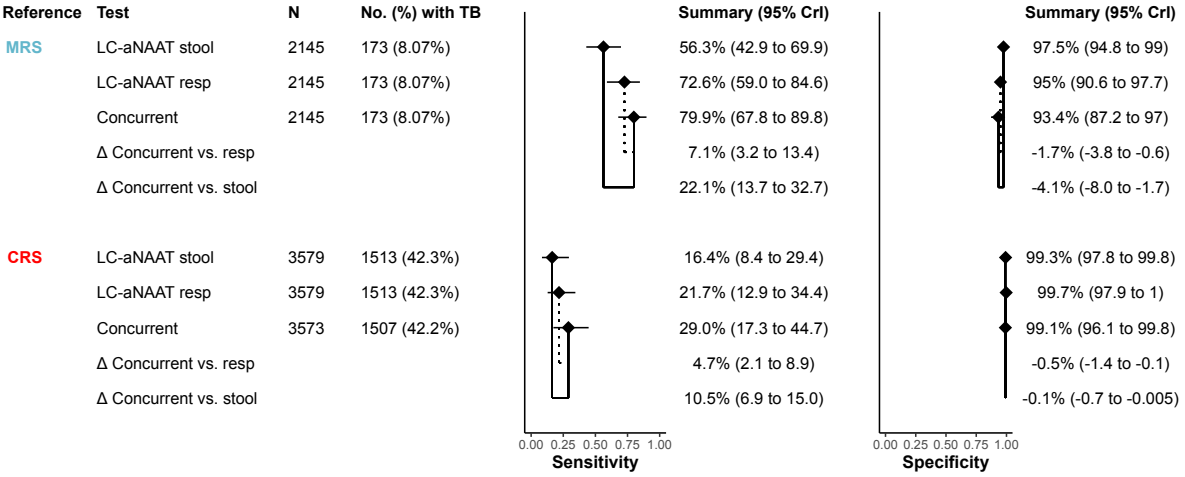
Eight studies (2145 participants, 173 [8.1%] of whom had TB disease) compared the accuracy of concurrent use of LC-aNAATs with respiratory and stool samples (LC-aNAATs combined) versus LC-aNAAT on one sample type (either respiratory or stool) against an MRS.

Compared with LC-aNAAT on respiratory samples alone, concurrent testing had 7.1 percentage points (95% CrI: 3.2 to 13.4) higher sensitivity and –1.7 percentage points (95% CrI: –3.8 to –0.6) lower specificity. Certainty of evidence for both sensitivity and specificity was low for comparison with LC-aNAAT on respiratory samples alone. Compared with LC-aNAAT on stool alone, concurrent testing had 22.1 percentage points (95% CrI: 13.7 to 32.7) higher sensitivity and –4.1 percentage points (95% CrI: –8.0 to –1.7) lower specificity. Certainty of evidence was moderate for sensitivity and low for specificity for comparison with LC-aNAAT on stool alone.

Twelve studies (3579 participants, 1464 [40.9%] of whom had TB disease) compared the accuracy of LC-aNAATs combined versus each LC-aNAAT alone against a CRS.

Compared with LC-aNAAT on respiratory samples alone, concurrent testing had 4.7 percentage points (95% CrI: 2.1 to 8.9) higher sensitivity and –0.5 percentage points (95% CrI: –1.4 to 0) lower specificity. Compared with LC-aNAAT on stool alone, concurrent testing had 10.5 percentage points (95% CrI: 6.9 to 15.0) higher sensitivity and –0.1 percentage points (95% CrI: –0.7 to –0.005) lower specificity. Certainty of evidence was very low for both sensitivity and specificity for both comparisons (concurrent testing versus respiratory sample alone and stool alone) under a CRS (**Fig. 2.3.2.1**). The data on Truenat MTB Plus and MTB-RIF Dx were unavailable.

Fig. 2.3.2.1 Forest plot of pooled sensitivity and specificity for all studies, by each index test



CrI: credible interval; CRS: composite reference standard; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis.

The diamonds represent pooled sensitivity and specificity, and the black horizontal line its 95% CrI. The difference in accuracy between index tests is indicated by solid lines (concurrent versus stool) or dotted lines (concurrent versus respiratory) connecting the diamonds.

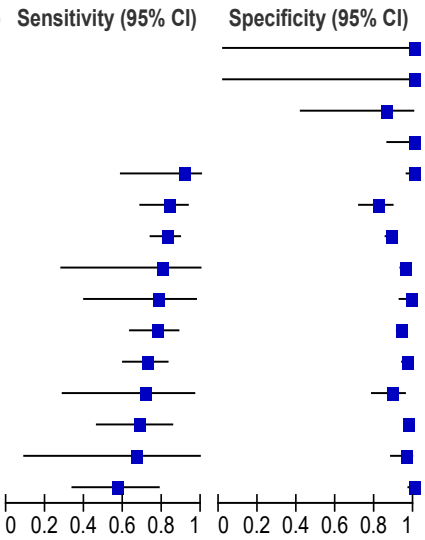
Single sample testing in children compared with the MRS

Should LC-aNAATs on respiratory samples be used to diagnose pulmonary TB in children with signs and symptoms or who screened positive for pulmonary TB, against an MRS?

Fifteen studies (3024 participants) evaluating sputum were identified, with sensitivities ranging between 57% and 91% and specificities between 82% and 100% (**Fig. 2.3.2.2**). Eleven studies (2990 participants) were included in the meta-analysis. The summary sensitivity was 75.3% (95% CI: 68.9–80.8) and summary specificity was 95.9% (95% CI: 92.3–97.9). Certainty of evidence was high for both sensitivity and specificity. The data on Truenat MTB Plus and MTB-RIF Dx were unavailable.

Fig. 2.3.2.2 Forest plot of LC-aNAAT sensitivity and specificity for detection of pulmonary TB in sputum samples and MRS

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aguilera-Alonso 2022	0	0	0	1	Not estimable	1.00 [0.03, 1.00]
Barcellini 2019	0	0	0	1	Not estimable	1.00 [0.03, 1.00]
Liu 2021	0	1	0	6	Not estimable	0.86 [0.42, 1.00]
NCT04121026	0	0	0	25	Not estimable	1.00 [0.86, 1.00]
Yenew 2023	10	0	1	93	0.91 [0.59, 1.00]	1.00 [0.96, 1.00]
Enimill 2022	35	16	7	71	0.83 [0.69, 0.93]	0.82 [0.72, 0.89]
Zar 2018-2023	89	66	19	499	0.82 [0.74, 0.89]	0.88 [0.85, 0.91]
Kabir 2020	4	18	1	379	0.80 [0.28, 0.99]	0.95 [0.93, 0.97]
Jaganath 2021	7	1	2	71	0.78 [0.40, 0.97]	0.99 [0.93, 1.00]
NCT03734172	38	36	11	519	0.78 [0.63, 0.88]	0.94 [0.91, 0.95]
Nicol 2018	47	9	18	261	0.72 [0.60, 0.83]	0.97 [0.94, 0.98]
NCT04899076	5	7	2	56	0.71 [0.29, 0.96]	0.89 [0.78, 0.95]
Sabi 2022	17	10	8	340	0.68 [0.46, 0.85]	0.97 [0.95, 0.99]
Ssengooba 2020	2	3	1	68	0.67 [0.09, 0.99]	0.96 [0.88, 0.99]
Sabi 2018	12	0	9	122	0.57 [0.34, 0.78]	1.00 [0.97, 1.00]



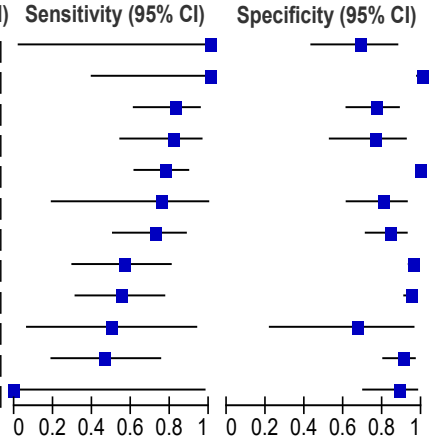
CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Should LC-aNAATs on gastric aspirate specimens be used to diagnose pulmonary TB in children with signs and symptoms or who screened positive for pulmonary TB, against an MRS?

Twelve studies (1959 participants) were identified, with sensitivities between 0% and 100% and specificities between 67% and 100% (**Fig. 2.3.2.3**). All 12 studies were included in the meta-analysis. The summary sensitivity was 69.6% (95% CI: 60.3–77.6) and summary specificity was 91.0% (95% CI: 82.5–95.6). Certainty of evidence was moderate for both sensitivity and specificity. The data on Truenat MTB Plus and MTB-RIF Dx were unavailable.

Fig. 2.3.2.3 Forest plot of LC-aNAAT sensitivity and specificity for detection of pulmonary TB in gastric aspirate and MRS[a]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
NCT04899076	1	6	0	13	1.00 [0.03, 1.00]	0.68 [0.43, 0.87]
NCT04121026	4	0	0	137	1.00 [0.40, 1.00]	1.00 [0.97, 1.00]
Aguilera-Alonso 2022	19	10	4	33	0.83 [0.61, 0.95]	0.77 [0.61, 0.88]
Parigi 2021	13	5	3	16	0.81 [0.54, 0.96]	0.76 [0.53, 0.92]
Yenew 2023	31	8	9	632	0.78 [0.62, 0.89]	0.99 [0.98, 0.99]
NCT03734172	3	6	1	24	0.75 [0.19, 0.99]	0.80 [0.61, 0.92]
Liu 2021	18	9	7	46	0.72 [0.51, 0.88]	0.84 [0.71, 0.92]
NCT04240990	9	25	7	504	0.56 [0.30, 0.80]	0.95 [0.93, 0.97]
Ssengooba 2020	11	12	9	212	0.55 [0.32, 0.77]	0.95 [0.91, 0.97]
Enimill 2022	2	2	2	4	0.50 [0.07, 0.93]	0.67 [0.22, 0.96]
Sabi 2022	6	6	7	56	0.46 [0.19, 0.75]	0.90 [0.80, 0.96]
Jaganath 2021	0	3	1	23	0.00 [0.00, 0.97]	0.88 [0.70, 0.98]



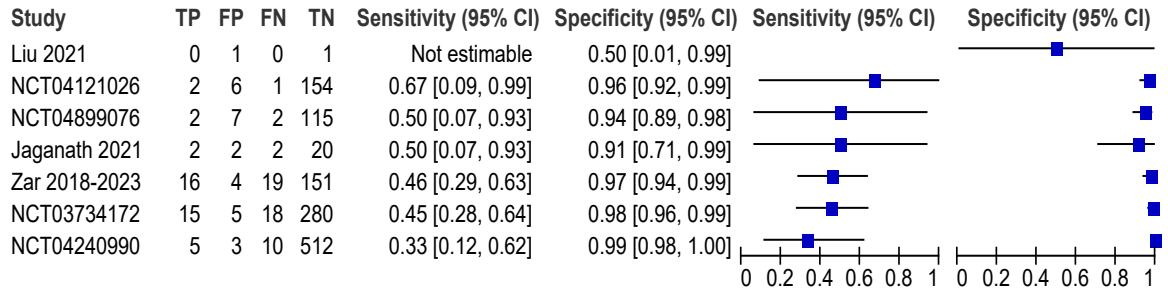
CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted on the plot by decreasing sensitivity and specificity.

Should LC-aNAATs on nasopharyngeal aspirate specimens be used to diagnose pulmonary TB in children with signs and symptoms or who screened positive for pulmonary TB, against an MRS?

Seven studies (1355 participants) were identified, with sensitivities between 33% and 67% and specificities between 50% and 99% (**Fig. 2.3.2.4**). Six studies (*1353*) were included in the meta-analysis. The summary sensitivity was 46.2% (95% CI: 34.9–57.9) and summary specificity was 97.5% (95% CI: 95.1–98.7). Certainty of evidence was moderate for sensitivity and high for specificity. The data on Truenat MTB Plus and MTB-RIF Dx were unavailable.

Fig. 2.3.2.4 Forest plot of LC-aNAAT sensitivity and specificity for detection of pulmonary TB in nasopharyngeal aspirate samples and MRS

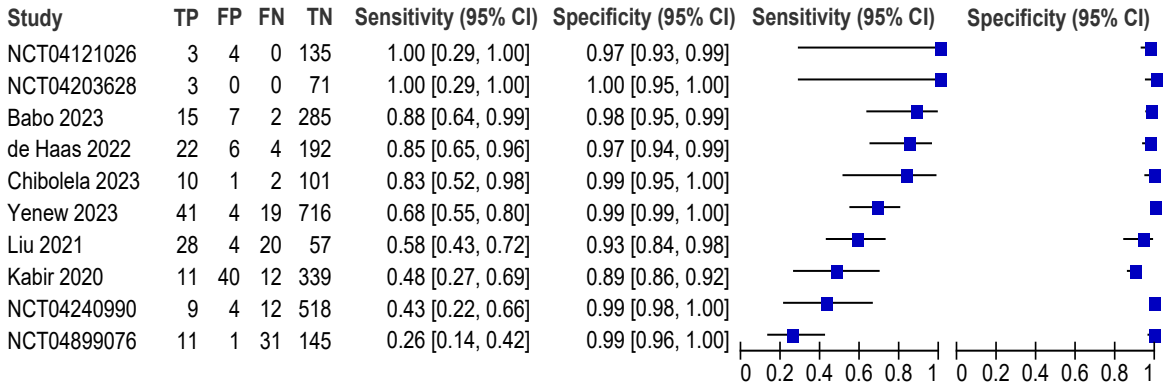


CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test; MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

Should LC-aNAATs on stool be used to diagnose pulmonary TB in children with signs and symptoms or who screened positive for pulmonary TB, against an MRS?

Ten studies (2855 participants) were identified, with sensitivities between 26% and 100% and specificities between 89% and 100% (**Fig. 2.3.2.5**). All 10 studies were included in the meta-analysis. The summary sensitivity was 68.0% (95% CI: 50.3–81.7) and summary specificity was 98.2% (95% CI: 96.3 to 99.1). Certainty of evidence was moderate for sensitivity and high for specificity. The data on Truenat MTB Plus and MTB-RIF Dx were unavailable.

Fig. 2.3.2.5 Forest plot of LC-aNAAT sensitivity and specificity for detection of pulmonary TB in stool and MRS[a]



CI: confidence interval; FN: false negative; FP: false positive; LC-aNAAT: low-complexity automated nucleic acid amplification test: MRS: microbiological reference standard; TB: tuberculosis; TN: true negative; TP: true positive.

[a] Studies are sorted on the plot by decreasing sensitivity.

Cost–effectiveness analysis

As part of the preparatory process for the GDG meeting in May 2024, WHO commissioned a modelled study to assess the cost–effectiveness of using LC-aNAATs (including Xpert Ultra, Truenat and other novel LC-aNAATs in the development pipeline) for the detection of TB when used concurrently among people living with HIV and children, including children with HIV, across two different country settings (Malawi and the Philippines).

A study objective was to assess the cost–effectiveness of concurrent use of LC-aNAATs on respiratory and stool samples for TB diagnosis and rifampicin-resistance detection among children (aged <10 years) with presumptive TB and without HIV infection, compared with a single LC-aNAAT on a respiratory sample alone.

In this hypothetical model, a cohort of children with presumptive TB progressed through a decision analytical framework. In the intervention arm, TB diagnosis involved the concurrent use of LC-aNAATs on both respiratory and stool samples, whereas the comparator arm solely used LC-aNAATs on respiratory specimens. The probability of being able to provide a respiratory sample was considered, and testing was conducted, either for both respiratory and stool samples concurrently or solely for stool. In both the intervention and comparator arms, participants not diagnosed through the diagnostic strategy had the opportunity for clinical diagnosis. Children with bacteriologically confirmed TB underwent DST for rifampicin and began either drug-susceptible TB or DR-TB treatment, depending on the DST result. All individuals were followed over time, including those with false negative or false positive diagnostic results, to account for unnecessary treatment or additional mortality due to missed diagnoses.

When using the high TB burden setting of Malawi to parametrize the model, cost–effectiveness modelling found that the use of an LC-aNAAT on a respiratory sample resulted in an average cost of US$ 144, with a corresponding average DALY of 0.93. In contrast, the concurrent use of LC-aNAATs on respiratory and stool samples yielded an average cost of US$ 204, and a DALY of 0.57, resulting in an incremental cost per DALY averted of US$ 253 (95% UR: 123–2317) (**Table 2.3.2.1**).

Similarly, in the Philippines, the cost of an LC-aNAAT on a respiratory sample was US$ 84, associated with a DALY of 1.04. Concurrent testing in the Philippines resulted in an average cost of US$ 149 and a DALY of 0.66, with an ICER of US$ 156 per DALY averted (95% UR: 79–888) (**Table 2.3.2.1**).

Table 2.3.2.1 Cost–effectiveness analysis of concurrent use of LC-aNAATs among children in Malawi and the Philippines

Country	Diagnostic strategy	Cost, US$	Effectiveness, DALYs	ICER (95% UR), US$
Malawi	LC-aNAAT on respiratory sample	114	0.93	Reference
	LC-aNAATs on respiratory and stool samples	204	0.57	253 (123–2317)
Philippines	LC-aNAAT on respiratory sample	84	1.04	Reference
	LC-aNAATs on respiratory and stool samples	149	0.62	156 (79–888)

DALY: disability-adjusted life year; ICER: incremental cost–effectiveness ratio; LC-aNAAT: low-complexity automated nucleic acid amplification test; UR: uncertainty range.

More information on the cost–effectiveness analysis of concurrent use of tests in children is available in **Web Annex B.9.**

User perspective

The GDG assessed whether concurrent testing of multiple samples would increase the diagnostic yield (i.e. the benefit to patients or the programme in terms of finding more people with TB). One of the PICO questions focused on the concurrent use of LC-aNAATs on respiratory and stool samples for the diagnosis of TB in children.

User preferences and values

As important outcomes of the diagnostic test, people in high TB burden settings value:

- getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me");
- avoiding diagnostic delays, as they exacerbate existing financial hardships and emotional and physical suffering and make people feel guilty for infecting others (especially children);
- having accessible facilities; and
- reducing diagnosis-associated costs (e.g. travel, missing work).

Participants appreciate that stool collection is far less invasive than gastric lavage and can thereby reduce physical and emotional suffering of children and their parents (see **Web Annex B.10**).

Equity

Concurrent specimen testing was not practiced in the interview study countries. However, it was believed to improve access to care by minimizing repeat visits and loss to follow-up.

According to the interview study respondents, using non-sputum specimens has the potential to improve access to care, especially with a test that can be performed at all levels of the health

care system. Challenges with producing a sufficient quality and quantity of sputum are well documented and can lead to repeat testing or false results.

Acceptability

Most participants, including health workers and caregivers, did not immediately understand why multiple samples would be tested concurrently at the same visit, if a respiratory sample is available. They highlighted that a sputum sample is the preferred choice, and they would only collect the second-best sample if that were not available. However, participants also thought that concurrent sample testing could be possible if there was a WHO recommendation, altered diagnostic algorithms and specific training and capacity strengthening to facilitate it (see **Web Annex B.10**).

For young children, stool seems to be an acceptable specimen, especially after adequate training in how to process it. Stool from adults is considered more difficult in terms of both acceptance and processing time. In general, participants had confidence in the results from stool tested by GeneXpert (see **Web Annex B.10**).

Feasibility

Important feasibility challenges are related to the deteriorating quality of stool, caused by delays between time of collection and time of processing in the laboratory (see **Web Annex B.10**).

Concurrent testing needs to be framed as a more efficient way of working (i.e. testing two samples concurrently during the same visit, instead of testing one sample during each of two separate visits) that also allows increasing access and reducing costs for patients. The practice of concurrent testing needs to be framed as generating sufficient benefit to justify the additional short-term workload and having the potential to reduce the workload in the longer term. Without such framing, there is a risk that already overburdened health care workers will avoid concurrent testing (see **Web Annex B.10**).

Prior investments made in frontrunner technologies, donor preferences, limited health systems thinking and unnecessary competition between manufacturers all pose challenges to policy adoption and implementation of novel molecular diagnostics. In addition, national in-country health technology and cost-efficacy assessments can delay decisions to implement newer technologies and diagnostic strategies using different samples (see **Web Annex B.10**).

More information on the qualitative evidence analysis and synthesis for concurrent use of tests in children is available from **Web Annex B.10.**

Implementation considerations

- Concurrent testing maximizes diagnostic opportunity and accuracy of case detection, is a more efficient way to address the needs of this population and is preferred even if the testing workload may increase.
- A positive result on either test is sufficient to confirm TB diagnosis.
- Patient loss to follow-up for the second test result should be monitored and prevented. Patients should be provided with information to understand the concurrent testing approach and the need for follow-up.

- Testing capacity should be secured for the second test, as volumes will increase.
- Adequate staffing capacity and training are needed to improve the collection of different sample types and laboratory processing of collected samples.
- Performing the same test on a new sample may need additional regulatory approval on a national and international level.
- Infrastructure and training on how to collect a stool sample privately should be available.
- As with all WHO-recommended TB diagnostics, quality assurance programmes for both sample types are required.
- At a primary health care level, in a situation of sputum paucity or absence, stool and nasopharyngeal aspirate may be feasible, whereas collection of more invasive specimen types (i.e. induced sputum, BAL and gastric aspirate) would require upward referral, depending local capacity and expertise. In these circumstances, performing stool testing at primary health care level and waiting for a test result before upward referral of the child may be appropriate.

Monitoring and evaluation

- Monitor simultaneous specimen collection and turnaround time for the test results in a concurrent testing approach.
- Monitor patient loss to follow-up from a second test in a concurrent testing approach.
- Monitor trends in the rate of indeterminate test results for both sample types with LC-aNAATs.
- Monitor trends in the discordance rate between the respiratory and stool LC-aNAAT results. If these differences vary from other local or regional patterns, or if the trends change, further investigation is required.

Research priorities

- Evaluate the impact of concurrent specimen testing on patient-important outcomes for children (cure, mortality, time to diagnosis and time to start of treatment).
- Evaluate the impact of concurrent specimen testing on affordability and cost–effectiveness in the intended settings of use.
- Evaluate the performance of other LC-aNAATs in concurrent testing approaches.
- Identify an improved reference standard that accurately defines TB disease in children and paucibacillary specimens because the sensitivity of all available diagnostics is suboptimal.
- Develop new tools that correctly diagnose a higher proportion of TB in children. Ideally, the new tools will be rapid, affordable, feasible and acceptable to children and their parents.
- Develop rapid point-of-care diagnostic tests and simpler alternative sample types for paucibacillary and extrapulmonary TB in children.
- Perform operational research to ensure that tests are used optimally in intended settings.
- Develop and apply standardized methods for assessment of costs and cost–effectiveness, to improve comparability and scope of economic evidence.

2.3.3 Concurrent use of tests in children with HIV

NEW

Recommendations

9. For children with HIV who have signs or symptoms or screen positive for pulmonary TB, concurrent testing using low-complexity automated NAATs on respiratory and stool samples and LF-LAM on urine may be used as the initial diagnostic strategy for diagnosing TB rather than low-complexity automated NAATs on respiratory or stool samples alone.

(Conditional recommendation, low certainty of evidence for test accuracy)

Remarks

- This recommendation prioritizes concurrent testing over the use of molecular testing and LF-LAM in isolation for diagnosis of TB in children with HIV.
- Use of LC-aNAATs on isolated specimens was also evaluated. The findings supported the use of LC-aNAATs for initial diagnostic testing for TB in HIV-positive children with signs or symptoms or who screen positive for pulmonary TB, using sputum, gastric aspirate, stool or nasopharyngeal aspirate, rather than smear or culture.
- This recommendation is conditional because the findings indicate moderate undesirable effects (i.e. decreased specificity, resulting in more false positive test results) when compared with a single test strategy.
- The product for which eligible data met the LC-aNAAT class-based performance criteria for this recommendation was Xpert MTB/RIF Ultra. The performance of Truenat MTB Plus and MTB-RIF Dx for this recommendation could not be assessed, as data were unavailable.

Justification and evidence

LC-aNAATs on respiratory and stool sample and LF-LAM on urine are recommended as the first test for symptomatic children with HIV presenting with presumptive TB disease, and should be used to diagnose TB.

Previous systematic reviews have traditionally assessed diagnostic accuracy of LC-aNAATs on two samples and LF-LAM on urine in isolation for the detection of TB in children, but in clinical practice the tests may be used concurrently (i.e. LC-aNAAT on a respiratory and stool sample and LF-LAM on urine) and together they increase sensitivity.

Incremental diagnostic accuracy

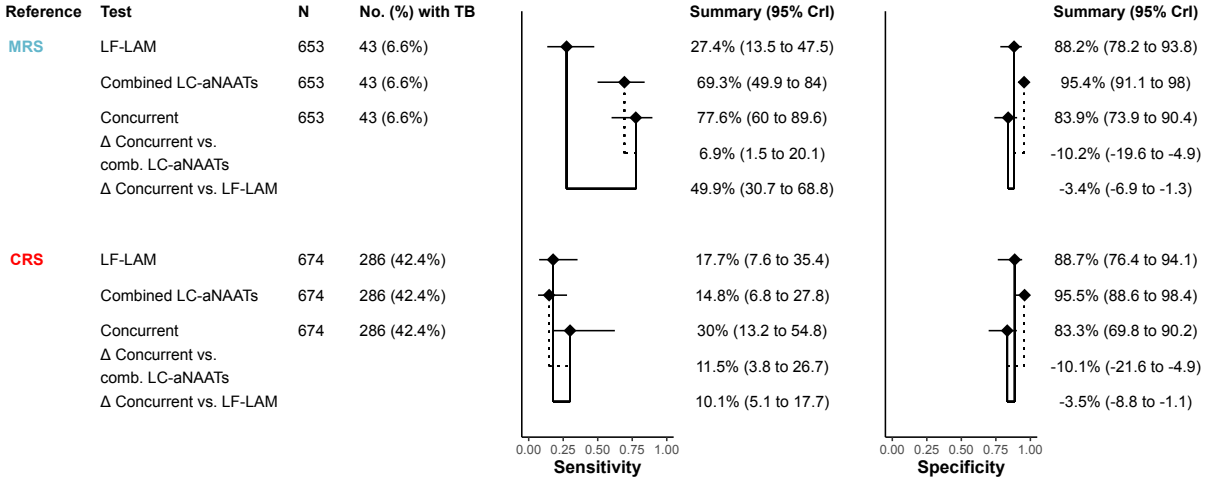
What is the incremental diagnostic accuracy of concurrent use of LC-aNAATs on respiratory and stool samples and LF-LAM on urine versus each sample type alone for diagnosis of pulmonary TB disease in children with HIV, with signs and symptoms or who screened positive for pulmonary TB, compared with any of the tests (either LC-aNAATs combined or LF-LAM) alone?

Based on six studies (653 participants, including 43 [6.6%] with TB) included in the meta-analysis for an MRS, the estimated diagnostic accuracy of the concurrent use of LC-aNAAT on respiratory

samples plus LC-aNAAT on stool and LF-LAM on urine had a pooled sensitivity of 77.8% (95% CrI: 59.9 to 89.8) and a pooled specificity of 83.9% (95% CrI: 73.9 to 90.4) (**Fig. 2.3.7**). Compared with LC-aNAAT on respiratory samples alone, concurrent testing had 6.9 percentage points (95% CrI: 1.5 to 20.1) higher sensitivity and –10.1 percentage points (95% CrI: –21.6 to –4.9) lower specificity. Certainty of evidence was low for specificity and moderate for sensitivity.

Based on six studies (674 participants, including 286 [42.4%] with TB) included in the meta-analysis for a CRS, the estimated diagnostic accuracy of the concurrent use of LC-aNAAT on respiratory samples plus LC-aNAAT on stool and LF-LAM on urine had a pooled sensitivity of 30.1% (95% CrI: 13.2 to 54.9) and a pooled specificity of 83.3% (95% CrI: 69.6 to 90.2) (**Fig. 2.3.3.1**). Compared with LC-aNAAT on respiratory samples alone, concurrent testing had 14.9 percentage points (95% CrI: 0 to 41.1) higher sensitivity and –12.0 percentage points (95% CrI: –27.0 to –2.6) lower specificity. Certainty of evidence was very low for sensitivity and low for specificity.

Fig. 2.3.3.1 Forest plot of pooled sensitivity and specificity for all studies, by each index test



CrI: credible interval; CRS: composite reference standard; LC-aNAAT: low-complexity automated nucleic acid amplification test; LF-LAM: lateral flow urine lipoarabinomannan assay; MRS: microbiological reference standard; TB: tuberculosis.

[a] The diamonds represent pooled sensitivity and specificity, and the black horizontal line its 95% CrI. The difference in accuracy between index tests is indicated by solid lines (concurrent versus stool) or dotted lines (concurrent versus respiratory) connecting the diamonds.

Cost–effectiveness analysis

In addition to the economic evidence regarding concurrent use of tests in people living with HIV and children (see Sections 2.3.1 and 2.3.2), WHO commissioned a third study that aimed to assess the cost–effectiveness of using LC-aNAATs (including Xpert Ultra, Truenat and other novel LC-aNAATs in the development pipeline) for the detection of TB when used concurrently among children with HIV, across two different country settings (Malawi and the Philippines).

An objective of this study was to assess the cost–effectiveness of concurrent use of LC-aNAATs on respiratory and stool samples and LF-LAM on urine for TB diagnosis and rifampicin-resistance

detection among children (aged <10 years) living with HIV and with presumptive TB, compared with a single LC-aNAAT on a respiratory sample alone.

In the hypothetical model that informed this study, a cohort of children with HIV and with signs and symptoms of TB progressed through a decision analytical framework. In the intervention arm, TB diagnosis involved the concurrent use of LC-aNAATs on both respiratory and stool samples, alongside LF-LAM on urine. The comparator arm used LC-aNAAT on respiratory samples alone. The probability of providing a respiratory sample was considered, and testing was conducted, either concurrently on respiratory and stool samples alongside LF-LAM, or on stool alone alongside LF-LAM. In both the intervention and comparator arm, participants not diagnosed through the diagnostic strategy had the opportunity for clinical diagnosis. Children with bacteriologically confirmed TB underwent DST for rifampicin and began either drug-susceptible TB or DR-TB treatment, depending on the DST result. All individuals were followed over time, including those with false negative or false positive diagnostic results, to account for unnecessary treatment or additional mortality due to missed diagnoses.

The study findings shown in **Table 2.3.3.1** show the cost–effectiveness of the concurrent use of LC-aNAATs on respiratory and stool samples and LF-LAM on urine among children with HIV in Malawi and the Philippines. In Malawi, the average cost of implementing an LC-aNAAT on a respiratory sample was US$ 319, with a corresponding average DALY of 5.08. When used concurrently, the average cost increased to US$ 460, while the average DALY decreased to 1.8. The resulting ICER per DALY averted was US$ 43 (95% UR: 28–89). Similarly, in the Philippines, implementation of an LC-aNAAT on a respiratory sample alone cost US$ 249, with an average DALY of 5.13, whereas concurrent use incurred an average cost of US$ 345 and an average DALY of 1.77. The ICER per DALY averted was US$ 29 (95% UR: 18–63).

Table 2.3.3.1 Cost–effectiveness analysis of concurrent use of LC-aNAATs among children living with HIV in Malawi and the Philippines

Country	Diagnostic strategy	Cost, US$	Effectiveness, DALYs	ICER (95% UR), US$
Malawi	LC-aNAAT on respiratory sample	320	5.08	Reference
	LC-aNAAT on respiratory and stool samples and LF-LAM	460	1.8	43 (28–89)
Philippines	LC-aNAAT on respiratory sample	249	5.13	Reference
	LC-aNAAT on respiratory and stool samples and LF-LAM	345	1.77	29 (18–63)

DALY: disability-adjusted life year; HIV: human immunodeficiency virus; ICER: incremental cost–effectiveness ratio; LC-aNAAT: low-complexity automated nucleic acid amplification test; LF-LAM: lateral flow urine lipoarabinomannan assay; UR: uncertainty range.

More information on the cost–effectiveness analysis of concurrent use of tests in children with HIV is available in **Web Annex B.9.**

User perspective

The GDG assessed whether concurrent testing of multiple samples would increase the diagnostic yield (i.e. the benefit to patients or the programme in terms of finding more people with TB). One of the PICO questions focused on concurrent use of LC-aNAATs on respiratory and stool samples and LF-LAM on urine for the diagnosis of TB in children living with HIV.

User preferences and values

The interview study and quality evidence synthesis produced no data on the use of LF-LAM in children living with HIV. However, in general, as important outcomes of the diagnostic test, patients in high TB burden settings value:

- getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me");
- avoiding diagnostic delays, as they exacerbate existing financial hardships and emotional and physical suffering and make people feel guilty for infecting others (especially children);
- having accessible facilities; and
- reducing diagnosis-associated costs (e.g. travel, missing work).

Participants appreciate that stool sample collection is far less invasive than gastric aspirate (see **Web Annex B.10**).

Equity

Concurrent sample testing was not practiced in the study countries. However, concurrent sample testing could improve access to care by minimizing repeat visits and loss to follow-up (see **Web Annex B.10**).

Using non-sputum samples can improve access to care, especially with a test that can be performed at all levels of the health carecare system. Challenges with producing sputum of sufficient quality and quantity are well documented and can lead to repeat testing or false results. Participants highlighted the impact that using stool has on increasing case-finding and access to care, particularly among destitute families (see **Web Annex B.10**).

Acceptability

The interview study produced no data on the use of LF-LAM in children living with HIV.

Most participants (including parents and legal representatives of children) did not immediately understand why multiple samples would be tested concurrently at the same visit, if a respiratory sample is available. They highlighted that a sputum sample is the preferred choice, and they would only collect the second-best sample if that were not available. However, participants also thought that concurrent sample testing could be possible if there was a WHO recommendation, altered diagnostic algorithms and specific training and capacity strengthening to facilitate it (see **Web Annex B.10**).

For young children, stool seems to be an acceptable specimen, especially after adequate training in how to process it. Stool from adults is considered more difficult in terms of both acceptance and processing time. There was general confidence among participants regarding results from stool tested by GeneXpert (see **Web Annex B.10**).

Feasibility

The interview study produced no data on the use of LF-LAM in children living with HIV. In younger and sicker children, urine sample collection is more cumbersome, as it requires both the child's and the caregiver's cooperation, and it may be affected by medical issues, such as dehydration (see **Web Annex B.10**).

Important feasibility challenges are related to the deteriorating quality of the stool sample caused by delays between time of collection and time of processing in the laboratory (see **Web Annex B.10**).

Concurrent testing needs to be framed as a more efficient way of working (i.e. testing two samples concurrently during the same visit, instead of testing one sample during each of two separate visits) that also allows increasing access and reducing costs for patients. According to a laboratory manager, this framing of the benefits outweighing the additional workload, and potentially resulting in reduced work in the long run, will be critical to avoid concurrent testing being perceived as additional work for already overburdened health care workers (see **Web Annex B.10**).

Prior investments made in frontrunner technologies, donor preferences, limited health systems thinking and unnecessary competition between manufacturers all pose challenges to policy adoption and implementation of novel molecular diagnostics. In addition, national in-country health technology and cost-efficacy assessments can delay decisions to implement newer technologies and diagnostic strategies using different samples (see **Web Annex B.10**).

Implementation considerations

- The implementation considerations are the same as those in Sections 2.3.1 and 2.3.2.

Monitoring and evaluation

- The monitoring and evaluation considerations are the same as those in Sections 2.3.1 and 2.3.2.

Research priorities

- The research priorities are the same as those in Sections 2.3.1 and 2.3.2.

2.4. Follow-on diagnostic tests for detection of additional drug-resistance after TB confirmation

2.4.1 Low complexity automated NAATs for detection of resistance to isoniazid and second-line anti-TB agents

Among 105 countries possessing representative data on resistance to fluoroquinolones from the past 15 years, the proportion of MDR/RR-TB cases with resistance to any fluoroquinolone for which testing was done was 20.1% (95% CI: 15.5–25.0%) *(1)*. Thus, rapid and early testing for the detection of fluoroquinolone resistance is essential for determining eligibility for treatment with the all-oral 9–12 month standardized shorter regimen for MDR/RR-TB. However, the current limitation with testing for fluoroquinolone resistance is the limited accessibility of current technologies (which are often only available at higher tiers of the health system) and poor yield in paucibacillary specimens.

Low complexity automated NAATs are a new class of diagnostics intended for use as a reflex test in specimens determined to be Mtb complex (MTBC)-positive; they offer rapid DST in intermediate and peripheral laboratories. The first product in this class simultaneously detects resistance to isoniazid, fluoroquinolones, ethionamide and amikacin. Results are available in under 90 minutes, leading to faster time to results than the current standard of care, which includes LPAs and culture-based phenotypic DST.

An additional value of the tests is the accurate and rapid detection of isoniazid resistance, which is relevant for both RR-TB and rifampicin-susceptible TB; the latter is often undiagnosed and contributes to a large burden of disease. Globally, rifampicin-susceptible TB is estimated to occur in 13.1% (95% CI: 9.9–16.9%) of new cases and 17.4% (95% CI: 0.5–54.0%) of previously treated cases *(1)*. Thus, this test could also be used as a reflex test to complement existing technologies that only test for rifampicin, allowing the rapid and accurate detection of isoniazid-resistant, rifampicin-susceptible TB.

Although these new technologies are excellent at detecting resistance to selected drugs, conventional culture-based phenotypic DST remains important to determine resistance to other anti-TB agents, particularly the new and repurposed medicines such as bedaquiline and linezolid.

Recommendations

10. In people with bacteriologically confirmed pulmonary TB, low complexity automated NAATs may be used on sputum for the initial detection of resistance to isoniazid and fluoroquinolones, rather than culture-based phenotypic DST.

(Conditional recommendation, moderate certainty of evidence for diagnostic accuracy)

11. In people with bacteriologically confirmed pulmonary TB and resistance to rifampicin, low complexity automated NAATs may be used on sputum for the initial detection of resistance to ethionamide, rather than DNA sequencing of the *inhA* promoter.

(Conditional recommendation, very low certainty of evidence for diagnostic accuracy)

12. In people with bacteriologically confirmed pulmonary TB and resistance to rifampicin, low complexity automated NAATs may be used on sputum for the initial detection of resistance to amikacin, rather than culture-based phenotypic DST.

(Conditional recommendation, low certainty of evidence for diagnostic accuracy)

There are several subgroups to be considered for these recommendations:

- The recommendations are based on the evidence of diagnostic accuracy in sputum of adults with bacteriologically confirmed pulmonary TB, with or without rifampicin resistance.
- The recommendations are extrapolated to adolescents and children, based on the generalization of data from adults.
- The recommendations apply to people living with HIV (studies included a varying proportion of such individuals); data stratified by HIV status were not available.
- The recommendations are extrapolated to people with extrapulmonary TB, and testing of non-sputum samples was considered appropriate, which affects the certainty. The panel did not evaluate test accuracy in non-sputum samples directly, including in children; however, extrapolation was considered appropriate given that WHO has recommendations for similar technologies for use on non-sputum samples (e.g. Xpert MTB/RIF and Xpert Ultra).
- Recommendations for detection of resistance to amikacin and ethionamide are only relevant for people who have bacteriologically confirmed pulmonary TB and resistance to rifampicin.

Justification and evidence

The WHO Global TB Programme initiated an update of the current guidelines and commissioned a systematic review on the use of low complexity automated NAATs for the detection of resistance to isoniazid and second-line TB drugs in people with signs and symptoms of TB.

The PICO questions were designed to form the basis for the evidence search, retrieval and analysis:

1. Should low complexity automated NAATs be used on sputum in people with signs and symptoms of pulmonary TB, irrespective of resistance to rifampicin, for detection of resistance to isoniazid, as compared with culture-based phenotypic DST?

2. Should low complexity automated NAATs be used on sputum in people with signs and symptoms of pulmonary TB, irrespective of resistance to rifampicin, for detection of resistance to fluoroquinolones, as compared with culture-based phenotypic DST?

3. Should low complexity automated NAATs be used on culture isolates in people with signs and symptoms of pulmonary TB, and detected resistance to rifampicin, for detection of resistance to ethionamide, as compared with genotypic sequencing of the *inh*A promoter?

4. Should low complexity automated NAATs be used on sputum in people with signs and symptoms of pulmonary TB, and detected resistance to rifampicin, for detection of resistance to amikacin, as compared with culture-based phenotypic DST?

The databases Ovid Medline (Ovid, 1946 to present) and Embase (Ovid, 1947 to present) were searched for studies evaluating cartridge-based tests using the following search terms: tuberculosis, pulmonary AND Xpert, GeneXpert, Truenat, cartridge, point-of-care systems, drug susceptibility test, isoniazid resistance, fluoroquinolone resistance and second-line injectable drug resistance. Clinicaltrials.gov and the WHO International Clinical Trials Registry Platform were also searched for trials in progress. Searches were run up to 6 September 2020 without language restriction. On 4 November 2020, an additional search was run using the search terms Zeesan and MeltPro.

Researchers at FIND, the WHO Global TB Programme, the manufacturer and other experts in the field of TB diagnostics were contacted for information about ongoing and unpublished studies. Data submitted in response to the WHO public call were reviewed.

Drug resistance was compared against a phenotypic reference standard (or a genotypic reference standard for ethionamide resistance), as well as a composite reference standard that was constructed by combining the results of phenotypic and genotypic DST results in studies where both had been performed.

Data synthesis was structured around the four preset PICO questions, as outlined below. Three web annexes give additional information, as follows:

- details of studies included in the current analysis (**Web Annex A.6:** *Low complexity automated NAATs*);
- a summary of the results and details of the evidence quality assessment (**Web Annex A.6:** *Low complexity automated NAATs*); and
- a summary of the GDG panel judgements (**Web Annex A.6:** *Low complexity automated NAATs*).

PICO 1: Should low complexity automated NAATs be used on sputum in people with signs and symptoms of pulmonary TB, irrespective of resistance to rifampicin, for detection of resistance to isoniazid, as compared with culture-based phenotypic DST?

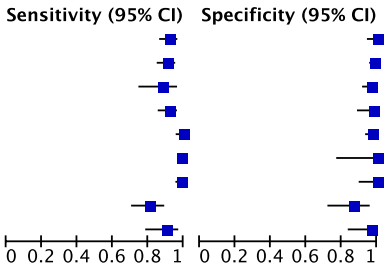
Three multinational studies with 1605 participants provided data for evaluating isoniazid resistance detection. The reference standard for each of these studies was culture-based phenotypic DST. Each study centre in the multinational studies was analysed as a separate study (**Fig. 2.4.1.1**).

Several concerns were expressed about indirectness in the study populations. First, the median prevalence of isoniazid resistance in the included studies was 67.2% (range, 26.8% [Diagnostics for Multidrug Resistant Tuberculosis in Africa – DIAMA, Benin] to 93.9% [FIND, Moldova]), which is higher than the global estimates for isoniazid resistance. Hence, applicability to settings

with a lower prevalence of isoniazid resistance comes with some uncertainty. Second, there are potential differences in the mutations present in isoniazid monoresistant strains and MDR strains; that is, some studies suggest that the mutations found in monoresistant strains are more diverse than the mutations found in MDR strains. Third, although the population for this PICO question is "irrespective of rifampicin resistance", enrolment criteria in the studies meant that most participants within the included studies had RR-TB. As a result of these concerns, certainty of evidence was downgraded one level for indirectness both for sensitivity and specificity, and the quality (certainty) of evidence was rated moderate both for sensitivity and specificity.

Fig. 2.4.1.1 Forest plot of included studies for isoniazid resistance detection, irrespective of rifampicin resistance with culture-based phenotypic DST as the reference standard

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Cepheid China 2020	117	0	10	60	0.92 [0.86, 0.96]	1.00 [0.94, 1.00]
Cepheid South Africa 2020	127	2	13	149	0.91 [0.85, 0.95]	0.99 [0.95, 1.00]
DIAMA Benin 2020	37	4	5	111	0.88 [0.74, 0.96]	0.97 [0.91, 0.99]
DIAMA Cameroon 2020	110	1	10	45	0.92 [0.85, 0.96]	0.98 [0.88, 1.00]
DIAMA Rwanda 2020	76	3	0	120	1.00 [0.95, 1.00]	0.98 [0.93, 0.99]
FIND Moldova 2020	213	0	3	14	0.99 [0.96, 1.00]	1.00 [0.77, 1.00]
FIND Mumbai 2020	143	0	2	33	0.99 [0.95, 1.00]	1.00 [0.89, 1.00]
FIND New Delhi 2020	63	5	15	33	0.81 [0.70, 0.89]	0.87 [0.72, 0.96]
FIND South Africa 2020	45	1	5	30	0.90 [0.78, 0.97]	0.97 [0.83, 1.00]



CI: confidence interval; DIAMA: Diagnostics for Multidrug Resistant Tuberculosis in Africa; DST: drug susceptibility testing; FIND: Foundation for Innovative New Diagnostics; FN: false negative; FP: false positive; TB: tuberculosis; TN: true negative; TP: true positive.

The sensitivity in these three studies ranged from 81% to 100% and the specificity from 87% to 100%. **The pooled sensitivity was 94.2% (95% CI: 89.3–97.0%) and the pooled specificity was 98.0% (95% CI: 95.2–99.2%).**

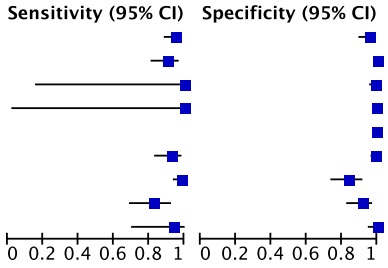
PICO 2: Should low complexity automated NAATs be used on sputum in people with signs and symptoms of pulmonary TB, irrespective of resistance to rifampicin, for detection of resistance to fluoroquinolones, as compared with culture-based phenotypic DST?

Three multinational studies with 1337 participants provided data for evaluation of detection of fluoroquinolone resistance. The reference standard for each of these studies was culture-based phenotypic DST. Each study centre in the multinational studies was analysed as a separate study (**Fig. 2.4.1.2**).

Specificity estimates were inconsistent, at 84% (FIND, Mumbai), 91% (FIND, New Delhi) and more than 96% for other studies. The heterogeneity in specificity estimates could not be explained. Consequently, the certainty of the evidence was downgraded one level for inconsistency; the quality (certainty) of the evidence was rated high for sensitivity and moderate for specificity.

Fig. 2.4.1.2 Forest plot of included studies for fluoroquinolone resistance detection, irrespective of rifampicin resistance with culture-based phenotypic DST as the reference standard

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Cepheid China 2020	90	4	5	87	0.95 [0.88, 0.98]	0.96 [0.89, 0.99]
Cepheid South Africa 2020	58	0	6	167	0.91 [0.81, 0.96]	1.00 [0.98, 1.00]
DIAMA Benin 2020	2	2	0	144	1.00 [0.16, 1.00]	0.99 [0.95, 1.00]
DIAMA Cameroon 2020	1	1	0	166	1.00 [0.03, 1.00]	0.99 [0.97, 1.00]
DIAMA Rwanda 2020	0	1	0	186	Not estimable	0.99 [0.97, 1.00]
FIND Moldova 2020	52	2	4	172	0.93 [0.83, 0.98]	0.99 [0.96, 1.00]
FIND Mumbai 2020	102	12	2	62	0.98 [0.93, 1.00]	0.84 [0.73, 0.91]
FIND New Delhi 2020	38	6	8	64	0.83 [0.69, 0.92]	0.91 [0.82, 0.97]
FIND South Africa 2020	15	0	1	64	0.94 [0.70, 1.00]	1.00 [0.94, 1.00]



CI: confidence interval; DIAMA: Diagnostics for Multidrug Resistant Tuberculosis in Africa; DST: drug susceptibility testing; FIND: Foundation for Innovative New Diagnostics; FN: false negative; FP: false positive; TB: tuberculosis; TN: true negative; TP: true positive.

The sensitivity for fluoroquinolone resistance in these three studies ranged from 83% to 100% and the specificity from 84% to 100%. **The pooled sensitivity was 93.1% (95% CI: 88.0–96.1%) and the pooled specificity was 98.3% (95% CI: 94.5–99.5%).**

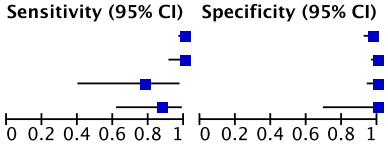
PICO 3: Should low complexity automated NAATs be used on culture isolates in people with signs and symptoms of pulmonary TB, and detected resistance to rifampicin, for detection of resistance to ethionamide, as compared with genotypic sequencing of the *inhA* promoter?

One multinational study with 434 participants provided data for evaluating resistance to ethionamide. The reference standard for this study was DNA sequencing of the *inhA* promoter. Each study centre in the multinational study was analysed as a separate study (**Fig. 2.4.1.3**).

The study was judged to be at very serious risk of bias in the reference standard domain because it did not include all loci (i.e. *ethA*, *ethR* and *inhA* promoter) required for the reference standard to classify the target condition correctly. Against a reference standard of phenotypic DST, the pooled sensitivity was considerably lower, at 51.7% (95% CI: 33.1–69.8%). Consequently, certainty of evidence was downgraded two levels for risk of bias for both sensitivity and specificity. In addition, the 95% CIs were wide for both sensitivity and specificity, which could lead to different decisions, depending on which confidence limits are assumed. Consequently, the certainty of the evidence was downgraded one level for imprecision for both sensitivity and specificity; the quality (certainty) of evidence was rated very low for both sensitivity and specificity.

Fig. 2.4.1.3 Forest plot of included studies for ethionamide resistance detection with genotypic DST as the reference standard

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
FIND Moldova 2020	103	3	0	103	1.00 [0.96, 1.00]	0.97 [0.92, 0.99]
FIND Mumbai 2020	39	0	0	94	1.00 [0.91, 1.00]	1.00 [0.96, 1.00]
FIND New Delhi 2020	7	0	2	57	0.78 [0.40, 0.97]	1.00 [0.94, 1.00]
FIND South Africa 2020	14	0	2	10	0.88 [0.62, 0.98]	1.00 [0.69, 1.00]



CI: confidence interval; DST: drug susceptibility testing; FIND: Foundation for Innovative New Diagnostics; FN: false negative; FP: false positive; TB: tuberculosis; TN: true negative; TP: true positive.

The sensitivity for ethionamide resistance in this study ranged from 78% to 100% and the specificity from 97% to 100%. **The pooled sensitivity was 98.0% (95% CI: 74.2–99.9%) and the pooled specificity was 99.7% (95% CI: 83.5–100.0%).**

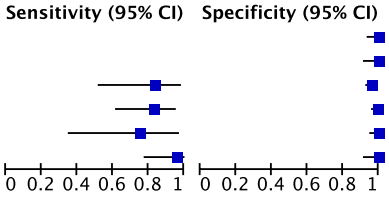
PICO 4: Should low complexity automated NAATs be used on sputum in people with signs and symptoms of pulmonary TB, and detected resistance to rifampicin, for detection of resistance to amikacin, as compared with culture-based phenotypic DST?

One multinational study with 490 participants provided data for evaluating resistance to amikacin. The reference standard for this study was culture-based phenotypic DST. Each study centre in this multinational study was analysed as a separate study (**Fig. 2.4.1.4**).

The 95% CI for sensitivity was wide, which could lead to different decisions around true positives and false negatives, depending on which confidence limits are assumed. Also, there were few participants with amikacin resistance contributing to this analysis for the observed sensitivity. Consequently, the certainty of the evidence was downgraded two levels for imprecision. Also, there were few participants with amikacin resistance contributing to this analysis for the observed sensitivity. Consequently, the certainty of the evidence was downgraded two levels for imprecision; the quality (certainty) of evidence was rated low for sensitivity and high for specificity.

Fig. 2.4.1.4 Forest plot of included studies for amikacin resistance detection with culture-based phenotypic DST as the reference standard

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
DIAMA Benin 2020	0	0	0	51	Not estimable	1.00 [0.93, 1.00]
DIAMA Cameroon 2020	0	0	0	39	Not estimable	1.00 [0.91, 1.00]
FIND Moldova 2020	10	8	2	192	0.83 [0.52, 0.98]	0.96 [0.92, 0.98]
FIND Mumbai 2020	19	1	4	120	0.83 [0.61, 0.95]	0.99 [0.95, 1.00]
FIND New Delhi 2020	6	0	2	65	0.75 [0.35, 0.97]	1.00 [0.94, 1.00]
FIND South Africa 2020	21	0	1	39	0.95 [0.77, 1.00]	1.00 [0.91, 1.00]



CI: confidence interval; DIAMA: Diagnostics for Multidrug Resistant Tuberculosis in Africa; DST: drug susceptibility testing; FIND: Foundation for Innovative New Diagnostics; FN: false negative; FP: false positive; TB: tuberculosis; TN: true negative; TP: true positive.

The sensitivity for amikacin resistance in this study ranged from 75% to 95% and the specificity from 96% to 100%. **The pooled sensitivity was 86.1% (95% CI: 75.0–92.7%) and the pooled specificity was 98.9% (95% CI: 93.0–99.8%).**

Cost–effectiveness analysis

This section answers the following additional question:

What is the comparative cost, affordability and cost–effectiveness of implementation of low complexity automated NAATs?

A systematic review was conducted, focusing on economic evaluations of low complexity automated NAATs. Four online databases (Embase, Medline, Web of Science and Scopus) were searched for new studies published from 1 January 2010 through 17 September 2020. The citations of all eligible articles, guidelines and reviews were reviewed for additional studies. Experts and test manufacturers were also contacted to identify any additional unpublished studies.

The objective of the review was to summarize current economic evidence and further understand the costs, cost–effectiveness and affordability of low complexity automated NAATs.

Two low complexity automated NAATs were identified: the MeltPro MTB/RIF (Xiamen Zeesan Biotech Co Ltd, China) and the Xpert MTB/XDR assay (Cepheid, Sunnyvale, USA). Only data concerning Xpert MTB/XDR are included in this review. As is the case with Xpert MTB/RIF, the novel XDR assay can be used to test either unprocessed or concentrated sputum. No published studies providing direct evidence on the cost or cost–effectiveness of low complexity automated NAATs were identified.

Through direct communication from the Xpert MTB/XDR manufacturer, Cepheid, the low- and middle-income country (LMIC) cost for the XDR cartridge is expected to be US$ 19.80 ex-works. Shipping and customs costs will be additional and will be borne by the ordering nations or organizations, as is currently the case for Xpert MTB/RIF and Ultra cartridges.

As with the Xpert MTB/RIF and Ultra assays, the test cartridge costs represent just one component of the total unit test costs that must be considered, with equipment being another important consideration. The Xpert MTB/XDR test will not work on existing six-colour modules and will require laboratories to upgrade to 10-colour GeneXpert modules. There will be different upgrade options for the 10-colour system, with different price points depending on the needs and resources available. Upgrade options include:

- a new 10-colour system – this is the most costly option, at US$ 9420 for one module to US$ 72 350 for 16 modules, including the GeneXpert platform, computer and scanner;
- a new 10-colour satellite instrument with the GeneXpert connected to an existing system – this costs from US$ 6495 for one module to US$ 69 525 for 16 modules; and
- converting an existing GeneXpert system from a six-colour to a 10-colour system by replacing modules – a 10-colour module kit costs US$ 3860.

Additional cost considerations for Xpert MTB/XDR include additional testing or repeated testing in the case of indeterminate or non-actionable results (indeterminate, non-determinate or invalid). The potential cost burden of this is likely to vary, depending on the proportion of indeterminate test results across settings and the associated re-testing protocols.

No studies that have directly assessed the cost–effectiveness of the Xpert MTB/XDR cartridge were identified. Although extrapolation from other platforms and testing approaches for costing may be appropriate, extrapolation of cost–effectiveness data from Xpert MTB/RIF (Ultra) or other NAATs is not advised because of differences in diagnostic accuracy, costs associated with XDR treatment, and the different testing and treatment cascade of care.

Several factors are likely to influence the cost–effectiveness of Xpert MTB/XDR; they include diagnostic accuracy, which may lead to more or fewer individuals being diagnosed compared with the standard of care (which in turn will vary, depending on the local standard of care). In addition to diagnostic accuracy associated with the test itself, the diagnostic algorithm and placement of the Xpert MTB/XDR test within the algorithm has important implications.

The novel Xpert MTB/XDR provides results in less than 90 minutes. Thus, introduction of this test is likely to result in faster time to a result for genotypic DST and could affect cost–effectiveness by improving the numbers of patients initiating treatment, reducing loss to follow-up and improving survival rates. Costs associated with XDR treatment are likely to be an important driver of cost

and cost–effectiveness because previous work has shown that these costs are high compared to diagnostic and other treatment costs. As larger numbers of XDR-positive individuals requiring treatment are identified, total resources required to treat these individuals will increase.

In the absence of transmission modelling studies, there is no information on the long-term population level impact of introducing Xpert MTB/XDR. Nevertheless, the benefits of identifying more cases earlier could lead to a reduction in ongoing transmission and potential cost-savings over the long term. This requires thorough investigations through transmission modelling.

How large are the resource requirements (costs)?

No published studies provided direct evidence about the total resources required. Resource requirements will include the purchase of cartridges (US$ 19.80/cartridge), upgrading of existing platforms to 10-colour modules (an upgrade that will eventually be required for all Xpert platforms: US$ 3860 to >US$ 72 350) and operational and programmatic costs associated with implementing the novel diagnostic. Resource requirements for XDR treatment (e.g. drugs, hospital capacity and staff) are also likely to increase as the number of people diagnosed increases. Total costs will vary, depending on testing volume and prevalence of XDR in the population; also, the impact on the budget will depend on the current standard of care and associated resource use.

What is the certainty of the evidence of resource requirements (costs)?

Direct costs related to the purchase of cartridges and machinery are provided from the manufacturer; however, several important items related to resource use for implementing Xpert MTB/XDR have not been investigated (e.g. staff time, overhead and operational costs). Differences in resource use between Xpert MTB/XDR and existing approaches will vary across settings using different phenotypic and genotypic DST. There is important variability in costs of staff time and operational costs (e.g. testing volume) across settings.

Does the cost–effectiveness of the intervention favour the intervention or the comparison?

No cost–effectiveness studies using Xpert MTB/XDR were identified. Extrapolation of cost–effectiveness data from Xpert MTB/RIF or other NAATs is not advised because of differences in diagnostic accuracy, and costs associated with XDR treatment and the testing and treatment cascade of care.

More details on economic evidence synthesis and analysis are provided in **Web Annex B.20**: *Systematic literature review of economic evidence for NAATs to detect TB and DR-TB in adults and children*.

User perspective

This section answers the following question about **key informants' views and perspectives on the use of low complexity automated NAATs:**

- Is there important uncertainty about or variability in how much end-users value the main outcomes?
- What would be the impact on health equity?
- Is the intervention acceptable to key stakeholders?
- Is the intervention feasible to implement?

The synthesis and analysis of qualitative evidence on end-users' perspectives are discussed above in the section "User perspective" for moderate complexity automated NAATs: chapter 2.1.2 of the current guidelines.

Findings of the review and interviews

The main findings of the systematic review and interviews are given below. Where information is from the review, a level of confidence in the QES is given; where it is from interviews, this is indicated with 'Interviews'.

Is there important uncertainty about or variability in how much end-users value the main outcomes?

- **Patients** in high burden TB settings value:
  - getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me");
  - avoiding diagnostic delays because they exacerbate existing financial hardships and emotional and physical suffering, and make patients feel guilty for infecting others (especially children);
  - having accessible facilities; and
  - reducing diagnosis-associated costs (e.g. travel, missing work) as important outcomes of the diagnostic.

    QES: moderate confidence
- Low complexity automated NAATs, when compared with existing tests or sputum microscopy, are appreciated by **health care professionals** because of:
  - the rapidity and accuracy of the results;
  - the confidence that a result generates to start treatment and motivate patients;
  - the diversity of sample types;
  - the ability to detect drug resistance earlier or at all, for as many drugs as possible (altering a clinician's risk perception of drug resistance in children), and the consequence of avoiding costlier investigations or hospital stays.

    QES: high confidence
  - Compared with other available diagnostic methods, the cartridge has a quicker turnaround time for first- and second-line DST. Health care professionals value the faster turnaround time, the potential ability to reflex samples from the Xpert MTB/RIF to the Xpert MTB/XDR cartridge, and receiving information on multiple drugs and high-level or low-level resistance simultaneously, because it could enable quicker diagnosis and optimized treatment for patients.

    Interviews
- **Laboratory technicians** appreciate low complexity automated NAATs for the following reasons:
  - Overall, the tests improve laboratory work compared with sputum microscopy in terms of ease of use, ergonomics and biosafety.

    QES: high confidence
  - These tests require minimal user steps, and the GeneXpert platform is a familiar system that people feel comfortable running and interpreting.

    Interviews

- **Laboratory managers** appreciate that monitoring of laboratory work and training is easier than with sputum microscopy, and that use of low complexity automated NAATs eases staff retention because it increases staff satisfaction and is symbolic of progress within the TB world.
  QES: low confidence

What would be the impact on health equity?

The impact on health equity would be similar to that of moderate complexity automated NAATs: Chapter 2.1.2 of the current guidelines.

Is the intervention acceptable to key stakeholders?

The acceptability to key stakeholders is similar to that of moderate complexity automated NAATs: Chapter 2.1.2 of the current guidelines.

The identified challenges in implementing the use of low complexity automated NAATs and accumulated delays at every step may compromise the added value and benefits identified by the users (e.g. avoiding delays, keeping costs low, accurate results, information on drug resistance and easing laboratory work), ultimately leading to use. *QES: high confidence* If these values are not met, it can be assumed that users are less likely to find low complexity automated NAATs acceptable.

Is the intervention feasible to implement?

- Low complexity automated NAATs may decrease the workload in the laboratory in terms of freeing up time for laboratory staff. However, based on experience with Xpert MTB/RIF (Ultra), the introduction of a new class of technologies may increase the workload of laboratory staff if added onto existing work without adjusting staffing arrangements or if the new technology does not replace existing diagnostic tests.
  QES: moderate confidence
- Low complexity automated NAATs require less user training than other DST methods (e.g. LPA and culture), making these tests more feasible to implement than methods with more user steps and those that require significant additional training.
  Interview study
  Implementation of new diagnostics must be accompanied by training for clinicians, to help them interpret results from new molecular tests and understand how this relates to the treatment of a patient. In the past, with the introduction of Xpert MTB/RIF (Ultra), this has been a challenge and has led to underuse.
  QES: high confidence and interview study
  Introduction of Xpert MTB/RIF (Ultra) has also led to overreliance on results of cartridge-based NAATs at the expense of clinical acumen.
  QES: moderate confidence
- Introduction of new diagnostics must also be accompanied by guidelines and algorithms that support clinicians and laboratories in communicating with each other; for example, these resources allow clinicians and laboratories to discuss discordant results, and interpret laboratory results in the context of drug availability, patient history and patient progress on a current drug regimen.
  Interviews

- An efficient sample transportation system, with sustainable funding mechanisms, is crucial for feasibility, especially if an algorithm requires multiple samples at different times from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, there is a risk that the sample may become contaminated and yield inconclusive results on molecular diagnostics. Participants cited good personnel skills, standardized operating procedures and significant laboratory infrastructure as essential in reducing sample contamination in their laboratory.
  Interviews
- The feasibility of low complexity automated NAATs is challenged if there is an accumulation of diagnostic delays or underuse (or both) at every step in the process, mainly because of health system factors:
  - non-adherence to testing algorithms, testing for TB or MDR-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor or delayed sample transport and sample quality, poor or delayed communication of results, delays in scheduling follow-up visits and recalling patients, and inconsistent recording of results;
  - lack of sufficient resources and maintenance (e.g. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners and sputum containers; dusty environment; and delayed or absent local repair option);
  - inefficient or unclear workflows and patient flows (e.g. inefficient organizational processes, poor links between providers, and unclear follow-up mechanisms or information on where patients need to go); and
  - lack of data-driven and inclusive national implementation processes.
    QES: high confidence
- The feasibility of using low complexity automated NAATs is also challenged by the value of diagnosing MTB over DR-TB at primary care. This situation makes the NAAT less feasible as a baseline test, although it would fit at a district or intermediate level laboratory.

Implementation considerations

Factors to consider when implementing low complexity automated NAATs for detection of resistance to isoniazid and second-line anti-TB agents are as follows:

- local epidemiological data on resistance prevalence should guide local testing algorithms, whereas pretest probability is important for the clinical interpretation of test results;
- the cost of a test varies depending on parameters such as the number of samples in a batch and the staff time required; therefore, a local costing exercise should be performed;
- low, moderate and high complexity tests have successive increase in technical competency needs (qualifications and skills) and staff time, which affects planning and budgeting;
- availability and timeliness of local support services and maintenance should be considered when selecting a provider;
- laboratory accreditation and compliance with a robust quality management system (including appropriate quality control) are essential for sustained service excellence and trust;
- training of both laboratory and clinical staff will ensure effective delivery of services and clinical impact;
- use of connectivity solutions for communication of results is encouraged, to improve efficiency of service delivery and time to treatment initiation;

- rapid and early testing for the detection of fluoroquinolone resistance is essential before starting treatment with the all-oral MDR/RR-TB shorter regimen (i.e. 6–9 months); this may also become relevant (depending on the epidemiological context) if new shorter drug-susceptible TB regimens that include fluoroquinolones are introduced;
- these tests can be used to rule in ethionamide resistance, but not to rule out resistance, because mutations conferring resistance to ethionamide are not limited to the *inh*A promoter region – they also include *eth*A, *eth*R and other genes;
- culture-based phenotypic DST may still be required, particularly among those with a high pretest probability of resistance when the low complexity automated NAATs does not detect drug resistance; in addition, culture-based phenotypic DST:
  - remains important to determine resistance to other anti-TB agents, particularly the new and repurposed medicines, and to monitor the emergence of additional drug resistance;
  - does not apply to ethionamide because it is unreliable and poorly reproducible;
- for second-line injectable drugs, the panel evaluated the performance in detecting resistance to amikacin only because both kanamycin and capreomycin are no longer recommended for the treatment of DR-TB; and
- culture-based phenotypic DST may be important to confirm amikacin susceptibility in situations where it is appropriate to use this medicine, to balance risk and benefit.

Research priorities

Research priorities for low complexity automated NAATs for detection of resistance to isoniazid and second-line anti-TB agents are as follows:

- diagnostic accuracy, in specific patient populations (e.g. children, people living with HIV, and patients with signs and symptoms of extrapulmonary TB) and in non-sputum samples;
- impact of diagnostic technologies on clinical decision-making and outcomes that are important to patients (e.g. cure, mortality, time to diagnosis and time to start treatment) in all patient populations;
- impact of specific mutations on treatment outcomes among people with DR-TB;
- use, integration and optimization of diagnostic technologies in the overall landscape of testing and care, as well as diagnostic pathways and algorithms;
- economic studies evaluating the costs, cost–effectiveness and cost–benefit of different diagnostic technologies;
- qualitative studies evaluating equity, acceptability, feasibility and end-user values of different diagnostic technologies;
- effect of non-actionable results (indeterminate, non-determinate or invalid) on diagnostic accuracy and outcomes that are important to patients;
- evaluation of low complexity automated NAATs for initial TB detection, in addition to its use as a follow-on test, in all people with signs and symptoms of TB, in children and in people living with HIV; and
- the potential utility of *katG* resistance detection to identify MDR-TB clones that may be missed because they do not have an RRDR mutation (e.g. the Eswatini MDR-TB clone, which has both the *katG S315T* and the non-RRDR *rpoB I491F* mutation).

2.4.2 First-line LPAs

In 2008, WHO approved the use of commercial LPAs for detecting MTBC in combination with resistance to rifampicin and isoniazid in sputum smear-positive specimens (direct testing) and in cultured isolates of MTBC (indirect testing). A systematic review at that time evaluated the diagnostic accuracy of two commercially available LPAs – the INNO-LiPA Rif.TB assay (Innogenetics, Ghent, Belgium), and the GenoType® MTBDR*plus* (version 1), hereafter referred to as Hain version 1 – and provided evidence for WHO's endorsement *(37, 38)*. Excellent accuracy was reported for both tests in detecting rifampicin resistance, but their diagnostic accuracy for isoniazid resistance had lower sensitivity, despite the high specificity. Because there were inadequate data to allow stratification by smear status, WHO's recommendation for using LPAs was limited to culture isolates or smear-positive sputum specimens. Further data have since been published on the use of LPAs; newer versions of LPA technology have now been developed, such as the Hain GenoType MTBDR*plus* version 2, hereafter referred to as Hain version 2; and other manufacturers have entered the market, including Nipro (Tokyo, Japan), which developed the Genoscholar™ NTM+MDRTB II, hereafter referred to as Nipro.

In 2015, FIND evaluated the Nipro and the Hain version 2 LPAs, and compared them with Hain version 1. The study demonstrated equivalence among the three commercially available LPAs for detecting TB and resistance to rifampicin and isoniazid *(5)*.

Table 2.4.2.1 Class criteria for LPAs

Purpose		Detection of resistance to first- and/ or second-line TB drugs
Principle of action		DNA-based reverse hybridization, or line probe, assays
Complexity	Reagents	Reagents are available within standardized kits and may have temperature requirements for storage.
	Skills	Advanced technical skills (i.e., multiple sample or reagent handling steps, precision pipetting, molecular workflows may be required)
	Pipetting	Multiple precision pipetting steps required by the procedure.
	Testing Procedure	May require multiple specimen treatment steps before transferring the specimen into a sealed container for multi-step testing. Manual or automated DNA extraction Manual or automated real-time PCR Instrument-based reverse hybridization
Type of test result reporting		Manual
Setting of use		Molecular laboratory (special infrastructure and separate of spaces for different parts of the testing procedure are required)

Recommendation

13. For persons with a sputum smear-positive specimen or a cultured isolate of MTBC, commercial molecular LPAs may be used as the initial test instead of phenotypic culture-based DST to detect resistance to rifampicin and isoniazid.

(Conditional recommendation, moderate certainty in the evidence for the test's accuracy)

Remarks

1. These recommendations apply to the use of LPAs for testing sputum smear-positive specimens (direct testing) and cultured isolates of MTBC (indirect testing) from both pulmonary and extrapulmonary sites.
2. LPAs are not recommended for the direct testing of sputum smear-negative specimens.
3. These recommendations apply to the detection of MTBC and the diagnosis of MDR-TB, but acknowledge that the accuracy of detecting resistance to rifampicin and isoniazid differs and, hence, that the accuracy of a diagnosis of MDR-TB is reduced overall.
4. These recommendations do not eliminate the need for conventional culture-based DST, which will be necessary to determine resistance to other anti-TB agents and to monitor the emergence of additional drug resistance.
5. Conventional culture-based DST for isoniazid may still be used to evaluate patients when the LPA result does not detect isoniazid resistance. This is particularly important for populations with a high pretest probability of resistance to isoniazid.
6. These recommendations apply to the use of LPA in children based on the generalization of data from adults.

Test description

LPAs are a family of DNA strip-based tests that can detect the MTBC strain and determine its drug resistance profile through the pattern of binding of amplicons (DNA amplification products) to probes targeting the following: specific parts of the MTBC genome (for MTBC detection), the most common resistance-associated mutations to first-line and second-line agents, or the corresponding wild-type DNA sequence (for detection of resistance to anti-TB drugs) *(38)*.

LPAs are based on reverse hybridization DNA strip technology and involve three steps: DNA extraction from *M. tuberculosis* culture isolates or directly from patient specimens, followed by multiplex PCR amplification and then reverse hybridization with visualization of amplicon binding (or lack thereof) to wild-type and mutation probes *(8)*.

Although LPAs are more technically complex to perform than the Xpert MTB/RIF assay, they can detect isoniazid resistance. Testing platforms have been designed for a reference laboratory setting and are thus most applicable to high TB burden countries. Results can be obtained in 5 hours.

Some of these steps can be automated, making the method quicker and more robust, and reducing the risk of contamination.

The Hain version 1 and version 2 assays include *rpoB* probes to detect rifampicin resistance, *katG* probes to detect mutations associated with high-level isoniazid resistance, and *inhA* promoter probes to detect mutations usually associated with low-level isoniazid resistance. The probes used to detect wild-type and specific mutations are the same for both versions of the Hain LPA.

Similarly, the Nipro assay allows for the identification of MTBC, and resistance to rifampicin and isoniazid. The Nipro assay also differentiates *M. avium*, *M. intracellulare* and *M. kansasii* from other non-tuberculous mycobacteria.

The *rpoB*, *katG* and *inhA* promoter mutation probes are the same for the three assays, with the exception of the *katG* S315N mutation, which is included in the Nipro assay but not in Hain version 1 or version 2. There are some minor variations in the codon regions covered for the wild type among Hain version 1 and version 2, and the Nipro.

Justification and evidence

In 2015, WHO commissioned an updated systematic review of the accuracy of commercial LPAs for detecting MTBC, and resistance to rifampicin and isoniazid. A total of 74 studies were identified, comprising 94 unique datasets (see **Web Annex A.7:** "FL-LPA"). Of these 94 datasets, 83 evaluated Hain version 1, five evaluated Hain version 2, and six evaluated the Nipro assay. Only one of the studies performed head-to-head testing of all three target LPAs on directly tested clinical specimens and indirectly tested isolates, and these data were included as six separate datasets *(9)*. No studies performed LPA testing on specimens and culture isolates from the same patients, precluding direct within-study comparisons.

Following the 2015 systematic review, the WHO Global TB Programme convened a GDG in March 2016 to assess the data and update the 2008 policy recommendations on using commercial LPAs to detect MTBC, and resistance to isoniazid and rifampicin. The PICO questions are given in **Box 2.4.2.1.**

LPAs were compared with a phenotypic culture-based DST reference standard, and a composite reference standard that combined the results from genetic sequencing with results from phenotypic culture-based DST. Phenotypic DST was the primary reference standard applied to all participants for all analyses. These analyses were stratified – first, by susceptibility or resistance to rifampicin or isoniazid (or both) and second, by type of LPA testing (indirect testing or direct testing).

Box 2.4.2.1 PICO questions

1. Should LPAs be used to guide clinical decisions to use rifampicin in the direct testing of specimens and the indirect testing of culture isolates from patients with signs and symptoms consistent with TB?
2. Should LPAs be used to guide clinical decisions to use isoniazid in the direct testing of specimens and the indirect testing of culture isolates from patients with signs and symptoms consistent with TB?
3. Should LPAs be used to diagnose MDR-TB in patients with signs and symptoms consistent with TB?
4. Should LPAs be used to diagnose TB in patients with signs and symptoms consistent with TB but for whom sputum-smear results are negative?

Several studies contributed to either sensitivity (no true positives and no false negatives) or specificity (no true negatives and no false positives) but not to both. For these studies, a univariate, random-effects meta-analysis of the estimates of sensitivity or specificity was performed separately, to make optimal use of the data. The results from the univariate analysis (using all studies) were compared with the results from the bivariate analysis of the subset of studies that contributed to estimates of both sensitivity and specificity.

If there were at least four studies for index tests with data that contributed only to sensitivity or specificity, a univariate, random-effects meta-analysis was performed to assess one summary estimate, assuming no correlation between sensitivity and specificity. In cases in which there were fewer than four studies, or where substantial heterogeneity was evident on forest plots that precluded a meta-analysis, a descriptive analysis was performed for these index tests. Forest plots were visually assessed for heterogeneity among the studies within each index test and in the summary plots, for variability in estimates and the width of the prediction region (a wider prediction region suggests more heterogeneity).

Implementation considerations

Adopting LPAs to detect rifampicin and isoniazid resistance does not eliminate the need for conventional culture and DST capacity. Culture and phenotypic culture-based DST have critical roles in monitoring patients' responses to treatment and detecting additional resistance to second-line agents.

- The adoption of LPA should be phased in, starting at national or central reference laboratories, or those with proven capability to conduct molecular testing. Expansion could be considered, within the context of a country's plans for laboratory strengthening, the availability of suitable personnel in peripheral centres and the quality of specimen transport systems.
- Adequate and appropriate laboratory infrastructure and equipment should be provided, to ensure that the required precautions for biosafety and the prevention of contamination are met – specimen processing for culture and procedures for manipulating cultures must be performed in biological safety cabinets in TB-containment laboratories.

- Laboratory facilities for LPAs require at least three separate rooms, one each for DNA extraction, pre-amplification procedures, and amplification and post-amplification procedures. To avoid contamination, access to molecular facilities must be restricted, a unidirectional workflow must be implemented and stringent cleaning protocols must be established.
- Appropriate laboratory staff should be trained to conduct LPA procedures. Staff should be supervised by a senior staff member with adequate training and experience in molecular assays. A programme for the external quality assessment of laboratories using LPAs should be developed as a priority.
- Mechanisms for rapidly reporting LPA results to clinicians must be established, to provide patients with the benefit of early diagnosis. The same infrastructure used for performing LPAs can be used also to perform second-line LPAs.
- LPAs are designed to detect TB and resistance to rifampicin and isoniazid in the direct testing of processed sputum samples, and in the indirect testing of culture isolates of MTBC. The use of LPAs with other respiratory samples (e.g. from BAL or gastric aspiration) or extrapulmonary samples (e.g. tissue samples, CSF or other body fluids) have not been adequately evaluated.
- The availability of second-line agents is critical in the event that resistance to rifampicin or isoniazid, or both, is detected.
- For patients with confirmed MDR/RR-TB, second-line LPAs are recommended to detect additional resistance to second-line anti-TB agents.

Research priorities

- Development of improved understanding of the correlation between the detection of resistance-conferring mutations using culture-based DST and patient outcomes.
- Review of evidence to confirm or revise different critical concentrations used in culture-based DST methods.
- Determination of the limit of detection for LPA in detecting heteroresistance.
- Determination of needs for training, assessing competency and ensuring quality assurance.
- Gathering of more evidence on the impact on mortality of initiating appropriate treatment for MDR-TB.
- Meeting the STARD for future diagnostic studies.
- Performance of country-specific cost–effectiveness and cost–benefit analyses of LPA use in different programmatic settings.

2.4.3 Second-line LPAs

Genotypic (molecular) methods have considerable advantages for scaling up programmatic management and surveillance of DR-TB, offering rapid diagnosis, standardized testing, potential for high throughput and fewer requirements for laboratory biosafety. Molecular tests for detecting drug resistance – for example, the GenoType MTBDR*sl* assay (Hain Lifescience, Nehren, Germany), hereafter referred to as MTBDR*sl* *(10)* – have shown promise for the diagnosis of DR-TB. These tests are rapid (can be performed in a single working day) and detect the presence of mutations associated with drug resistance. MTBDR*sl* belongs to a category of molecular genetic tests called second-line LPAs (SL-LPAs).

MTBDR*sl* (version 1.0) was the first commercial SL-LPA for detection of resistance to second-line TB drugs. In 2015, the manufacturer developed and made commercially available version 2.0 of the MTBDR*sl* assay. Version 2.0 detects the mutations associated with fluoroquinolones and second-line injectable drug (SLID) resistance detected by version 1.0, and additional mutations. Once a diagnosis of MDR/RR-TB has been established, an SL-LPA can be used to detect additional resistance to second-line drugs.

The MTBDR*sl* assay incorporates probes to detect mutations within genes that are associated with resistance to either fluoroquinolones or SLIDs (*gyr*A and *rrs* for version 1.0 and those genes plus *gyrB* and the *eis* promoter for version 2.0). The presence of mutations in these regions does not necessarily imply resistance to all the drugs within a particular class. Although specific mutations within these regions may be associated with different levels of resistance (i.e. different minimum inhibitory concentrations) to each drug within these classes, the extent of cross-resistance is not completely understood.

Recommendations

14. For patients with confirmed MDR/RR-TB, SL-LPA may be used as the initial test, instead of phenotypic culture-based DST, to detect resistance to fluoroquinolones.

(conditional recommendation, moderate certainty in the evidence for test accuracy)

15. For patients with confirmed MDR/RR-TB, SL-LPA may be used as the initial test, instead of phenotypic culture-based DST, to detect resistance to the SLIDs.

(conditional recommendation, low certainty in the evidence for test accuracy)

Remarks

- These recommendations apply to the use of SL-LPA for testing sputum specimens (direct testing) and cultured isolates of *M. tuberculosis* (indirect testing) from both pulmonary and extrapulmonary sites. Direct testing on sputum specimens allows for the earlier initiation of appropriate treatment.
- These recommendations apply to the direct testing of sputum specimens from MDR/RR-TB, irrespective of the smear status, while acknowledging that the indeterminate rate is higher when testing smear-negative sputum specimens than with smear-positive sputum specimens.
- These recommendations do not eliminate the need for conventional phenotypic DST capacity, which will be necessary to confirm resistance to other drugs and to monitor the emergence of additional drug resistance.
- Conventional phenotypic DST can still be used in the evaluation of patients with negative SL-LPA results, particularly in populations with a high pretest probability for resistance to fluoroquinolones or SLID (or both).
- These recommendations apply to the use of SL-LPA in children with confirmed MDR/RR-TB, based on the generalization of data from adults.
- Resistance-conferring mutations detected by SL-LPA are highly correlated with phenotypic resistance to ofloxacin and levofloxacin.

- Resistance-conferring mutations detected by SL-LPA are highly correlated with phenotypic resistance to SLID.
- Given the high specificity for detecting resistance to fluoroquinolones and SLID, the positive results of SL-LPA could be used to guide the implementation of appropriate infection control precautions.

Test description

The SL-LPA is based on the same principle as the first-line LPA. The assay procedure can be performed **directly** using a processed sputum sample or **indirectly** using DNA isolated and amplified from a culture of *M. tuberculosis.* Direct testing involves the following steps:

1. Decontamination (e.g. with sodium hydroxide) and concentration of a sputum specimen by centrifugation.
2. Isolation and amplification of DNA.
3. Detection of the amplification products by reverse hybridization.
4. Visualization using a streptavidin-conjugated alkaline phosphatase colour reaction.

Indirect testing includes only Steps 2–4. The observed bands, each corresponding to a wild-type or resistance-genotype probe, can be used to determine the drug susceptibility profile of the analysed specimen. The assay can be performed and completed within a single working day. Further details on the test process and practical support for implementation can be found in the WHO operational handbook. Module 3: diagnosis.

The index test used was MTBDR*sl* versions 1.0 and 2.0. These SL-LPAs detect specific mutations associated with resistance to the class of fluoroquinolones (including ofloxacin, levofloxacin, moxifloxacin and gatifloxacin) and SLIDs (including kanamycin, amikacin and capreomycin) in the MTBC. The MTBDR*sl* LPA detects mutations in the *gyrA* quinolone resistance-determining region (codons 85–97) and *rrs* (codons 1401, 1402 and 1484), and version 2.0 of the test added detection of mutations in the *gyrB* quinolone resistance-determining region (codons 536–541) and the *eis* promoter region (codons –10 to –14) *(40).* Version 2.0 is therefore expected to have improved sensitivity for resistance detection to these classes of drugs. Lastly, while version 1.0 included detection of mutations in *embB* that may encode for resistance to ethambutol, it was omitted from version 2.0 due to its status as a first line anti-TB drug. Therefore, this review did not determine the accuracy for ethambutol resistance.

More data are needed to better understand the correlation of the presence of certain fluoroquinolone resistance-conferring mutations with phenotypic DST resistance and with patient outcomes.

Justification and evidence

In March 2016, the WHO Global TB Programme convened a GDG to assess available data on the use of the MTBDR*sl* assay. WHO commissioned a systematic review on the accuracy and clinical use of assays for the detection of mutations associated with resistance to fluoroquinolones and SLID in people with MDR/RR-TB.

The PICO questions in **Box 2.4.3.1** were designed to form the basis for the evidence search, retrieval and analysis.

Box 2.4.3.1 PICO questions

1. **Should the MTBDRsl test be used to guide clinical decisions to use fluoroquinolones in patients with confirmed MDR/RR-TB?**
   - ➔ Direct testing (stratified by smear grade: smear negative; scanty; 1+; ≥2+).
   - ➔ Indirect testing.
2. **Should the MTBDRsl test be used to guide clinical decisions to use SLIDs in patients diagnosed with MDR/RR-TB?**
   - ➔ Direct testing (stratified by smear grade: smear negative; scanty; 1+; ≥2+).
   - ➔ Indirect testing.

Twenty-nine unique studies were identified; of these, 26 evaluated the MTBDR*sl* version 1.0 assay (including 21 studies from the original Cochrane review). Three studies (one published and two unpublished) evaluated version 2.0. Data for version 1.0 and version 2.0 of the MTBDR*sl* assay were analysed separately. A phenotypic culture-based DST reference standard was used for the primary analyses. These analyses were stratified first by susceptibility or resistance to a particular drug, and second by type of SL-LPA testing (indirect testing or direct testing).

Performance of SL-LPA on sputum specimens and culture isolates

In patients with MDR/RR-TB, a positive SL-LPA result for fluoroquinolone resistance (as a class) or SLID resistance (as a group) can be treated with confidence. The diagnostic accuracy of SL-LPA is similar when performed directly on sputum specimens or indirectly on cultured isolates of *M. tuberculosis*.

Given the confidence in a positive result and the ability of the test to provide rapid results, the GDG felt that SL-LPA may be considered for use as an initial test for resistance to the fluoroquinolones and when relevant SLIDs. However, when the test shows a negative result, phenotypic culture-based DST may be necessary, especially in settings with a high pretest probability for resistance to either fluoroquinolones or SLIDs (or both). The use of SL-LPA in routine care should improve the time to the diagnosis of fluoroquinolone and where relevant SLIDs, especially when used for the direct testing of sputum specimens of patients with confirmed MDR/RR-TB. Early detection of drug resistance should allow for the earlier initiation of appropriate patient therapy and improved patient health outcomes. Overall, the test performs well in the direct testing of sputum specimens from patients with confirmed MDR/RR-TB, although the indeterminate rate is higher when testing smear-negative sputum specimens compared with smear-positive sputum specimens.

When the MTBDR*sl* assay is used in the direct testing of smear-negative sputum specimens from a population of patients with confirmed DR-TB, up to 44% of the results may be indeterminate

(less with version 2.0, although very limited data) and hence require repeat or additional testing. However, if the same test were to be applied to the testing of smear-negative sputum specimens from patients without confirmed TB or DR-TB (i.e. patients suspected of having DR-TB), the indeterminate rate for the test would be significantly higher. Given the test's sensitivity and specificity when an SL-LPA is done directly on sputum, the GDG felt that SL-LPAs can be used for the testing of all sputum specimens from patients with confirmed MDR/RR-TB, irrespective of whether the microscopy result is positive or negative.

For the reasons mentioned above (inadequate data owing to too few studies on version 2.0), results are not presented here for version 2.0. For MTBDR*sl* version 2.0, the data were either too sparse or too heterogeneous to combine in a meta-analysis or to compare indirect and direct testing.

Three studies evaluated the MTBDR*sl* version 2.0 in 562 individuals, including 111 confirmed cases of TB with fluoroquinolone resistance by indirect testing on a culture of *M. tuberculosis* compared with a phenotypic culture-based DST reference standard. Estimates of sensitivity ranged from 84% to 100% and specificity from 99% to 100%.

See **Web Annex B.15:** *Drug concentrations used in culture-based DST SL-LPA* for details of the drug concentrations used in culture-based DST to evaluate the performance of SL-LPAs in each included study.

Implementation considerations

The SL-LPA should only be used to test specimens from patients with confirmed MDR/RR-TB. Adoption of SL-LPAs does not eliminate the need for conventional culture and DST capability. Despite good specificity of SL-LPAs for the detection of resistance to fluoroquinolones and the SLIDs, culture and phenotypic DST is required to completely exclude resistance to these drug classes as well as to other second-line drugs. The following implementation considerations apply:

- SL-LPAs cannot determine resistance to individual drugs in the class of fluoroquinolones. Resistance-conferring mutations detected by SL-LPAs are highly correlated with phenotypic resistance to ofloxacin and levofloxacin.
- Mutations in some regions (e.g. the *eis* promoter region) may be responsible for causing resistance to one drug in a class more than other drugs within that class. For example, the *eis* C14T mutation is associated with kanamycin resistance in strains from Eastern Europe.
- SL-LPAs should be used in the direct testing of sputum specimens, irrespective of whether samples are smear negative or smear positive.
- SL-LPAs are designed to detect TB and resistance to fluroquinolones and SLIDs from sputum samples. Other respiratory samples (e.g. BAL and gastric aspirates) or extrapulmonary samples (tissue samples, CSF or other body fluids) have not been adequately evaluated.
- Culture and phenotypic DST plays a critical role in the monitoring of a patient's response to treatment, and in detecting additional resistance to second-line drugs during treatment.
- SL-LPAs are suitable for use at the central or national reference laboratory level; they can also be used at the regional level if the appropriate infrastructure can be ensured (three separate rooms are required).
- All patients identified by SL-LPAs should have access to appropriate treatment and ancillary medications.

Research priorities

- Development of improved understanding of the correlation between the detection of resistance-conferring mutations with phenotypic DST results and with patient outcomes.
- Development of improved knowledge of the presence of specific mutations detected with SL-LPA correlated with minimum inhibitory concentrations for individual drugs within the classes of fluoroquinolones and SLIDs.
- Determination of the limit of detection of SL-LPA for the detection of heteroresistance.
- Gathering of more evidence on the impact of MTBDR*sl* on appropriate MDR-TB treatment initiation and mortality.
- Strongly encourage that future studies follow the recommendations in the STARD *(11)* statement to improve the quality of reporting.
- Performance of country-specific cost–effectiveness and cost–benefit analyses of the use of SL-LPA in different programmatic settings.

2.4.4 High complexity reverse hybridization-based NAATs for detection of pyrazinamide resistance

Pyrazinamide is an important antibiotic for the treatment of both drug-susceptible TB and DR-TB because of its unique ability to eradicate persisting bacilli and its synergistic properties with other antibiotics. Mono-resistance to pyrazinamide is rare; however, pyrazinamide resistance is strongly associated with MDR/RR-TB, with an estimated 30–60% of MDR/RR-TB also resistant to pyrazinamide. Thus, for people diagnosed with RR-TB, it is important to detect the presence of pyrazinamide resistance so that clinicians can make an informed decision on whether to include or exclude pyrazinamide in the treatment regimen. The high complexity hybridization-based NAAT may be used for diagnosis of pyrazinamide resistance on patient isolates; however, performance of this test requires appropriate infrastructure and skilled staff.

Recommendation

16. In people with bacteriologically confirmed TB, high complexity reverse hybridization-based NAATs may be used on *Mtb* culture isolates for detection of pyrazinamide resistance rather than culture-based phenotypic DST.

(Conditional recommendation, very low certainty of evidence for diagnostic accuracy)

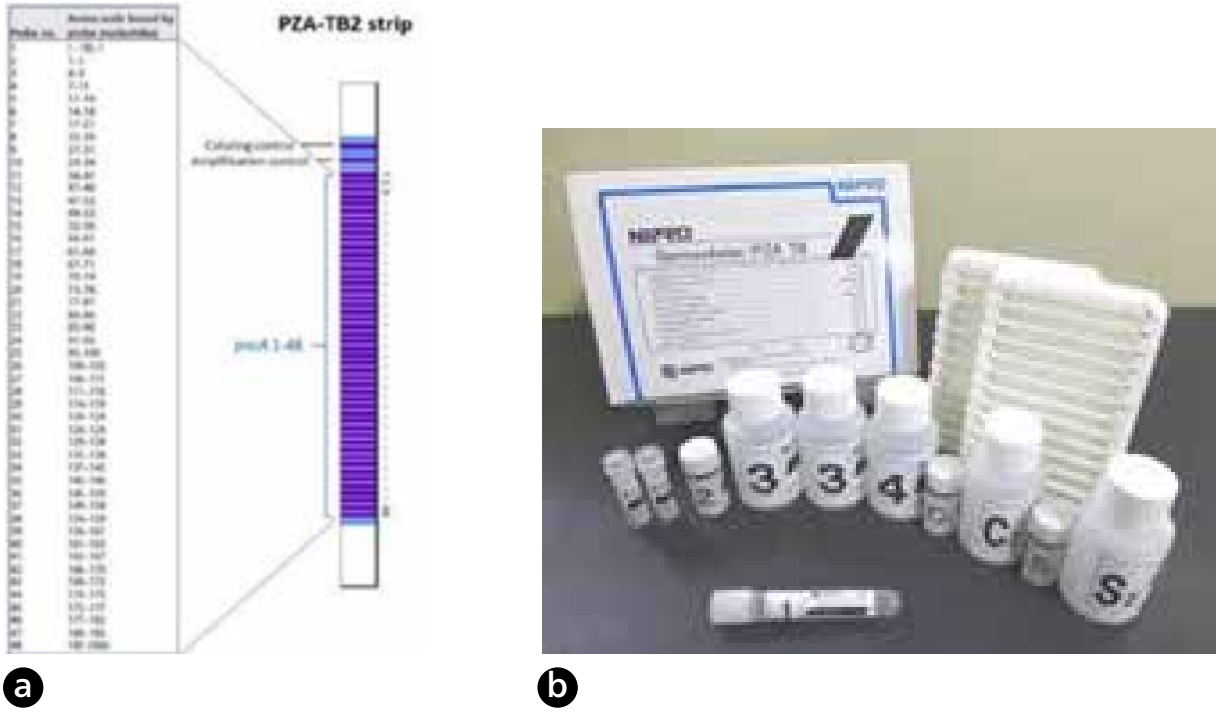
In terms of subgroups to be considered for this recommendation, no special considerations are required (e.g. for children, people living with HIV and those with extrapulmonary TB), given that the test is recommended for use on culture isolates.

Test description

Nipro (Osaka, Japan) developed Genoscholar™ PZA-TB, an LPA with reverse hybridization-based technology for detection of pyrazinamide resistance *(12)*. This assay is a commercially available rapid molecular test for detection of pyrazinamide resistance. Compared with MTBDR*plus* and MTBDR*sl* LPA, the Genoscholar PZA-TB LPA does not include specific mutant

probes because resistance mutations are widespread across the entire *pncA* gene with no predominant mutations. Instead, the Genoscholar PZA-TB assay targets a 700 base pair (bp) fragment covering the entire *pncA* gene and promoter region up to nucleotide –18 of the wild-type H37Rv reference strain.

Fig. 2.4.4.1 Nipro GenoScholar PZA-TB II strip (a) and Nipro GenoScholar PZA-TB II kit contents (b)



DNA extracted from cultures is amplified with primers by PCR. Amplified DNA is then hybridized to complementary oligonucleotide probes that are bound on a membrane strip. Streptavidin labelled with alkaline phosphatase is then added, to bind to any hybrids formed in the previous step. Next, a substrate is added, and an enzymatic reaction results in purple bands, which are visually interpreted. The absence of wild-type probe binding indicates the presence of a mutation. The first version of the assay contained 47 probes, which covered the *pncA* promoter and open reading frame. The second version contained 48 probes, three of which (*pncA* 16, 17 and 35) represent silent mutations known to be genetic markers not associated with pyrazinamide resistance: Gly60Gly (probe 16), Ser65Ser (probe 17) and Thr142Thr (probe 35).

Justification and evidence

The Genoscholar PZA-TB LPA assay, which is already commercially available, could potentially be implemented for diagnosis of pyrazinamide resistance in routine care. However, limited data have been published on the diagnostic accuracy of the assay. This systematic review with meta-analysis aimed to assist in collating all the available data to understand the diagnostic accuracy of the pyrazinamide LPA assay for detection of pyrazinamide resistance in TB patients, to guide policy-makers and clinicians.

The WHO Global TB Programme initiated an update of the current guidelines and commissioned a systematic review on the use of high complexity reverse hybridization-based NAATs for detection of pyrazinamide resistance in people with signs and symptoms of TB.

Two PICO questions were designed to form the basis for the evidence search, retrieval and analysis:

1. Should high complexity reverse hybridization-based NAATs on sputum be used to diagnose pyrazinamide resistance in patients with microbiologically confirmed pulmonary TB, irrespective of resistance to rifampicin, as compared with culture-based phenotypic DST or composite reference standard?

2. Should high complexity reverse hybridization-based NAATs on isolates be used to diagnose pyrazinamide resistance in patients with microbiologically confirmed pulmonary TB, irrespective of resistance to rifampicin, as compared with culture-based phenotypic DST?

The databases searched were PubMed, Web of Science and Embase, and they were searched without language or date restrictions. The search query was (PZA OR pyrazinamide OR pncA) AND (tuberculosis) AND ("line-probe assay" OR LPA OR "hybridization-based technology"). In addition, we approached Nipro (Osaka, Japan) to identify non-published data.

The microbiological reference standard was defined either as phenotypic culture-based DST performed using BD MGIT 960 PZA liquid assay or another acceptable phenotypic assay, or as genotypic DST performed using either targeted sequencing of the *pncA* gene or whole genome sequencing. In the case of genotypic DST, all samples with a *pncA* wild type were defined as being susceptible, while any variant in *pncA* was considered resistant, which implicitly would categorize "silent" mutations as resistant. In contrast, the composite reference standard was defined by classifying all samples with *pncA* wild type, *pncA* silent mutations and neutral mutations as being susceptible, while any other variant in *pncA* was considered resistant *(13)*.

Data synthesis was structured around the two preset PICO questions, as outlined below. Three web annexes give additional information, as follows:

- details of studies included in the current analysis (**Web Annex A.9:** *High complexity reverse hybridization-based NAATs*);
- a summary of the results and details of the evidence quality assessment (**Web Annex A.9:** *High complexity reverse hybridization-based NAATs*); and
- a summary of the GDG panel judgements (**Web Annex A.9:** *High complexity reverse hybridization-based NAATs*).

PICO 1: Should high complexity reverse hybridization-based NAATs on sputum be used to diagnose pyrazinamide resistance in patients with microbiologically confirmed pulmonary TB, irrespective of resistance to rifampicin, as compared with culture-based phenotypic DST or composite reference standard?

Three studies with a total of 122 participants provided data for evaluation of these NAATs for detection of pyrazinamide resistance, including two studies (101 participants) with phenotypic culture-based reference standard and one study (21 participants) with genotypic reference

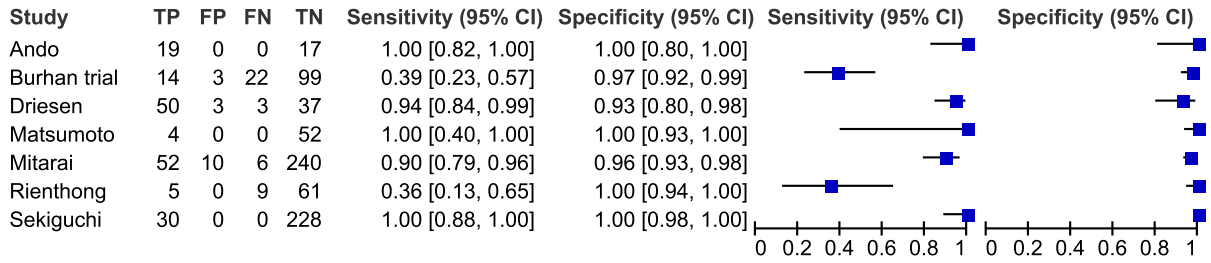
standard. **The number of studies and participants were considered insufficient to make a conclusion on a diagnostic accuracy of high complexity reverse hybridization-based NAATs on sputum.**

PICO 2: Should high complexity reverse hybridization-based NAATs on isolates be used to diagnose pyrazinamide resistance in patients with microbiologically confirmed pulmonary TB, irrespective of resistance to rifampicin, as compared with culture-based phenotypic DST?

Seven studies with a total of 964 participants provided data for evaluation of these NAATs for detection of pyrazinamide resistance compared with a phenotypic culture-based reference standard (**Fig. 2.4.4.2**).

The studies suffered from selection bias because they selected isolates with a wide range of different *pnc*A mutations rather than a representative sample from a population. Thus, the evidence was downgraded by one level for risk of bias. The included studies did not directly address the review question; hence, the evidence was downgraded one level for indirectness. The Burhan trial and the Rienthong study are outliers for their sensitivities compared with the other studies; hence, the evidence was downgraded one level for inconsistency. Taking these judgements together, the quality (certainty) of evidence was rated very low for sensitivity and low for specificity.

Fig. 2.4.4.2 Forest plot of included studies for pyrazinamide resistance detection, irrespective of rifampicin resistance with culture-based phenotypic DST as the reference standard



CI: confidence interval; DST: drug susceptibility testing; FN: false negative; FP: false positive; TB: tuberculosis; TN: true negative; TP: true positive.

The overall sensitivity for pyrazinamide resistance in these seven studies ranged from 36% to 100% and the specificity from 96% to 100%. **The pooled sensitivity was 81.2% (95% CI: 75.4–85.8%) and specificity was 97.8% (95% CI: 96.5–98.6%).**

More details on diagnostic accuracy of the high complexity reverse hybridization-based NAATs, including comparison with genotypic and composite reference standards are available in **Web Annex 4.17:** *High complexity reverse hybridization-based NAATs: diagnostic accuracy for detection of resistance to pyrazinamide. A systematic review*.

Cost–effectiveness analysis

This section answers the following additional question:

What is the comparative cost, affordability and cost–effectiveness of implementation of high complexity reverse hybridization-based NAATs?

A systematic review was carried out, focusing on economic evaluations of high complexity reverse hybridization-based NAATs. Four online databases (Embase, Medline, Web of Science and Scopus) were searched for new studies published from 1 January 2010 through 17 September 2020. The citations of all eligible articles, guidelines and reviews were reviewed for additional studies. The experts and test manufacturers were also contacted to identify any additional unpublished studies.

The objective of the review was to summarize current economic evidence and further understand the costs, cost–effectiveness and affordability of high complexity reverse hybridization-based NAATs.

No published studies were identified assessing costs or cost–effectiveness using the commercially available high complexity hybridization-based NAAT (Genoscholar PZA-TB II, Nipro Japan). Indirect evidence was available from several sources. Four studies examining other commercially available LPAs (Genotype MTBDR*sl* and MTBDR*plus*, Hain Lifescience) were identified.

The Genoscholar PZA LPA was developed for use with the Nipro automated MultiBlot; however, a recent unpublished trial[12] demonstrated that the Twincubator by Hain Lifescience could be used successfully with this LPA. This finding could make it easier to implement the Genoscholar PZA LPA in selected settings where Hain Lifescience equipment is already in use.

How large are the resource requirements (costs)?

No direct evidence from published studies was found regarding the total resources required. Resource requirements will include the purchase of test kits (Genoscholar PZA LPA: US$ 16/test kit consumables only), and the equipment, which is available for US$ 14 000. Operational costs are frequently several times greater than test kit costs (and will vary across settings), but are not accounted for usually. Nipro hopes that further reductions in test costs can be achieved when the Genoscholar PZA-TB II product is distributed globally.

Unit test costs for the Genotype MTBDR*sl* and MTBDR*plus* ranged from US$ 23.46 to US$ 108.70 *(14–15)*, with higher unit test costs in countries such as China and South Africa, largely driven by higher staff wages and operational costs. Extrapolations from unit test costs using different LPAs should be done with caution, and they are not intended to be directly transferrable estimates. Nevertheless, these indirect data do suggest that the total unit test cost of the Genoscholar PZA-TB II is likely several-fold higher than the unit test kit consumable cost of US$ 16.

Total costs will vary, depending on testing volume, numbers eligible for testing and prevalence of pyrazinamide resistance in the population. The impact on the budget will depend on the current standard of care, diagnostic and care pathways, and associated resource use.

[12] Leen Rigouts: Validation study of Genoscholar PZA LPA in three Supranational TB Reference Laboratories.

What is the certainty of the evidence of resource requirements (costs)?

Direct costs related to test kits and machinery are available, whereas several important items related to resource use (e.g. staff time, and overhead and operational costs associated with implementing Genoscholar PZA-TB II) have not been investigated. Differences in resource use between Genoscholar PZA-TB II and existing approaches will vary across settings that are using different phenotypic and genotypic DST. Also, there is important variability in costs of staff time and operation (e.g. testing volume) across settings.

Does the cost–effectiveness of the intervention favour the intervention or the comparison?

No cost–effectiveness studies were identified using the Genoscholar PZA-TB II. Extrapolation of cost–effectiveness data from other LPAs is not advised owing to differences in diagnostic accuracy, resistance prevalence, and the testing and treatment cascade of care.

More details on economic evidence synthesis and analysis are given in **Web Annex 4.9:** *Systematic literature review of economic evidence for NAATs to detect TB and DR-TB in adults and children*.

User perspective

This section answers the following questions about **key informants' views and perspectives on the use of high complexity reverse hybridization-based NAATs:**

- Is there important uncertainty about or variability in how much end-users value the main outcomes?
- What would be the impact on health equity?
- Is the intervention acceptable to key stakeholders?
- Is the intervention feasible to implement?

Findings of the review and interviews

The main findings of the systematic review and interviews are given below. Where information is from the review, a level of confidence in the QES is given; where it is from interviews, this is indicated with 'Interviews'.

Is there important uncertainty about or variability in how much end-users value the main outcomes?

- **Patients** in high burden TB settings value:
  - getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me");
  - avoiding diagnostic delays because they exacerbate existing financial hardships and emotional and physical suffering, and make patients feel guilty for infecting others (especially children);
  - having accessible facilities; and
  - reducing diagnosis-associated costs (e.g. travel, missing work) as important outcomes of the diagnostic.
    QES: moderate confidence

- The high complexity reverse hybridization-based NAATs meet some preferences and values of **laboratory staff** and **clinicians**, in that the current test:
  - provides quicker results about pyrazinamide resistance than other available methods (e.g. culture DST);
  - can provide information on different concentration levels; and
  - targets a drug that is widely used in first-line TB treatment.

  Interviews

What would be the impact on health equity?

The impact on health equity would be similar to that of moderate complexity automated NAATs (Section 2.1.2), plus the following:

- Lengthy diagnostic delays, underuse of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions hamper access to prompt and accurate testing and treatment, particularly for vulnerable groups.
  QES: high confidence
  Applicability to three index tests also confirmed in interviews
- Staff and managers voiced concerns about the sustainability of funding and maintenance, complex conflicts of interest between donors and implementers, and the strategic and equitable use of resources, which makes it difficult to ensure equitable access to cartridge-based diagnostics.
  QES: high confidence
- For patients, access to clear, comprehensible and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component of equity; lack of such access represents an important barrier for patients.
  Interviews
- New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, and to improve access to diagnostics for new treatment options.
  Interviews
- The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment:
  - between countries (i.e. between those that can and cannot quickly keep up with the rapidly changing TB diagnostic environment); and
  - within countries (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies).

  Interviews

Is the intervention acceptable to key stakeholders?

- Acceptability of a high complexity reverse hybridization-based NAAT depends on how well the test performs on different samples, because laboratory staff question how well LPA methods work on smear-negative samples. If samples need to be cultured before the pyrazinamide LPA is run, this may undermine the benefits of this method's quicker turnaround time compared with phenotypic DST for pyrazinamide. Acceptability also depends on how well the test actually detects mutations specific to pyrazinamide resistance; clinicians and

laboratory staff may require further clarification and justification in some settings as to why this specific drug test is being prioritized, given that it is not currently part of routine DST.
- Specific feasibility challenges (training and infrastructure requirements, sample quality result interpretation system), general feasibility challenges (as identified in the interview study and QES, respectively) and accumulated delays risk undoing the added value and benefits identified by the users (e.g. avoiding delays and drug-resistance information). *QES high confidence and interviews*

Is the intervention feasible to implement?

- The feasibility of implementing the pyrazinamide LPA is challenged by the significant training and laboratory infrastructure required to implement this method. Feasibility also hinges on the availability of an automated interpretation system, because the result is difficult to interpret. *Interviews*

Implementation considerations

Factors to consider when implementing a high complexity hybridization-based NAAT for detection of pyrazinamide resistance are as follows:

- There are specific concerns about the complexity and difficulty of interpretation. The large number of bands makes it difficult to read the result of the high complexity reverse hybridization-based NAAT.
- Local epidemiological data on resistance prevalence should guide local testing algorithms, whereas pretest probability is important for the clinical interpretation of test results.
- The cost of a test varies, depending on the number of samples in a batch, staff time and other parameters requiring a local costing exercise to be performed.
- Low, moderate, and high complexity tests have a successive increase in technical competency needs (qualifications and skills) and staff time, impacting planning and budgeting.
- Availability and timeliness of local support service and maintenance should be considered when selecting a provider.
- Laboratory accreditation and compliance with a robust quality management system (including appropriate quality control) is essential for sustained service excellence and trust.
- Training of both laboratory and clinical staff will ensure effective delivery of services and clinical impact.
- Use of connectivity solutions for communication of results is encouraged, to improve efficiency of service delivery and time to treatment initiation.
- Based on a multinational, population-based study, levels of pyrazinamide resistance varied widely in the surveyed settings (3.0–42.1%). In all settings, pyrazinamide resistance was significantly associated with rifampicin resistance *(49)*.
- Implementation of a high complexity hybridization-based NAAT requires laboratories with the required infrastructure, space and functional sample referral systems.
- Because there are several manual steps involved, well-trained staff are needed to set up assays and maintain instruments. Special training and experience are required for reading of banding patterns on the strip.

Research priorities

Research priorities for a high complexity hybridization-based NAAT for detection of pyrazinamide resistance are as follows:

- diagnostic accuracy of high complexity hybridization-based NAATs indirect testing on sputum and non-sputum samples in people with signs and symptoms of TB, with or without resistance to rifampicin;
- impact of diagnostic technologies on clinical decision-making and outcomes important to patients (e.g. cure, mortality, time to diagnosis and time to start treatment) in all patient populations;
- impact of specific mutations on treatment outcomes among people with DR-TB;
- use, integration and optimization of diagnostic technologies in the overall landscape of testing and care, as well as diagnostic pathways and algorithms;
- economic studies evaluating the costs, cost–effectiveness and cost–benefit of diagnostic technologies;
- qualitative studies evaluating equity, acceptability, feasibility and end-user values of diagnostic technologies; and
- interpretation of the results from a high complexity hybridization-based NAAT compared with sequencing and newer evidence on genotypic and phenotypic associations.

2.4.5 Targeted next-generation sequencing

Targeted NGS technology couples amplification of selected genes with NGS technology to detect resistance to many drugs with a single test. Also, since targeted NGS can interrogate entire genes to identify specific mutations associated with resistance, tests based on this technology may be more accurate than existing WRDs. In addition, new tests based on NGS can detect resistance to new and repurposed drugs that are not currently included in any other molecular assays. Hence, tests based on targeted NGS offer great potential to provide comprehensive resistance detection matched to modern treatment regimens.

Recommendations

17. In people with bacteriologically confirmed pulmonary TB disease, targeted next-generation sequencing technologies may be used on respiratory samples to diagnose resistance to rifampicin, isoniazid, fluoroquinolones, pyrazinamide and ethambutol rather than culture-based phenotypic drug susceptibility testing.

(Conditional recommendation, certainty of evidence moderate [isoniazid and pyrazinamide],low [rifampicin, fluoroquinolones and ethambutol])

Remarks

- Priority should be assigned to those at higher risk of resistance to first-line treatment medications, including individuals who:
  - continue to be smear or culture positive after 2 or more months of treatment, or experience treatment failure;
  - have previously had TB treatment,

- are in contact with a person known to have resistance to TB drugs; or
- reside in settings or belong to subgroups where there is a high probability of resistance to either rifampicin, isoniazid or fluoroquinolone (used in new shorter regimens), or where there is a high prevalence of *M. tuberculosis* strains harbouring mutations not detected by other rapid molecular tests.

- This recommendation is conditional because of the lack of data on health benefits, the variable certainty of evidence on diagnostic accuracy, and the fact that accuracy is suboptimal for certain drugs. In addition, because this is a new technology that has not yet been widely implemented, there is still limited and variable evidence on costs, cost–effectiveness and feasibility of implementation.

18. In people with bacteriologically confirmed rifampicin-resistant pulmonary TB disease, targeted NGS technologies may be used on respiratory samples to diagnose resistance to isoniazid, fluoroquinolones, bedaquiline, linezolid, clofazimine, pyrazinamide, ethambutol, amikacin and streptomycin rather than culture-based phenotypic drug susceptibility testing.

(Conditional recommendation, certainty of evidence high [isoniazid, fluoroquinolones and pyrazinamide], moderate [ethambutol], low [bedaquiline, linezolid, clofazimine and streptomycin], very low [amikacin])

Remarks

- Priority should be given to those at a higher risk of resistance to medications used for the treatment of RR-TB, including individuals who:
  - continue to be smear or culture positive after 2 months or more of treatment or have experienced treatment failure;
  - have previously had TB treatment, including with the new and repurposed drugs;
  - are in contact with a person known to have resistance to TB drugs, including the new and repurposed drugs; or
  - have pre-XDR-TB with resistance to fluoroquinolones.
- As above, this recommendation is conditional because of the lack of data on health benefits, the variable certainty of evidence on diagnostic accuracy, the fact that accuracy is suboptimal for certain drugs, and limited and variable evidence on costs, cost–effectiveness and feasibility of implementation.

Box 2.4.5.1

The products and drugs for which eligible data met the class-based performance criteria are listed below:

Deeplex® Myc-TB (Genoscreen, France): rifampicin, isoniazid, pyrazinamide, ethambutol, fluoroquinolones, bedaquiline, linezolid, clofazimine, amikacin and streptomycin

AmPORE-TB® (Oxford Nanopore Diagnostics, United Kingdom): rifampicin, isoniazid, fluoroquinolones, linezolid, amikacin and streptomycin

TBseq® (Hangzhou ShengTing Medical Technology Co., China): ethambutol

Where a product has not yet met the requirements for a specific drug (i.e., the drug is not listed), further improvements to the product are needed, and a review of the evidence is necessary before clinical use.

Test description

Three products met the inclusion criteria for detection of drug resistance to at least one of the anti-TB drugs under evaluation.

- The **Deeplex® Myc-TB test** (Genoscreen, France) is a targeted NGS-based kit for the simultaneous identification of mycobacterial species, genotyping and prediction of drug resistance of MTBC strains, directly applicable on sputum samples *(50)*. The assay relies on deep sequencing of a 24-plex amplicon mix, and it targets 18 MTBC gene regions associated with resistance to anti-TB drugs (rifampicin, isoniazid, pyrazinamide, ethambutol, fluoroquinolones, amikacin, kanamycin, capreomycin, streptomycin, ethionamide, bedaquiline, clofazimine and linezolid). Mycobacterial species identification is performed by targeting the *hsp*65 gene; the spoligotyping target (CRISPR/Direct Repeat locus) and phylogenetic single nucleotide polymorphisms (SNPs) in targets associated with drug resistance are used for MTBC strain genotyping. The assay is performed using the Nextera XT and DNA Flex library preparation kits on the iSeq 100, MiniSeq, MiSeq and NextSeq sequencing platforms (Illumina). The solution includes an automated analysis pipeline of the sequencing data in a secure online application with integrated databases for results interpretation.
- The **AmPORE-TB® test** (Oxford Nanopore Diagnostics, United Kingdom) – previously referred to as Nano-TB) – is a targeted NGS-based kit for the simultaneous identification of mycobacterial species and the detection of MTBC genetic variants associated with antimicrobial resistance in DNA extracted from sputum samples.[13] The assay relies on sequencing of a 27-plex amplicon mix: 24 drug-resistance targets, a genotyping target, a non-tuberculous mycobacteria (NTM) identification target (*hsp*65) and an internal control. The 24 drug-resistance targets are MTBC gene regions that are associated with resistance to various TB drugs (rifampicin, isoniazid, pyrazinamide, ethambutol, fluoroquinolones,

[13] Oxford Nanopore Diagnostics provided a draft protocol for the test.

amikacin, kanamycin, capreomycin, streptomycin, ethionamide, bedaquiline, clofazimine, linezolid and delamanid). Mycobacterial species identification is performed by targeting the *hsp*65 gene; the spoligotyping target (CRISPR/Direct Repeat locus) is used for MTBC strain genotyping. The assay is performed using the OND AmPORE-TB kit (OND-TBDR001-XX) and Flow Cells (OND-FLO-MIN001-XX) on the GridION Diagnostic Sequencing System (OND). The sequencing control software on the device can automatically start and report the results for the analysis workflows installed. The AmPORE-TB includes analysis software pre-installed on a device that processes readouts produced by the sequencing control software and creates an easy-to-interpret report, all performed locally on the device.

- The **TBseq® test** (Hangzhou ShengTing Medical Technology Co., China) is a kit based on targeted NGS that is used for the simultaneous identification of mycobacterial species and the prediction of drug resistance of MTBC strains; it is directly applicable to clinical specimens such as sputum and BAL fluid *(51)*. The assay relies on deep sequencing of a multiplex amplification mix and it targets 21 MTBC genes associated with resistance to TB drugs (rifampicin, isoniazid, pyrazinamide, ethambutol, fluoroquinolones, amikacin, kanamycin, capreomycin, streptomycin, para-aminosalicylic acid, cycloserine, ethionamide or prothionamide, bedaquiline, clofazimine and linezolid). Mycobacterial species identification is performed by targeting the 16S and *hsp*65 gene regions. The assay is performed using the Universal Gene Sequencing Kit (ShengTing) to generate libraries that are sequenced on either a MinION or a GridION platform (Oxford Nanopore Technologies). The solution includes automated analysis software (Nano TNGS V1.0) for sequencing data processing and a secure online application (TBseq® Web App) with integrated databases for interpretation of results.

Justification and evidence

Diagnostic accuracy and health benefits

Two health questions were designed using the PICO approach, to form the basis for the evidence search, retrieval and analysis.

1. Should targeted NGS as the initial test be used to diagnose drug resistance in individuals with bacteriologically confirmed pulmonary TB disease?
   This question applies to:
   - rifampicin, using a composite reference standard of phenotypic DST and whole genome sequencing (WGS), and Xpert MTB/RIF® or Xpert Ultra®;
   - isoniazid, using phenotypic DST as the reference standard;
   - levofloxacin, using phenotypic DST as the reference standard;
   - moxifloxacin, using phenotypic DST as the reference standard;
   - pyrazinamide, using a composite reference standard of phenotypic DST and WGS; and
   - ethambutol, using a composite reference standard of phenotypic DST and WGS.

2. Should targeted NGS be used to diagnose drug resistance in individuals with bacteriologically confirmed rifampicin-resistant pulmonary TB disease?
   This question applies to:
   - isoniazid, using phenotypic DST as the reference standard;
   - levofloxacin, using phenotypic DST as the reference standard;

- moxifloxacin, using phenotypic DST as the reference standard;
- pyrazinamide, using a composite reference standard of phenotypic DST and WGS;
- bedaquiline, using phenotypic DST as the reference standard;
- linezolid, using phenotypic DST as the reference standard;
- clofazimine, using phenotypic DST as the reference standard;
- amikacin, using phenotypic DST as the reference standard;
- ethambutol, using a composite reference standard of phenotypic DST and WGS; and
- streptomycin, using phenotypic DST as the reference standard.

A broad search was conducted to find, appraise and synthesize evidence about health benefits and the diagnostic test accuracy of targeted NGS compared with phenotypic drug sensitivity testing for patients with bacteriologically confirmed TB or with bacteriologically confirmed rifampicin-resistant pulmonary TB disease. A comprehensive search of three databases (Medline, Ovid Embase and Scopus) for relevant citations was performed. No date restriction was applied and the search was initially performed on 7 September 2022 and repeated on 17 January 2023. In addition, WHO made a public call for data and contacted well-known experts in the field to ask whether they had, or knew of, unpublished data that could contribute.

No data were found for the impact of targeted NGS on patient-level health effects. For the analysis of diagnostic accuracy, because few data were available in the literature, all data identified from the literature were included after correspondence with the authors. Hence, no manual data extraction from publications was required. A post-hoc decision was made to perform only an individual patient data (IPD) meta-analysis; thus, any study that could not provide IPD was excluded. Two report authors made independent assessments of methodological quality using QUADAS-2. Disagreements were resolved by discussion and uncertainties or disagreements were reviewed by an independent third party.

Subanalyses were performed to assess the diagnostic test accuracy in PLHIV and for semiquantitative results (derived from cycle thresholds) from Xpert MTB/RIF® or Xpert Ultra®, where "very low" or "low" concentrations of *M. tuberculosis* were compared with "medium" or "high" concentrations. The very low or low semiquantitative categories represent paucibacillary disease states, such as those frequently observed in paediatric TB.

Data were included from both published and unpublished prospective, observational clinical studies of targeted NGS platform diagnostic accuracy. All studies where targeted NGS had been performed directly from processed clinical samples were included, whereas those performed exclusively on cultured isolates were excluded. All studies were required to have comparator phenotypic DST data as a reference; in the cases of rifampicin, ethambutol and pyrazinamide, studies were required to also have WGS, to allow a composite reference to be generated. Rifampicin resistance results and semiquantitative results from Xpert MTB/RIF® or Xpert Ultra® were requested from all studies.

Given that this was a review of the diagnostic accuracy of a class of diagnostic platforms, all the data from each platform alone were analysed to assess which to include in an analysis to inform a class recommendation. Where the performance of any one platform appeared to be an outlier for sensitivity or specificity, that platform was excluded from subsequent meta-analyses. A platform was considered to be an outlier for a particular drug if the point estimate

for sensitivity was more than 10 percentage points worse than the best performing platform, or where the point estimate for specificity was more than 5 percentage points worse.

An IPD meta-analysis was performed instead of a classical meta-analysis, because the studies identified in the literature were generally too small to contribute to a classical meta-analysis, particularly for the new and repurposed drugs. In addition, this type of approach allowed for relevant co-variables to be included in the model; it could also control for repeated testing on the same samples using different platforms, which was the case for much of the available data.

For each dependent variable, a multivariable model included a number of co-variables as fixed effects. These included rifampicin resistance as determined by Xpert MTB/RIF® or Xpert Ultra® for all drugs other than rifampicin; semiquantitative cycle threshold (CT) value from Xpert MTB/RIF® or Xpert Ultra®; and a co-variable to indicate which samples featured in duplicate, meaning that some samples were sequenced on two different platforms and thus were represented twice in the analysis. For models looking specifically at diagnostic test accuracy in PLHIV, the HIV test result was included as a co-variable. Finally, the study site was included as a random effect. The models were run in Stata (version 17) using the melogit command, and the outputs were transformed using the margins command. Models were run for all PICO questions for sensitivity and specificity.

The certainty of the evidence of the pooled studies was assessed systematically for each of the PICO questions using the GRADE approach, which produces an overall quality assessment (or certainty) of evidence and has a framework for translating evidence into recommendations.

The GRADEpro Guideline Development Tool software *(16)* was used to generate summary of findings tables for the sensitivity and specificity of each drug. The numbers of samples classified as true, false positive or negative were then calculated across a range of three prevalences of drug resistance, chosen to be representative of different global settings. The quality of evidence was rated as high (not downgraded), moderate (downgraded one level), low (downgraded two levels) or very low (downgraded more than two levels), based on five factors: risk of bias, indirectness, inconsistency, imprecision and other considerations. The quality (certainty) of evidence was downgraded by one level when a serious issue was identified and by two levels when a very serious issue was identified in any of the factors used to judge the quality of evidence.

The data sources for the IPD data analysis are shown in **Fig. 2.4.5.1.** The analysis included data from published studies, a large multicountry trial conducted by FIND, and several other studies across multiple countries. Most of the studies only evaluated the Deeplex assay, while the FIND trial evaluated both the Deeplex and the AmPORE-TB. Only one study evaluated TBseq. For each drug, one or two platforms were dropped from the analysis based on the overall number of resistant or susceptible samples available for that platform and drug, or because the accuracy of the platform did not meet the diagnostic test accuracy criteria for inclusion when compared with the best performing platform.

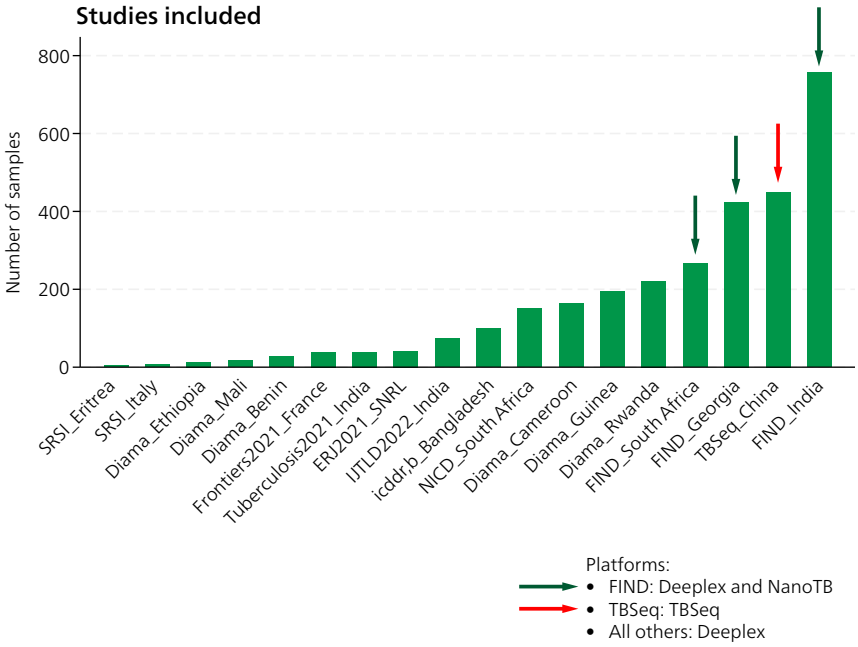
Fig. 2.4.5.1 Studies included in the IPD meta-analysis for targeted NGS

From the literature:

- ERJ 2021 (SNRL Germany)
- Frontiers 2021 (France)
- Tuberculosis 2021 (India)
- IJTLD 2022 (India)

Unpublished:

- FIND
  - Georgia
  - India
  - South Africa
- Diama
  - Benin
  - Guinea
  - Cameroon
  - Rwanda
  - Mali
  - Ethiopia
- NICD, South Africa
- icddr,b (Bangladesh)
- San Raffaele Scientific Institute, Italy
  - Italy
  - Eritrea
- TBSeq, China



ERJ: European Respiratory Journal; FIND: Foundation for Innovative New Diagnostics; IPD: individual patient data; NGS: next-generation sequencing; NICD: National Institute for Communicable Diseases; UTLD: International Union Against Tuberculosis and Lung Diseases.

Data synthesis was structured around the two preset PICO questions, as outlined below.

PICO 1: Should targeted NGS as the initial test be used to diagnose drug resistance in patients with bacteriologically confirmed pulmonary TB disease?

The available evidence included in the final pooled analysis varied by drug, from 12 studies with 1440 participants for the sensitivity of isoniazid to three studies with 269 participants for the specificity of pyrazinamide (**Table 2.4.5.1**). The pooled estimates were determined using a multivariable, mixed-effects model. All drugs were downgraded by one level for indirectness for sensitivity and specificity, because all studies were enriched for rifampicin resistance, leading to applicability concerns. In addition, for rifampicin, levofloxacin and pyrazinamide, specificity was downgraded a further level for imprecision; however, for ethambutol, it was downgraded for risk of bias because different samples were used for the index and reference tests. The overall certainty of the evidence for test accuracy ranged from moderate to very low.

The test performance was determined to be accurate for all drugs included in the assessment, with a **pooled sensitivity of at least 95% for isoniazid, moxifloxacin and ethambutol, more than 93% for rifampicin and levofloxacin, and 88% for pyrazinamide. The pooled specificity was at least 96% for all drugs.**

The reference standard was culture-based phenotypic DST for isoniazid, levofloxacin and moxifloxacin, and a combination of phenotypic DST and WGS for rifampicin, pyrazinamide and ethambutol. The percentage of tests with indeterminate results ranged from 9% (levofloxacin and moxifloxacin) to 18% (pyrazinamide), with higher indeterminate rates in samples with lower bacterial load (semiquantitative category low or very low).

Table 2.4.5.1 The accuracy and certainty of evidence of targeted NGS for the detection of resistance to anti-TB drugs among bacteriologically confirmed pulmonary TB

Drug	Reference standard	Accuracy % (95% CI)	Studies (persons)	Certainty in evidence
Rifampicin	Phenotypic DST+WGS	Se: 93.1 (87.0–99.2)	9 (1436)	Moderate
	Phenotypic DST+WGS	Sp: 96.2 (88.6–100)	7 (271)	Low
Isoniazid	Phenotypic DST	Se: 95.8 (92.8–98.7)	12 (1440)	Moderate
	Phenotypic DST	Sp: 97.0 (95.1–98.9)	12 (517)	Moderate
Levofloxacin	Phenotypic DST	Se: 94.2 (88.4–99.9)	6 (654)	Low
	Phenotypic DST	Sp: 96.2 (93.4–98.9)	7 (913)	Moderate
Moxifloxacin	Phenotypic DST	Se: 95.6 (92.4–98.7)	6 (652)	Moderate
	Phenotypic DST	Sp: 96.3 (93.2–99.5)	8 (921)	Moderate
Pyrazinamide	Phenotypic DST+WGS	Se: 88.4 (85.2–91.7)	3 (346)	Moderate
	Phenotypic DST+WGS	Sp: 98.5 (97.1–100)	3 (269)	Moderate
Ethambutol	Phenotypic DST+WGS	Se: 95.8 (94.0–97.6)	4 (432)	Low
	Phenotypic DST+WGS	Sp: 99.3 (98.2–100)	4 (268)	Low

CI: confidence interval; DST: drug susceptibility testing; NGS: next-generation sequencing; Se: sensitivity; Sp: specificity; TB: tuberculosis; WGS: whole genome sequencing.

There were no data on the impact of targeted NGS on patient outcomes such as time to treatment or treatment outcome.

PICO 2: Should targeted NGS be used to diagnose drug resistance in patients with bacteriologically confirmed rifampicin-resistant pulmonary TB disease?

The available evidence varied by drug, from 12 studies with 1440 participants for sensitivity of isoniazid to three studies with 31 participants for sensitivity of bedaquiline (**Table 2.4.5.2**). The pooled estimates were determined using a multivariable, mixed-effects model.

The overall certainty was high for some of the drugs. Levofloxacin was downgraded one level for inconsistency. Bedaquiline and linezolid were downgraded by two levels for imprecision in sensitivity because the number of resistant samples was below the threshold set and the confidence intervals were wide. Clofazimine was also downgraded by two levels, one for inconsistency (because two studies were outliers) and another level for imprecision (because the confidence intervals were wide). Amikacin was downgraded by one level for sensitivity and specificity because critical concentrations outside those recommended by WHO were used for a large proportion of samples. Amikacin sensitivity was further downgraded by two more levels, one for inconsistency and the other for imprecision. Ethambutol was downgraded by one level for risk of bias because different samples were used for the index and reference tests. Streptomycin specificity was downgraded by two levels, one for inconsistency and the other for imprecision. The overall certainty of the evidence for test accuracy ranged from high to very low.

The test performance among people with RR-TB was determined to be **accurate for isoniazid, levofloxacin, moxifloxacin, ethambutol and streptomycin (pooled sensitivity ≥95%) and acceptable for pyrazinamide (90%), bedaquiline (68%), linezolid (69%), clofazimine (70%) and amikacin (87%). The pooled specificity was 95% or greater for all drugs except streptomycin (75%).** The reference standard was culture-based phenotypic DST for all drugs except for ethambutol and pyrazinamide, where a combination of phenotypic DST and WGS was used. The percentage of tests with indeterminate results ranged from 9% (levofloxacin and moxifloxacin) to 21% (ethambutol); indeterminate rates were higher in samples with a lower bacterial load (semiquantitative category low or very low).

Table 2.4.5.2 The accuracy and certainty of evidence of targeted NGS for the detection of resistance to anti-TB drugs among bacteriologically confirmed rifampicin-resistant pulmonary TB

Drug	Reference standard	Accuracy % (95% CI)	Studies (persons)	Certainty in evidence
Isoniazid	Phenotypic DST	Se: 96.5 (93.8–99.2)	12 (1440)	High
	Phenotypic DST	Sp: 95.8 (91.8–99.8)	12 (517)	High
Levofloxacin	Phenotypic DST	Se: 95.8 (90.4–100)	6 (654)	Moderate
	Phenotypic DST	Sp: 96.0 (93.1–98.9)	7 (913)	High
Moxifloxacin	Phenotypic DST	Se: 96.5 (93.6–99.5)	6 (652)	High
	Phenotypic DST	Sp: 95.2 (91.0–99.4)	8 (921)	High
Pyrazinamide	Phenotypic DST+WGS	Se: 90.0 (86.8–93.2)	3 (346)	High
	Phenotypic DST+WGS	Sp: 98.6 (96.8–100)	3 (269)	High
Bedaquiline	Phenotypic DST	Se: 67.9 (42.6–93.2)	3 (31)	Low
	Phenotypic DST	Sp: 97.0 (94.3–99.7)	4 (519)	High
Linezolid	Phenotypic DST	Se: 68.9 (38.7–99.1)	4 (31)	Low
	Phenotypic DST	Sp: 99.8 (99.6–100)	6 (1093)	High
Clofazimine	Phenotypic DST	Se: 70.4 (34.6–100)	4 (36)	Low
	Phenotypic DST	Sp: 96.3 (93.2–99.3)	6 (789)	High
Amikacin	Phenotypic DST	Se: 87.4 (74.5–100)	5 (115)	Very low
	Phenotypic DST	Sp: 99.0 (98.4–99.6)	8 (1003)	Moderate
Ethambutol	Phenotypic DST+WGS	Se: 96.7 (95.0–98.4)	4 (431)	Moderate
	Phenotypic DST+WGS	Sp: 98.4 (96.1–100)	4 (123)	Moderate
Streptomycin	Phenotypic DST	Se: 98.1 (96.1–100)	5 (493)	High
	Phenotypic DST	Sp: 75.0 (59.5–90.5)	5 (250)	Low

CI: confidence interval; DST: drug susceptibility testing; NGS: next-generation sequencing; Se: sensitivity; Sp: specificity; TB: tuberculosis; WGS: whole genome sequencing.

There were no data on the impact of targeted NGS on patient outcomes such as time to treatment or treatment outcome.

Three web annexes give additional information, as follows:

- details of studies included in the current analysis (**Web Annex A.10:** *Review of the diagnostic accuracy of targeted NGS technologies for detection of drug resistance among people diagnosed with TB*);
- a summary of the results and details of the evidence quality assessment (**Web Annex A.10:** *GRADE profiles of targeted next-generation sequencing for detection of TB drug resistance*); and
- a summary of the GDG panel judgements (**Web Annex A.10:** *Evidence to decision tables: targeted next-generation sequencing for detection of TB drug resistance*).

Cost–effectiveness analysis

The cost and cost–effectiveness data for targeted NGS were assessed through a systematic review of the published literature and a generalized model-based cost–effectiveness analysis commissioned by WHO.

The systematic review on the cost and cost–effectiveness of using either targeted NGS or WGS to diagnose DR-TB searched three databases: PubMed, Embase and Scopus. The search was run on 30 October 2022 and had no time restriction. All costing data were inflated to 2021 US dollars. Findings were synthesized descriptively, given the considerable degree of heterogeneity in study methodology and outcomes. Among the studies included in the systematic review, three were on targeted NGS only, three were on targeted NGS and WGS, and four were on WGS only. For targeted NGS based on a single study (n=1), the cost per sample was between US$ 69.64 for Illumina MiSeq on 24 samples, and US$ 73.47 for Nanopore MinION on 12 samples; however, this costing was limited to only some components and did not include human resource costs or overhead costs. For WGS (n=5), cost per sample ranged from US$ 63.00 on Nanopore MinION to US$ 277.00 on Illumina MiSeq; given that studies used an inconsistent number of component costs, comparisons were challenging. Based on the review, the most significant cost component was the sequencing step, and the largest component costs were reagents and consumables, including those necessary for sequencing, sample processing and targeted NGS steps library preparation. Study authors identified four major cost drivers: use of different sequencers, depth and breadth of coverage, inefficiencies in initial sample runs, and economies of scale via batching or cross-batching.

The cost data from the systematic review were limited; therefore, an empirical unit costing was performed, in consultation with manufacturers and FIND. At the time of this work, only pricing for Deeplex Myc-TB was available and it was used for estimation of cost for the class. Unit costs included consumables, equipment, staffing and overheads (where available); also, costs assumed targeted NGS testing for all drugs. Based on the empirical analysis, the cost of targeted NGS was estimated to be:

- US$ 134 to US$ 257 in South Africa;
- US$ 120 to US$ 198 in Georgia; and
- US$ 121 to US$ 175 in India.

These costs are dependent on patient volume, batching and negotiated cost per targeted NGS kit.

Recognizing the lack of economic evidence on this topic, a hypothetical cost–effectiveness modelling study was undertaken to assess the cost–effectiveness (Objective 1) and affordability (Objective 2) of these tests for the diagnosis of DR-TB in various high TB burden settings.

Objective 1: To assess the potential cost–effectiveness of introducing the targeted NGS technology for the diagnosis of DR-TB in Georgia, India and South Africa.

This assessment included modelling the cost–effectiveness of targeted NGS in three separate scenarios with distinct comparison options:

a) Cost–effectiveness of targeted NGS for DST among individuals with RR-TB after a rapid molecular test for rifampicin resistance as a replacement for phenotypic DST (PICO 2).

b) Cost–effectiveness of targeted NGS for DST among individuals with RR-TB after a rapid molecular test for rifampicin resistance as a replacement for current in-country DST practice (PICO 2).

c) Cost–effectiveness of targeted NGS as the initial test for TB drug resistance in patients with bacteriologically confirmed TB compared with rapid molecular testing for drug resistance and phenotypic DST in a high DR-TB burden setting (PICO 1).

In the first scenario, targeted NGS was compared with universal phenotypic DST; in the second scenario, targeted NGS was compared with current in-country phenotypic DST practice among individuals with detected rifampicin resistance (PICO 2). This was done across three countries: Georgia, India and South Africa. Current DST practice in Georgia and South Africa includes Xpert XDR® followed by phenotypic DST; in India it includes LPAs and phenotypic DST done in parallel. A final scenario included targeted NGS compared to rapid molecular testing for drug resistance and phenotypic DST as initial tests for TB drug resistance among all TB patients (PICO 1) but was modelled for only one setting, Georgia – a high DR-TB burden setting. Epidemiological data were sourced from published literature; targeted NGS diagnostic accuracy data were sourced from the systematic review and IDP analysis conducted for this guideline. Economic data were sourced from published literature and a systematic and scoping review done in parallel by our team and supplemented with empirical data collection.

A decision analysis modelling approach was used to estimate the incremental cost–effectiveness of using targeted NGS for the diagnosis of DR-TB compared with various existing DST scenarios. This was done from the perspective of the health care system and accounts only for the health care system costs required to diagnose and treat TB. The estimation did not account for societal costs, or any direct or indirect costs incurred by patients. In addition, costs for sample transportation were not included in this analysis. The primary outcome was the incremental cost–effectiveness ratio (ICER), which was calculated as the incremental cost in US dollars per disability-adjusted life year (DALY) averted.

Main findings for PICO 1: Using targeted NGS as an initial test

Using targeted NGS as an initial test for DST in the high DR-TB burden setting of Georgia led to more health gains (DALYs=0.49) compared with Xpert MTB/RIF or Xpert Ultra, followed by

phenotypic DST (DALY=0.51). The ICER per DALY averted was US$ 9261 (95% uncertainty range [UR]: US$ 5258–32 040/DALY averted), which was considered cost effective at a willingness-to-pay (WTP) threshold of three times the country GDP per capita (US$ 15 609), with 80% of simulated iterations falling below the WTP threshold.

Main findings for PICO 2: Using targeted NGS among those with RR-TB

Using targeted NGS as a *replacement for universal phenotypic DST* among RR-TB patients, targeted NGS was dominated by phenotypic DST, with targeted NGS having higher costs and leading to fewer health gains. This finding was driven by the high diagnostic accuracy of phenotypic DST (which was assumed to be universal in this scenario), and an assumption of no difference in loss to follow-up between targeted NGS and phenotypic DST. When in-country DST practice was used as the comparator (instead of universal phenotypic DST), targeted NGS led to more health gains than in-country DST across all three countries. Targeted NGS was cost effective in South Africa (ICER: US$ 15 619/DALY averted, 95% UR: cost saving –US$ 114 782, at a WTP threshold of US$ 21 165), but was not cost effective in Georgia (ICER: US$ 18,375/DALY averted, UR: cost saving –US$ 158 972/DALY averted, at a WTP threshold of US$ 15 065). In India, where LPA, liquid culture and DST are being used as part of in-country DST, targeted NGS dominated the country's current DST practice, with lower costs and more health gains (95% UR: cost saving –US$ 60 083).

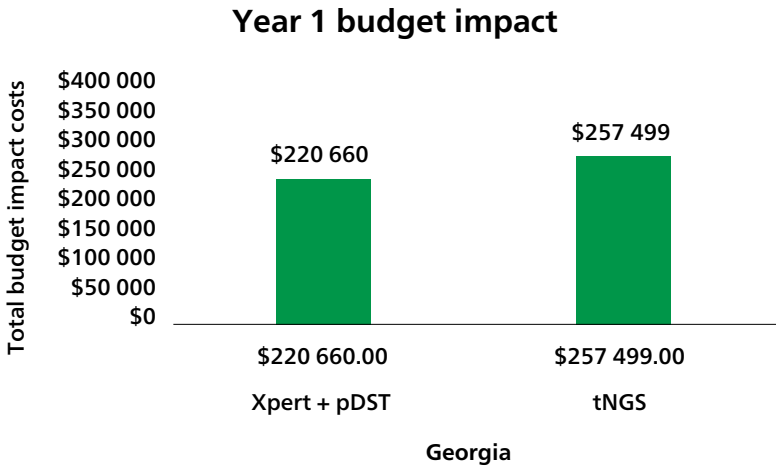
Main findings: scenario analyses

Several key scenario analyses were investigated. In the base case approach, loss to follow-up was assumed to be equivalent between phenotypic DST and targeted NGS; in a scenario where there was no loss to follow-up in targeted NGS compared with 10% in phenotypic DST, targeted NGS was cost effective in South Africa (ICER: US$ 13 004/DALY averted, WTP: US$ 21 165) and Georgia (ICER: US$ 13 640/DALY averted, WTP: US$ 15 069) and targeted NGS still dominated in-country DST practice in India. In scenarios where sequencing platforms are used for multiple different diseases to reduce the unit test cost of targeted NGS, the cost–effectiveness of targeted NGS improves in all three countries. A batching scenario was investigated, with an assumed 20% fewer samples per targeted NGS run, and led to an increased unit test cost for targeted NGS; in this scenario, the targeted NGS approach retained cost–effectiveness only in South Africa. When a 50% price reduction in targeted NGS test kit cost was assumed, targeted NGS cost–effectiveness further improved in all countries.

Objective 2: To assess the financial impact of introducing targeted NGS as a replacement for existing DST for diagnosis of DR-TB among TB patients across three countries: Georgia, India and South Africa.

A budget impact assessment was undertaken to estimate the financial consequences of adopting targeted NGS for DST for all patients diagnosed with TB, and replacing in-country DST practice in Georgia (PICO 1). The analysis suggested that implementing targeted NGS for all patients diagnosed with TB would be more expensive than testing all patients with Xpert MTB/RIF or Xpert Ultra, followed by phenotypic DST (see **Fig. 2.4.5.2**).

Fig. 2.4.5.2 Budget impact assessment results comparing current standard practice for DST with implementation of targeted NGS for all patients diagnosed with TB in Georgia

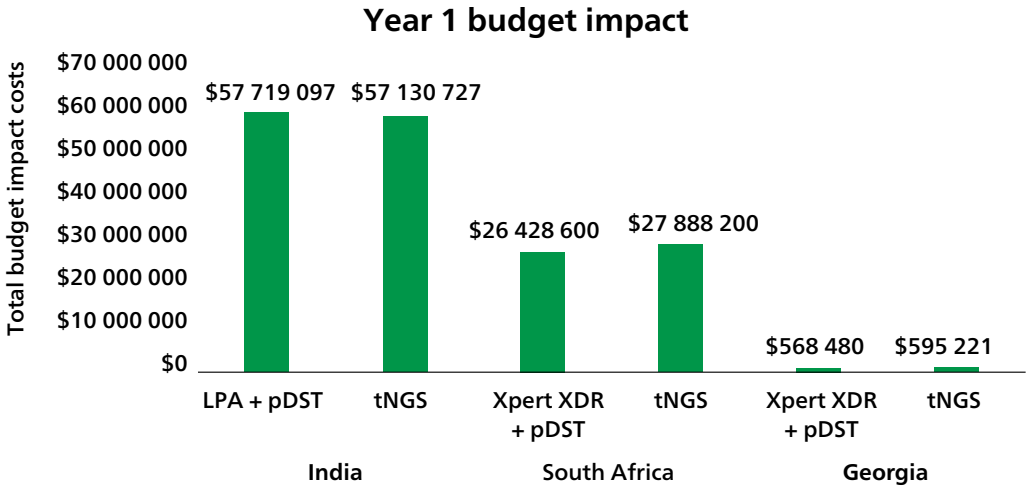


DST: drug susceptibility testing; NGS: next-generation sequencing; pDST: phenotypic DST; TB: tuberculosis; tNGS: targeted NGS.

A budget impact assessment was undertaken to estimate the financial consequences of adopting targeted NGS for DST after a rapid molecular test for rifampicin resistance, and replacing in-country DST practice in Georgia, India and South Africa (PICO 2). In-country DST practice included Xpert XDR combined with phenotypic DST in Georgia and South Africa, and Xpert XDR combined with LPA in Georgia over a 1-year and 5-year period. It was assumed that the eligible RR-TB patient populations requiring DST were 58 837, 8200 and 187 in South Africa, India and Georgia, respectively, and that the TB reduction rate over the 5 years was stable *(2)*. To estimate the impact on the country-specific budget, the economic costs generated by the model were multiplied by the number of patients.

Results from a 1-year budget impact assessment for PICO 2 are presented in **Fig. 2.4.5.3** In India, it was estimated that implementing targeted NGS would cost about US$ 57 130 727 – slightly lower than the current practice of LPA combined with phenotypic DST, which has a cost of US$ 57 719 097. In South Africa, it was estimated that implementing targeted NGS would result in a rise in budget to about US$ 27 888 200, slightly more than LPA combined with phenotypic DST, which has a cost of US$ 26 428 600. Finally in Georgia, where there are fewer bacteriologically confirmed patients, it was estimated that implementing targeted NGS would cost about US$ 592 221, slightly more than LPA combined with phenotypic DST, which has a cost of US$ 568 480.

Fig. 2.4.5.3 Budget impact assessment results comparing current standard practice for DST to implementing targeted NGS for patients with RR-TB in India, South Africa and Georgia



DST: drug susceptibility testing; LPA: line probe assay; NGS: next-generation sequencing; pDST: phenotypic DST; RR-TB: rifampicin-resistant TB; TB: tuberculosis; tNGS: targeted NGS.

User perspective

A rapid review was commissioned to identify and synthesize qualitative evidence on the use of targeted NGS for the detection of TB drug resistance; in particular, the aim was to examine the implementation considerations related to acceptability, feasibility, and values, preferences and equity. The review searched Medline with no year or language limits. The search was run on 19 August 2022, and then rerun on 10 October 2022 to include WGS-related studies for the detection of TB drug resistance. The review did not identify any eligible studies for analysis and synthesis. Based on the systematic search, three records were identified; in addition, based on the open, hand and expert searches, 27 records were found. On full-text review of the 30 records, none were found to be eligible for inclusion. Given that no direct evidence was found, note was made of a Cochrane qualitative evidence synthesis published in 2022 that examined recipient and provider perspectives on rapid molecular tests for TB and drug resistance *(52)*; that study provides relevant (though indirect) evidence on the subject. The authors noted that people with TB valued reaching diagnostic closure with an accurate diagnosis, avoiding diagnostic delays and keeping diagnostic associated costs low, whereas health care providers valued aspects of accuracy and the resulting confidence in low complexity NAAT results, rapid turnaround times and low costs to people seeking a diagnosis.

To address the direct evidence gap, WHO commissioned an additional qualitative cross-sectional study comprising semi-structured interviews, primarily with laboratory staff and management personnel directly involved with implementing targeted NGS in the three FIND trial sites, as well as with three global experts involved in TB care and diagnostics. In total, there were 17 respondents, and the work was conducted during September to October 2022. The objective was to explore the perceptions and experiences of those implementing targeted NGS technology, with respect to acceptability, feasibility, and values, preferences and equity. The main findings are summarized below.

Acceptability

A consistently positive sentiment was expressed for the acceptability and potential utility of targeted NGS technology. Targeted NGS was seen as a "major advancement" in molecular MDR-TB diagnostics.

1. The main reasons for the **high level of acceptability were the comprehensiveness** (resistance diagnosis for more drugs and for the newest and repurposed drugs), **the convenience of using a sputum sample** (as compared with culture samples), **and the rapidity** (quick results compared with phenotypic testing times; 3–5 days as compared with 4–6 weeks).

2. There was also the sense that there is **a good window of opportunity to benefit from the utility of targeted NGS technology;** that is, the technology is arriving at the right time, given that resistance to newer TB drugs is likely to increase as the use of these drugs becomes routine.

Feasibility

Although there was high praise for the capability and potential utility of targeted NGS technology, **several challenges were identified when testing samples using the targeted NGS platforms,** which may limit the feasibility of targeted NGS for routine uptake at the present time. The overall sentiment was that the targeted NGS technology needs to be further developed before it can be considered fully ready for operational use.

The following feasibility challenges were identified:

- **Start-up and setting-up challenges:** Multiple problems were identified with starting and setting up the technology. These problems related to the newness of the technology and the trial setting, importing technology and specialist supplies, lack of in-country technical assistance for problem-solving and need for more hands-on training practice.
- **High technical complexity of the test:** Targeted NGS technology was seen as a high complexity molecular test that was technically challenging. For example, preparing the sample for sequencing involves multiple steps that require attention to detail and precision, leaving little room for error. Preparation of the library is particularly complex for the Deeplex platform, although both the Deeplex and the Nanopore platforms are quite complex. In both platforms, it was thought that there were too few opportunities for early recognition and correction of errors, increasing the risk of failed runs.
- **Specialized laboratory infrastructure and human resource requirements:** Because targeted NGS is a molecular-based testing platform, it requires highly specialized laboratory infrastructure (e.g. multiple rooms to prevent amplicon contamination and specialized cold storage facilities). Also, highly specialized molecular and medical scientists are needed to perform the tests. In LMIC settings, such specialized laboratory infrastructure and staff may only be available at centralized laboratories (i.e. not at regional laboratories).
- **Special requirements for operating the test:** In addition to highly specialized laboratory infrastructure and staff, the testing technology also requires an uninterrupted supply of electricity, high internet connectivity, high computer capacity, clean water and temperature controls – requirements that may pose challenges in some LMIC settings.

- **Supply chain challenges:** Major challenges were reported relating to the required supply chain for implementing targeted NGS. Procurement bottlenecks and delays coupled with shelf-life limitations of reagents jeopardize continuous access to specialist supplies.
- **Data management and storage requirements:** There were concerns that data analysis and data storage requirements were not fully developed, including systems for backing up data, ownership of data and security of data. Another issue that needs to be considered is how targeted NGS and routine laboratory information systems can be interlinked.
- **Continuous updating of the WHO catalogue of mutations is required:** There was agreement that the usefulness of the targeted NGS technology depends on the informational support provided by the WHO catalogue of mutations *(53)*, which allows for meaningful interpretation of resistance data; thus, there is a need for the WHO catalogue to be continuously updated.

- **Feasibility concerns differed for the different targeted NGS platforms:** The overall sentiment was that all targeted NGS platforms needed to be further developed before they are fully ready for operational use, some more than others. The high level of technical complexity of the sample preparation stages (mainly the library preparation stage) was considered a key challenge for the Deeplex platform, and the need for improved computer analysis and storage capacity was a challenge for the Oxford Nanopore platform, although both required a high level of precision and attention to detail. There is also a need to incorporate steps for early error recognition.

Values, preferences and equity

The overall sentiment is that **MDR-TB diagnostic technology needs to balance accuracy, speed, affordability, equity and cost–effectiveness**, and that targeted NGS technology would need to address these considerations before it can be implemented in LMIC settings. These considerations were consistent across the different stakeholder groups who participated in the study.

1. **Centralized versus decentralized placement may have equity implications for access:** Given the high-level specialized laboratory infrastructure, specialized human resources and technical complexity needed for targeted NGS, the technology may be suitable for placement only at centralized, reference laboratories. This may have equity access considerations if it means less access for some regions of a country that lack reference laboratories. This may also have implications for costs (e.g. costs for transport of sputum), probability of sample loss and time to results.

2. **Affordability and cost–effectiveness are major concerns:** There was a major concern about the financial costs of the targeted NGS technology and the affordability for LMIC. Participants were worried about the cost of the equipment and the costs of ongoing specialist supplies (especially reagents), as well as the cost of maintaining equipment. They noted that costing calculations should be comprehensive and should include the cost of special consumables, extra general laboratory consumables and additional infrastructure needs (e.g. extra space, temperature control and internet connectivity). There were concerns that cost–effectiveness calculations should be comprehensive and should include assessment of the impact of the use of targeted NGS testing on improving TB outcomes.

3. **The MDR/RR-TB case burden of a country could influence equitable access at centralized levels.** In some settings with high caseloads, the targeted NGS technology capacity in central laboratories may not be sufficient for processing large caseloads in good time; also, in settings with low caseloads, waiting for sufficient samples to batch-test will cause delays.

Implementation considerations

Although the evidence that is available supports the use of targeted NGS to detect drug resistance after TB diagnosis, to guide clinical decision-making for DR-TB treatment, the following factors need to be considered when implementing these tests:

- Regulatory approval from national regulatory authorities or other relevant bodies is required before implementation of these diagnostic tests.
- In its current format, targeted NGS is a high complexity test that is most suitable for centralized laboratories equipped with specialized skills and infrastructure.
- Targeted NGS tests do not replace existing rapid tests that are more accessible and easier to perform for detecting resistance to rifampicin, isoniazid and fluoroquinolones. However, if targeted NGS can be performed rapidly, it can be considered as an alternative initial option for prioritized populations. Those who will benefit most from these tests are individuals who require rapid and comprehensive DST but have limited access to phenotypic DST.
- Priority should be given to samples with a high bacillary load as determined by initial bacteriological tests (e.g. semiquantitative high/medium or smear-positive grading). In situations where the bacillary load is low (e.g. semiquantitative low/very low/trace or smear-negative grading), the recommendations still hold, although rates of indeterminate results are likely to be higher; therefore, phenotypic DST is likely still required for samples with a low bacillary load.
- Similarly, the recommendations apply to children, adolescents and PLHIV populations because these populations have a higher frequency of samples with low bacterial load.
- The recommendation is based on data obtained from sputum and BAL specimens, and can be extrapolated to other lower respiratory tract samples (e.g. endotracheal aspirates). However, further research is needed to evaluate the use of these tests on alternative sample types for diagnosing pulmonary TB in children (e.g. nasopharyngeal and stool samples) and diagnosing extrapulmonary TB.
- Since sensitivity for bedaquiline, linezolid and clofazimine resistance is suboptimal, consideration of the pretest probability is important in interpreting the targeted NGS results for these drugs. Further testing of samples with a susceptible result (using culture-based phenotypic DST) would be warranted, particularly when the risk of resistance is high. Since specificity is high, a result that indicates resistance may be used to guide the therapy, particularly among those at risk for resistance. In the case of pretomanid, the basis for resistance has not been fully elucidated; hence, culture-based DST is also required for this drug.

Research priorities

Several key research priorities emerged from the reviews of the available evidence on targeted NGS for detecting TB drug resistance. They fall into three main categories: clinical research, implementation research, and monitoring and evaluation.

Clinical research:

- Conduct clinical trials to assess the impact of targeted NGS on patient-important outcomes[14].
- Evaluate the accuracy and impact on patient-important outcomes of targeted NGS among populations of individuals diagnosed with TB, across a range of prevalences of rifampicin or other drug resistance).
- Assess the accuracy and impact on patient-important outcomes of targeted NGS for detecting resistance to new and repurposed drugs, including pretomanid, across varied geographical and epidemiological settings.
- Assess the accuracy and impact on patient-important outcomes of targeted NGS for analysing extrapulmonary samples, including CSF for meningitis, non-sputum samples (e.g. nasopharyngeal aspirate, gastric aspirate or stool) for children, and alternative sample types (e.g. tongue swabs) in both adults and children.
- Undertake additional qualitative and quantitative research to further understand the perspectives of end-users and clinicians regarding the acceptability and feasibility of using targeted NGS.

Implementation research:

- Develop and evaluate effective and efficient implementation models by integrating targeted NGS into laboratory networks and optimizing algorithms, with the aim of enhancing timely access to testing and treatment initiation, and improving patient outcomes.
- Develop strategies to enhance the efficiency of targeted NGS testing, including sample processing and concentration techniques, determining optimal thresholds of bacterial load from initial tests before performing targeted NGS, and employing molecular transport medium for the ambient storage and transfer of samples to testing sites.
- Regularly update the WHO catalogue of mutations *(53)*, incorporating additional genetic targets and including new drugs (e.g. pretomanid) to enhance the sensitivity and specificity of targeted NGS.
- Explore technological advancements to simplify the testing process, automate steps (especially library preparation), develop decentralized targeted NGS solutions and investigate potential synergies with existing initial tests (e.g. using leftover DNA or smear-positive slides).
- Conduct comprehensive mapping of sequencing capacity within countries and perform diagnostic network optimization exercises. Placement of the technology should consider the demand for sequencing across multiple diseases, facilitating cross-disciplinary use of the machines and shared costs.
- Compile and use lessons learned from applying targeted NGS technology in other diseases (e.g. COVID-19) to develop effective implementation strategies for TB.

[14] Mortality, Cure, Lost to follow up; Time to diagnosis; Time to treatment.

Monitoring and evaluation:

- Standardize the nomenclature for reporting of results across different targeted NGS technologies, for integration into health information data systems.
- Ensure separate recording of true failures and unclassified mutations, and monitor trends over time as an essential component of result reporting.
- Regularly monitor performance data, including overall resistance rates, resistance rates by specific drugs or targets and turnaround times (both total and in-laboratory).
- Incorporate quality monitoring measures, such as tracking indeterminate rates, sequencing coverage and depth, and participating in external quality assurance programmes.
- Establish an external quality assurance programme for sequencing that covers all relevant targets of interest.
- Integrate the sequencing data generated into existing surveillance systems to monitor the prevalence and trends in drug resistance effectively. Share the data to update the WHO mutation catalogue.
- Collect cost data to address important questions, such as the costs associated with introducing and scaling up targeted NGS in different settings, the trade-offs between turnaround time and batching, and the optimal balance in various settings.
- Assess the impact of multidisease testing on programme operations and costs, including disease-specific testing volumes, turnaround times, costing, resource sharing and resource requirements.
- Evaluate the impact of time to treatment initiation or modification, treatment outcomes and overall cost–effectiveness of targeted NGS implementation.

2.5. References

1. Global tuberculosis report 2024. Geneva: World Health Organization; 2024 (https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024).
2. Resolution 78/L.4. Political declaration of the high-level meeting of the General Assembly on the fight against tuberculosis. New York: United Nations; 2023 (https://digitallibrary.un.org/record/4022582).
3. Implementing the End TB Strategy: the essentials. Geneva: World Health Organization; 2015 (https://apps.who.int/iris/handle/10665/206499).
4. Boyer S, March L, Kouanfack C, Laborde-Balen G, Marino P, Aghokeng AF, Mpoudi-Ngole E, Koulla-Shiro S, Delaporte E, Carrieri MP, Spire B, Laurent C, Moatti JP; Stratall ANRS 12110/ESTHER Study Group. Monitoring of HIV viral load, CD4 cell count, and clinical assessment versus clinical monitoring alone for antiretroviral therapy in low-resource settings (Stratall ANRS 12110/ESTHER): a cost-effectiveness analysis. Lancet Infect Dis. 2013 Jul;13(7):577–86. doi: 10.1016/S1473–3099(13)70073–2. Epub 2013 Apr 18. PMID: 23602084.
5. Eckman MH, Ward JW, Sherman KE. Cost effectiveness of universal screening for hepatitis C virus infection in the era of direct-acting, pangenotypic treatment regimens. Clin Gastroenterol Hepatol. 2019;17:930–9. e9. doi: https://doi.org/10.1016/j.cgh.2018.08.080.
6. Wang J-H, Chen C-H, Chang C-M, Feng W-C, Lee C-Y, Lu S-N. Hepatitis C virus core antigen is cost-effective in community-based screening of active hepatitis C infection in Taiwan. J Formos Med Assoc. 2020;119:504–8. doi: https://doi.org/10.1016/j.jfma.2019.07.011.
7. Report for WHO: non-inferiority evaluation of Nipro NTM+MDRTB and Hain GenoType MTBDRplus V2 line probe assays. Geneva: Foundation for Innovative New Diagnostics; 2015.
8. Nathavitharana RR, Cudahy PG, Schumacher SG, Steingart KR, Pai M, Denkinger CM. Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. Eur Respir J. 2017;49:1601075. doi: https://doi.org/10.1183/13993003.01075–2016.
9. Rapid diagnosis of tuberculosis brochure. Nehren, Germany: Hain Lifescience; 2015 (http://www.hain-lifescience.de/uploadfiles/file/produkte/mikrobiologie/mykobakterien/tb_eng.pdf).
10. Bossuyt P, Reitsma J, Bruns D, Gatsonis C, Glasziou P, Irwig L et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. doi: https://doi.org/10.1136/bmj.h5527.
11. Sekiguchi J, Nakamura T, Miyoshi-Akiyama T, Kirikae F, Kobayashi I, Augustynowicz-Kopec E et al. Development and evaluation of a line probe assay for rapid identification of pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis strains. J Clin Microbiol. 2007;45:2802–7. doi: https://doi.org/10.1128/jcm.00352–07.
12. Köser CU, Cirillo DM, Miotto P. How to optimally combine genotypic and phenotypic drug susceptibility testing methods for pyrazinamide. Antimicrob Agents Chemother. 2020;64:e01003–20. doi: https://doi.org/10.1128/AAC.01003–20.
13. Groessl EJ, Ganiats TG, Hillery N, Trollip A, Jackson RL, Catanzaro DG et al. Cost analysis of rapid diagnostics for drug-resistant tuberculosis. BMC Infect Dis. 2018;18:102. doi: https://doi.org/10.1186/s12879–018–3013–0.
14. Li X, Deng Y, Wang J, Jing H, Shu W, Qin J et al. Rapid diagnosis of multidrug-resistant tuberculosis impacts expenditures prior to appropriate treatment: a performance and diagnostic cost analysis. Infect Drug Resist. 2019;12:3549–55. doi: https://doi.org/10.2147/idr.S224518.
15. GRADEpro GDT [website]. Hamilton, Ontario: McMaster University; 2020 (https://gradepro.org/).

3. Recommendations for diagnosis of TB infection

3.1. Mycobacterium tuberculosis antigen-based skin tests for the diagnosis of TB infection

Since 2011, the World Health Organization (WHO) has issued recommendations on the use of IGRAs for the diagnosis of TB infection. In 2018, WHO updated the recommendations to stipulate that the TST or IGRAs (or both) can be used to test for TB infection in LMIC. The TST is a widely used point-of-care test that involves intradermal injection of purified protein derivative (PPD), a crude mixture of different mycobacterial antigens, which stimulates a delayed-type hypersensitivity response and causes induration at the injection site within 48–72 hours. This test has relatively low specificity in those with recent bacille Calmette-Guérin (BCG) vaccination and low sensitivity in immunosuppressed individuals (e.g. people living with HIV [PLHIV]); hence, interpretive cut-offs must be adapted for these populations. A follow-up clinic visit is required after the placement of the TST, and results must be read within the suggested time frame to be valid. In contrast, IGRAs are in vitro tests that measure release of interferon- gamma (IFN-γ) by T-cells following stimulation by the early secretory antigenic target 6 kDa protein (ESAT-6) and culture filtrate protein 10 (CFP-10) antigens that are specific to Mtb. Unlike the TST, IGRAs are not affected by prior BCG vaccination, or by infection with nontuberculous mycobacteria (NTM), with few exceptions. However, IGRA platforms are more expensive to run and require specialized facilities and trained personnel; consequently, the TST is the most commonly used test for TB infection globally. Recent global shortages of PPD have underscored the need for alternatives.

In addition to the TB skin tests and interferon gamma release assays previously recommended by WHO, Mtb antigen-based skin tests (TBSTs) based on specific antigens have recently been developed, using the same ESAT-6 and CFP-10 antigens; these tests combine the simpler skin-test platform with the specificity of IGRAs. TBSTs include the Cy-Tb (Serum Institute of India, India), Diaskintest® (Generium, Russian Federation) and C-TST (formerly known as ESAT6-CFP10 test, Anhui Zhifei Longcom, China). All tests use intradermal injection of antigen and, like the TST, are read after 48–72 hours as induration in millimetres, using the method suggested by Mantoux. Emerging evidence suggests that, compared with IGRAs, the tests may have similar specificity and provide more reliable results in children and adolescents as well as in PLHIV than the TST. However, the evidence had not been systematically reviewed.

In 2021, WHO commissioned a systematic review of published and unpublished data on this new class of tests for TB infection not previously revieed by WHO. The systematic review included

data on diagnostic accuracy, safety, economic aspects and qualitative evidence on feasibility, acceptability, equity, end-user values and preferences. A Guideline Development Group (GDG) was convened by WHO from 31 January to 3 February 2022, to discuss the findings of the systematic reviews and to make recommendations on this class of diagnostic technologies for TB infection.

The following technologies were included in the evaluation:

- Cy-Tb (Serum Institute of India, India);
- Diaskintest (Generium, Russian Federation); and
- C-TST (formerly known as ESAT6-CFP10 test, Anhui Zhifei Longcom, China)

Table 3.1.1 PICO questions for assessment of TBSTs

Population	Intervention	Comparator	Outcome
• PLHIV • Children aged <5 years • Household and other close contacts • Other at-risk groups: • Immune compromised (e.g. individuals receiving anti-TNF-α treatment – or dialysis; individuals undergoing preparation for an organ or haematological transplant; patients with silicosis; pregnant women; or individuals who are malnourished, have diabetes mellitus, use steroids or smoke tobacco) – High risk of prior TB exposure (e.g. prisoners, health workers, immigrants from high TB burden countries, individuals with CXR abnormalities, homeless people and people who use drugs, and inhabitants of high TB burden settings)[a] • BCG-vaccinated versus non-vaccinated (in identified groups at risk of TB infection – stratified or in combination, as appropriate)	• TBSTs: • Diaskintest • Cy-Tb • C-TST • Others	TST or IGRAs	• Efficacy of TPT based on diagnostic test results • Predictive value for progression to TB disease • Correlation with exposure gradient • Sensitivity and specificity[b] for TB infection[c] • Concordance with the TST • Concordance with IGRAs • Proportion started on TPT

The current recommendations are based on the evaluation of data for the tests that were included in the present evaluation. The findings cannot be extrapolated to other brand-specific tests; also, any new in-class technologies will need to be specifically evaluated by WHO.

Guidelines are disseminated through the WHO Global TB Programme (WHO/GTB) listservs to WHO regional offices, Member States, the Stop TB Partnership and other stakeholders (e.g. the Global Laboratory Initiative and the TB Supranational Reference Laboratory Network); they are also published on the websites of the WHO/GTB and Global Laboratory Initiative. The updated policy is incorporated into the WHO TB Knowledge Sharing Platform – an online reference resource for global TB policies and derivative products.

Recommendation

18. Mycobacterium tuberculosis antigen-based skin tests (TBSTs) may be used to test for TB infection.

(Conditional recommendation for the intervention, very low certainty of the evidence)

Evidence base

In 2021, WHO commissioned a systematic review of published and unpublished data on the new class of tests for TB infection not previously reviewed by WHO. The overarching policy question was: Should Mtb antigen-based skin tests (TBSTs) for TB infection be used as an alternative to the tuberculin skin test (TST) or WHO-endorsed interferon-γ release assays (IGRA) to identify individuals most at risk of progression from TB infection to TB disease? Based on the ov erarching policy question, four domains for evidence search and generation were included: diagnostic accuracy, safety, economic aspects and qualitative aspects. For each domain, specific population, intervention, comparator and outcome (PICO) or research questions were defined.

Domain 1 – *Diagnostic accuracy (PICO question)*: Do TBSTs have similar or better diagnostic performance than the TST or IGRAs to detect TB infection?

Domain 2 – *Safety*: Do TBSTs for TB infection cause more adverse reactions than the TST or IGRAs?

- What is the risk of adverse events of TBSTs compared with the current TST or IGRAs?
- Consider data on both local and systemic reactions graded by type, severity and seriousness, and stratified by subgroup.
- Compute relative risks where possible; however, if there is no control group receiving a comparator test, report frequency (%) of adverse events.

Domain 3 – *Cost–effectiveness analysis*: What are economic considerations of TBSTs compared with the TST or IGRAs?

- How large are the resource requirements (costs)?
- What is the certainty of the evidence on resource requirements (costs)?
- Does the cost–effectiveness of the intervention favour the intervention or the comparison?

Domain 4 – *User perspective*: What are end-user[4] views and perspectives on use of novel skin-based in vivo tests for TB infection use?

- Is there important uncertainty about, or variability in, how much end-users value the main outcomes?
- What would be the impact on health equity?
- Is the intervention acceptable to key stakeholders?
- Is it feasible to implement the intervention?

The certainty of the evidence of the pooled studies was assessed systematically through PICO questions, using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach *(2, 3)*. The GRADE approach produces an overall quality assessment (or certainty) of evidence, and has a framework for translating evidence into recommendations; also, under this approach, even if diagnostic accuracy studies are of observational design, they start as high-quality evidence.

GRADEpro Guideline Development Tool software *(4)* was used to generate summary of findings tables. The quality of evidence was rated as high (not downgraded), moderate (downgraded one level), low (downgraded two levels) or very low (downgraded more than two levels), based on five factors: risk of bias, indirectness, inconsistency, imprecision and other considerations. The quality (certainty) of evidence was downgraded by one level when a serious issue was identified and by two levels when a very serious issue was identified in any of the factors used to judge the quality of evidence. For data from the systematic reviews that were of a qualitative nature, the GRADE-CERQual tool was used. The tool examines the methodological limitations of the included studies, the coherence of each review finding, the adequacy of the data in support of a review finding and the relevance of the included studies to the review research questions; it is used to assess data quality from qualitative research studies.

Data synthesis was structured around the preset PICO question, as outlined above. The following web annexes provide additional information to evidence synthesis and analysis:

- ***Web Annex A.*** *Accuracy of Mycobacterium tuberculosis antigen-based skin tests: a systematic review and meta-analysis*
- ***Web Annex B.*** *Safety of Mycobacterium tuberculosis antigen-based skin tests: a systematic review and meta-analysis*
- ***Web Annex C.*** *GRADE profiles of Mycobacterium tuberculosis antigen-based skin tests*
- ***Web Annex D.*** *Cost–effectiveness of Mycobacterium tuberculosis antigen-based skin tests: a systematic review*
- ***Web Annex E.*** *Modelling for economic evidence for the use of Mycobacterium tuberculosis antigen-based skin tests*
- ***Web Annex F.*** *Qualitative evidence for the use of Mycobacterium tuberculosis antigen-based skin tests*
- ***Web Annex G.*** *Mycobacterium tuberculosis antigen-based skin tests: evidence-to-decision table*

Diagnostic accuracy

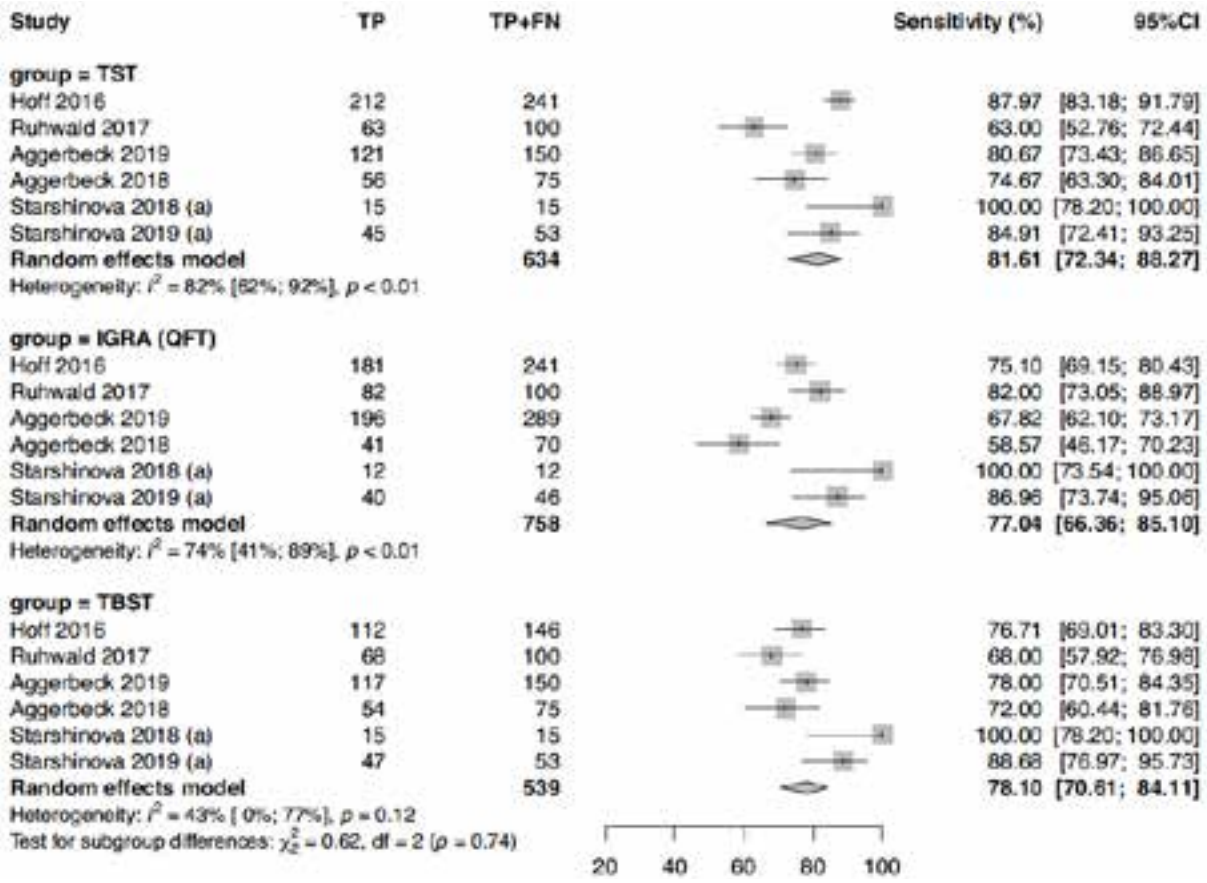
Diagnostic accuracy studies evaluating sensitivity, specificity and concordance (agreement) of TBSTs were identified. There were no identified studies on the efficacy of TPT based on diagnostic test results, on the predictive value for progression to TB disease or on the proportion started on TPT.

The assessed evidence for Cy-Tb and C-TST has included a manufacturer-recommended induration of at least 5 mm as the cut-off. According to the Diaskintest instructions for use, the presence of induration of any size is considered a positive response. However, the assessed evidence also included some studies for Diaskintest that used an induration of at least 5 mm as a cut-off, specified where applicable.

Sensitivity

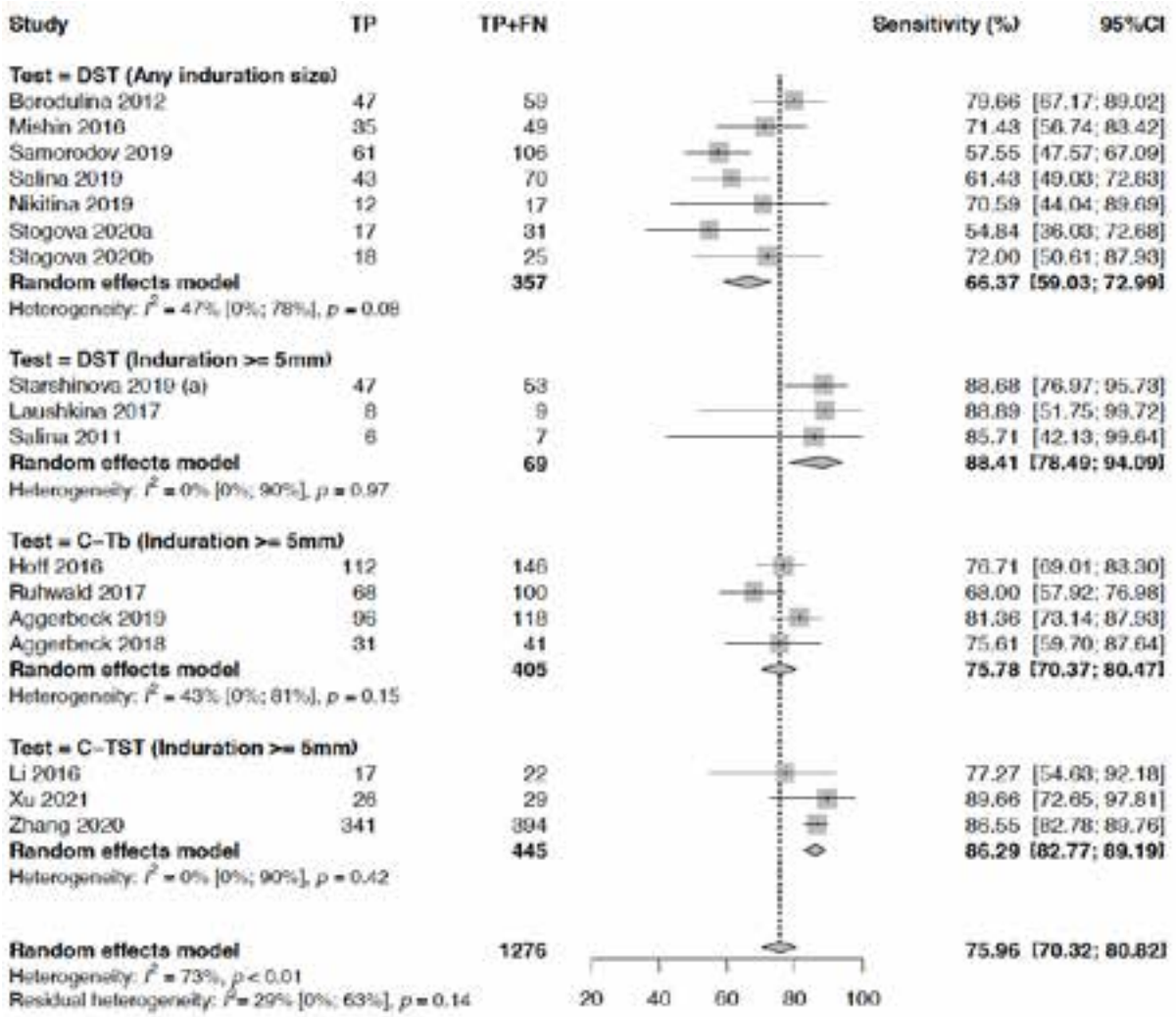
A total of 20 studies involving 1627 participants provided data for evaluating the sensitivity of TBSTs in people with microbiologically confirmed TB, which was used as a proxy for sensitivity to diagnose TB infection. Of these, six studies with 539 participants were head-to-head comparisons with the TST or IGRAs (or both); 17 studies included 1276 participants who were HIV-negative or whose HIV status was unknown; five studies included 317 PLHIV; and four studies included 34 participants aged under 18 years. Of the included studies, 14 evaluated Diaskintest, four Cy-Tb and three C-TST, as shown in **Figs. 3.1.1.1–3.1.1.2.**

Fig. 3.1.1.1 Sensitivity of TBSTs in head-to-head studies



The pooled sensitivity against the microbiological reference standard for TB disease in six head- to-head studies (**Fig. 3.1.1.1**) was 78.1% (95% confidence interval [CI]: 70.6–84.1%). The evidence was considered to be of high certainty and was not downgraded. Starshinova 2018 *(5)* and Starshinova 2019 *(6)* evaluated Diaskintest results with a cut-off of induration of at least 5 mm; the rest of the studies were head-to-head studies evaluating Cy-Tb. The assessed evidence for Cy-Tb included a cut-off of at least 5 mm in all studies. The TST cut-off was 5 mm for PLHIV and 15 mm for people who were HIV-negative in four studies *(7–10)*. Only studies on Diaskintest and Cy-Tb were included in this analysis.

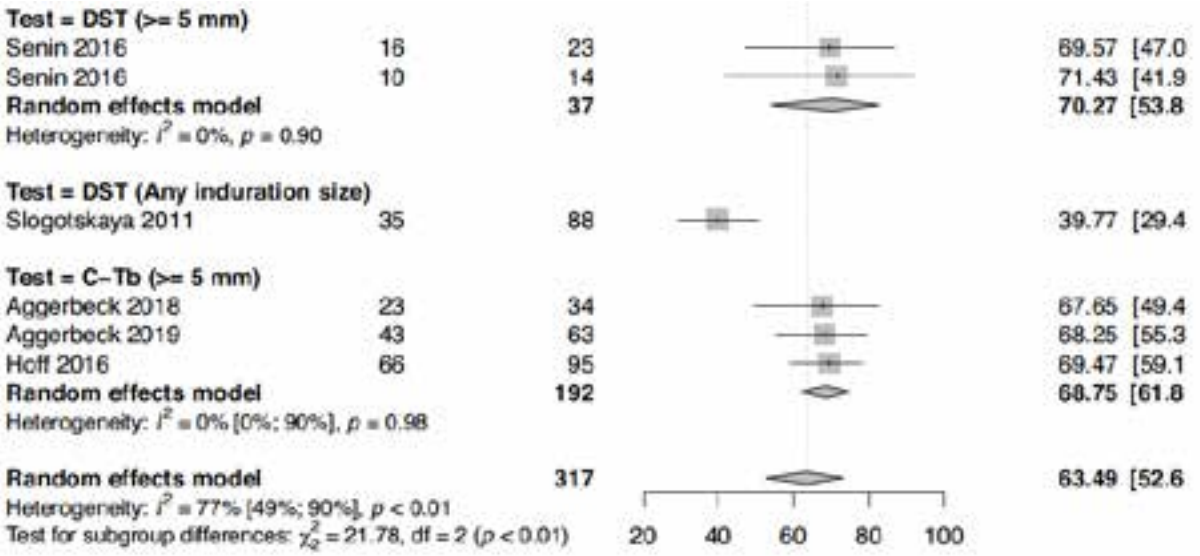
Fig. 3.1.1.2 Sensitivity of TBSTs in all studies in individuals with HIV-negative or unknown status



The pooled sensitivity in 17 studies presented in **Fig. 3.1.1.2** among participants who were HIV-negative or HIV status unknown was 76.0% (95% CI: 70.3–80.8%). The sensitivity estimates were lower in the studies using Diaskintest (any induration size). The reason for this is unclear; it may reflect different study populations or study quality. As a result, the evidence certainty was downgraded one level for inconsistency and another level for imprecision. Consequently, the certainty of the evidence was considered very low. Despite the manufacturer's recommendation to use induration of any size as a positive result, the sensitivity in studies using a Diaskintest result of at least 5 mm as the cut-off was more closely aligned with the other tests in the class, which all use a cut-off of at least 5 mm.

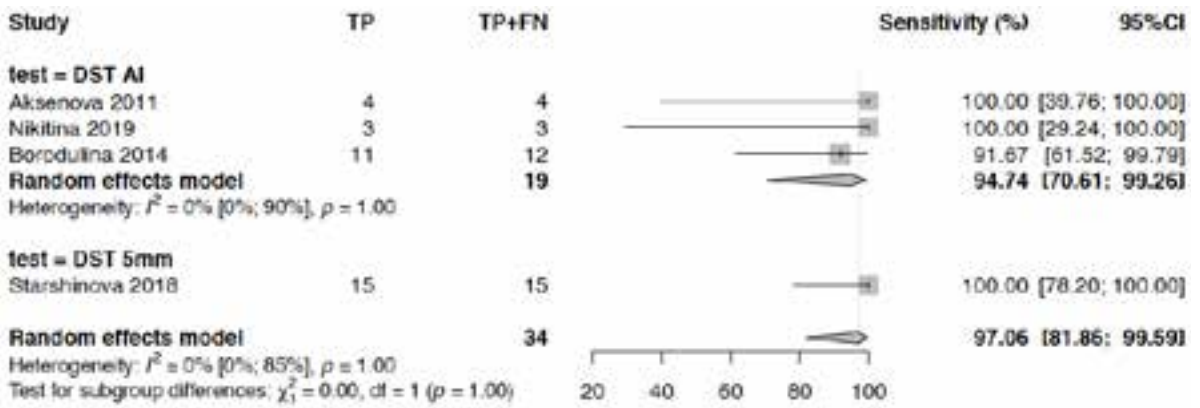
Risk of bias was considered serious due to the person having knowledge of the reference standards when interpreting the results of index tests. In most Diaskintest studies, the selection of participants and of the reference standard were unclear; hence, the certainty of the evidence was downgraded one level for risk of bias. The sensitivity ranged from 55% to 100% (the reasons for this heterogeneity are unknown); consequently, the certainty of the evidence was downgraded one level for inconsistency. Thus, the overall certainty of the evidence was considered low.

Fig. 3.1.1.3 Sensitivity of TBSTs in PLHIV



Only studies on Diaskintest and Cy-Tb were included in the analysis presented in **Fig. 3.1.1.3** The pooled sensitivity among PLHIV in five studies was 63.5% (95% CI: 52.6–73.2%). Risk of bias was considered serious for Diaskintest studies because of the person having knowledge of the reference standards when interpreting the results of index tests; hence, the evidence certainty was downgraded one level for risk of bias. The sensitivity estimates were lowest (39.8%) in the one study that used Diaskintest (any induration size). The reason for low sensitivity for Diaskintest (any induration size) is unclear, and the evidence certainty was downgraded one level for inconsistency. Certainty was also downgraded one level for imprecision. Consequently, the certainty of the evidence was considered to be very low.

Fig. 3.1.1.4 Sensitivity of TBSTs in children and adolescents

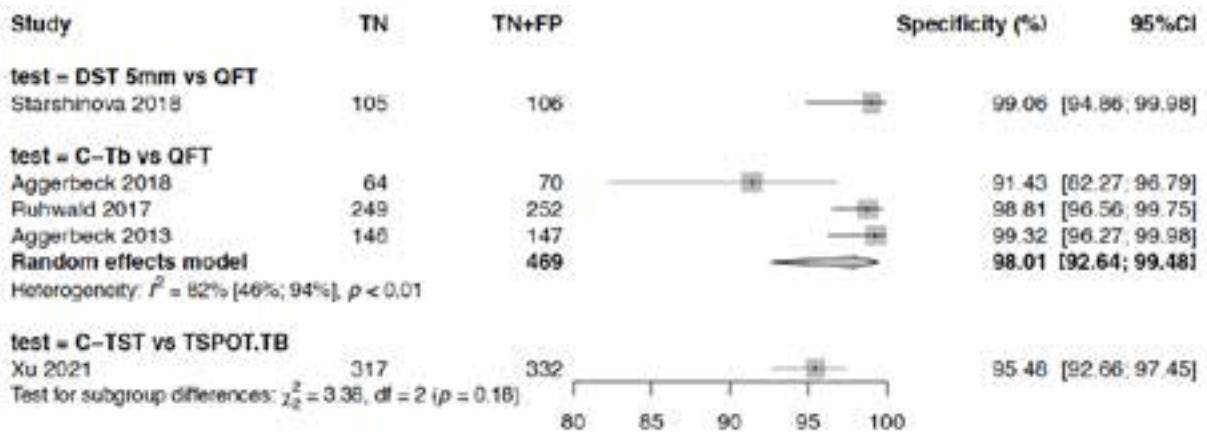


Sensitivity of TBSTs among children and adolescents is shown in **Fig. 3.1.1.4.** The pooled sensitivity in four studies for this class of tests was 97.1% (95% CI: 81.9–99.6%). The number of participants included in this analysis was small – only 34 participants in four studies; hence the studies were downgraded two levels for imprecision. Therefore, the evidence certainty was considered low. Only studies on Diaskintest were available for this analysis. Aggerbeck *(7)* estimated the sensitivity of Cy-Tb in 12 children and adolescents with TB, of whom only two were bacteriologically confirmed and were not included in the figure.

Specificity

A total of 14 studies involving 3792 participants provided data for evaluating specificity of TBSTs (including difference in specificity compared with the reference test); three of the studies included 1104 children and adolescents and three included 587 BCG-vaccinated individuals. Specificity was measured in healthy individuals with negative IGRA results. Difference in specificity was used as an alternative specificity measure, and was calculated as the difference in the proportion of negative results between TBSTs and the TST or IGRAs in healthy populations.

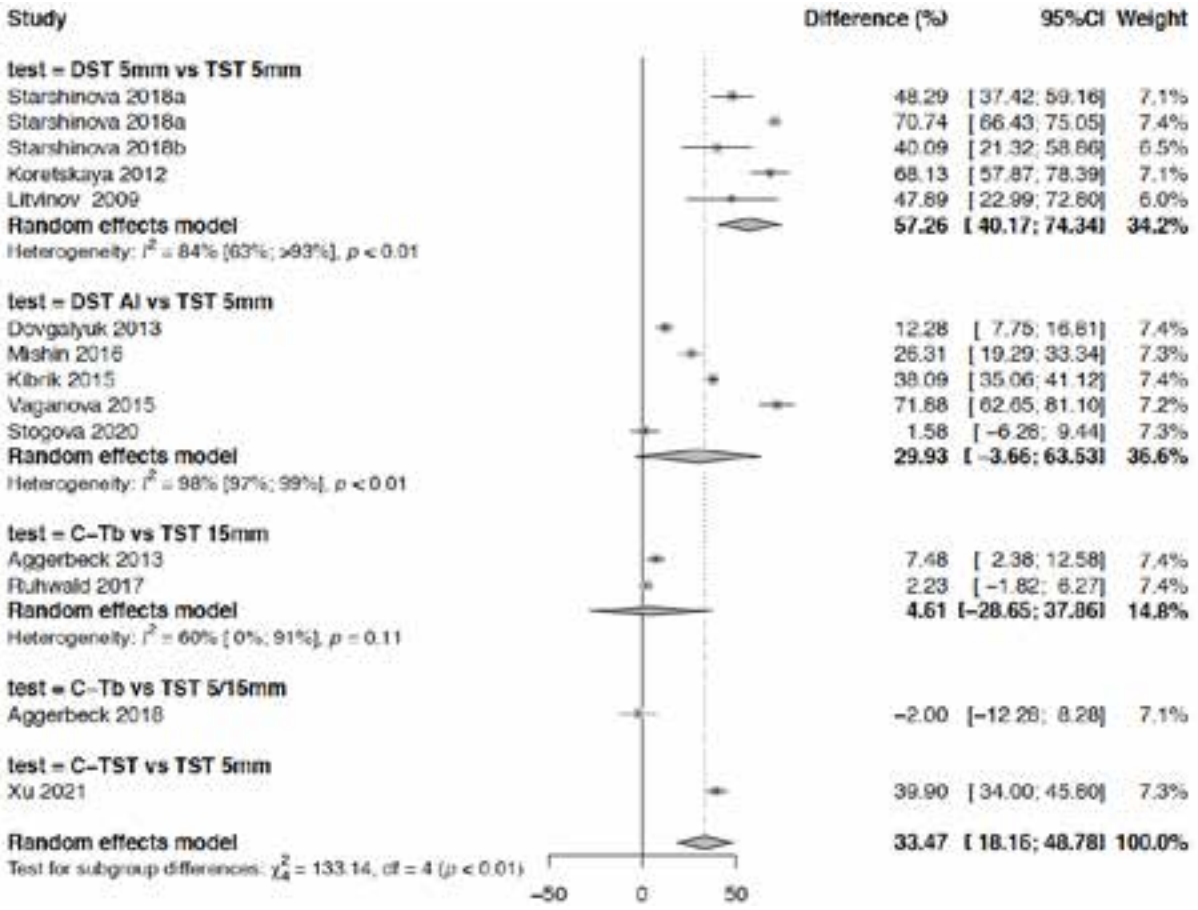
Fig. 3.1.1.5 Specificity in healthy individuals with negative IGRA results



The specificity assessed in the five studies presented in **Fig. 3.1.1.5** was high for all three tests in the TBST class. For Diaskintest it was 99.1% (95% CI: 93.6–99.9%), as compared with QFT; for Cy-Tb it was 98.0% (95% CI: 92.6–99.5%), as compared with QFT; and for C-TST it was 95.5% (95% CI: 92.6–97.3%), as compared with T-Spot. During the GDG meeting, participants noted that – considering the totality of evidence (which included studies of very low quality) – the overall certainty of the evidence on tests' effects for specificity was very low.

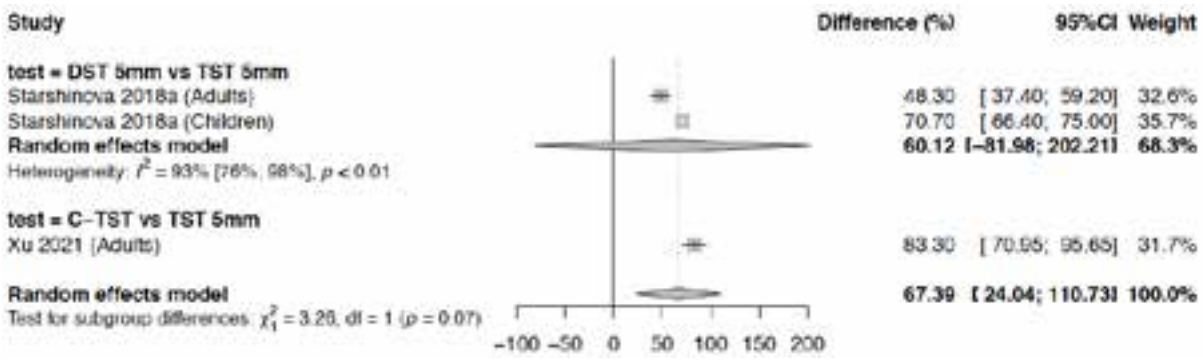
Specificity in children and adolescents (2 studies, 176 patients), as determined in individuals with negative IGRA results, was high. For Diaskintest with a cut-off of at least 5 mm it was 99.1% (95% CI: 94.9–99.9%), as compared with QFT, and for Cy-Tb it was 91.4% (95% CI: 82.2–96.1%), as compared with QFT. Specificity in BCG-vaccinated individuals (3 studies, 292 patients), as determined in healthy individuals with negative IGRA results, was also high, being 97–99% (depending on the test), with a pooled value of 99.0% (95% CI: 96.9–99.7%). More details can be found in **Web Annex A**.

Fig. 3.1.1.6 Difference in specificity – TBSTs versus the TST



The overall pooled difference in specificity in 14 studies (**Fig. 3.1.1.6**) comparing TBSTs and the TST was 33.5% (95% CI: 18.2–48.8%) higher for TBSTs. In studies of Diaskintest and C-TST done in high TB incidence settings, the differences in specificity were higher for Diaskintest versus the TST (with both tests having a cut-off of at least 5 mm) (57.3%, 95% CI: 40.2–74.3%), than with Diaskintest (any induration size) versus the TST with a cut-off of at least 5 mm (29.9%, 95% CI: –3.66–63.5%). For C-TST versus the TST with a cut-off of at least 5 mm, the difference in specificity was 39.9% (95% CI: 34.0–45.8%). In contrast, in studies of Cy-Tb undertaken in low TB incidence settings, the difference in specificity between Cy-Tb and the TST was less prominent, but was greater with the TST with a cut-off of at least 15 mm (4.61%, 95% CI: –28.6–37.9%) than with the TST with a cut-off of 5 or 15 mm (–2.0%, 95% CI: –12.3–8.3%). The difference may be explained by the background level of BCG in the study populations or by the cut-offs that were used. **Fig. 3.1.1.7** has more details on the specificity of TBSTs versus the TST in BCG-vaccinated people. Overall risk of bias was considered serious because test allocation by arm was not blinded in any of the studies except those for Cy-Tb. In most Diaskintest studies, the selection of participants and the diagnosis of the reference standard were unclear. The certainty of the evidence was therefore downgraded one level for risk of bias. The difference in specificity ranged from –2% to 72%; hence, the certainty of the evidence was downgraded one more level for inconsistency. Consequently, the certainty of the evidence for difference in specificity between TBSTs and the TST was low.

Fig. 3.1.1.7 Difference in specificity – TBSTs versus the TST in BCG-vaccinated population



Two studies (three analyses) provided data on difference in specificity in BCG-vaccinated populations, which was even higher for this population than in populations where only some people had received BCG vaccination; the pooled difference in specificity was 67.4% (95% CI: 24.0–110.7%). Overall risk of bias was considered serious because test allocation by arm was not blinded; hence, the certainty of the evidence was downgraded one level for risk of bias. The CI was broad, ranging from 24.0% to 110.7%, so the certainty of the evidence was downgraded one more level for imprecision. Consequently, certainty of the evidence for difference in specificity between TBSTs and the TST in BCG-vaccinated populations was low.

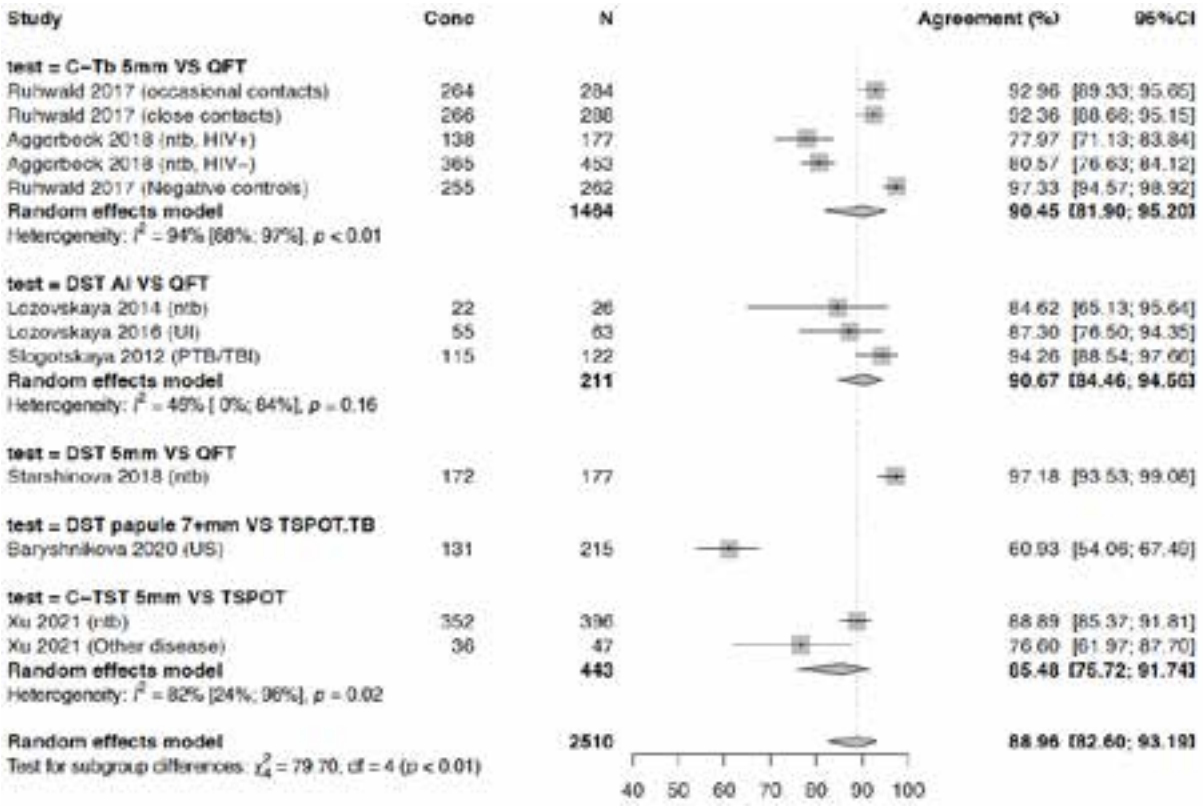
The pooled difference in specificity in six studies comparing TBSTs and IGRAs was low, at 2.3% (95% CI: –1.6–6.2%), meaning that TBSTs were similar to IGRAs in terms of specificity.

Agreement

Overall, 16 studies involving 3198 participants (among which four studies with 1307 participants recruited people aged under 18 years) were included to assess agreement of the index tests with comparator tests (the TST or IGRAs, or both).

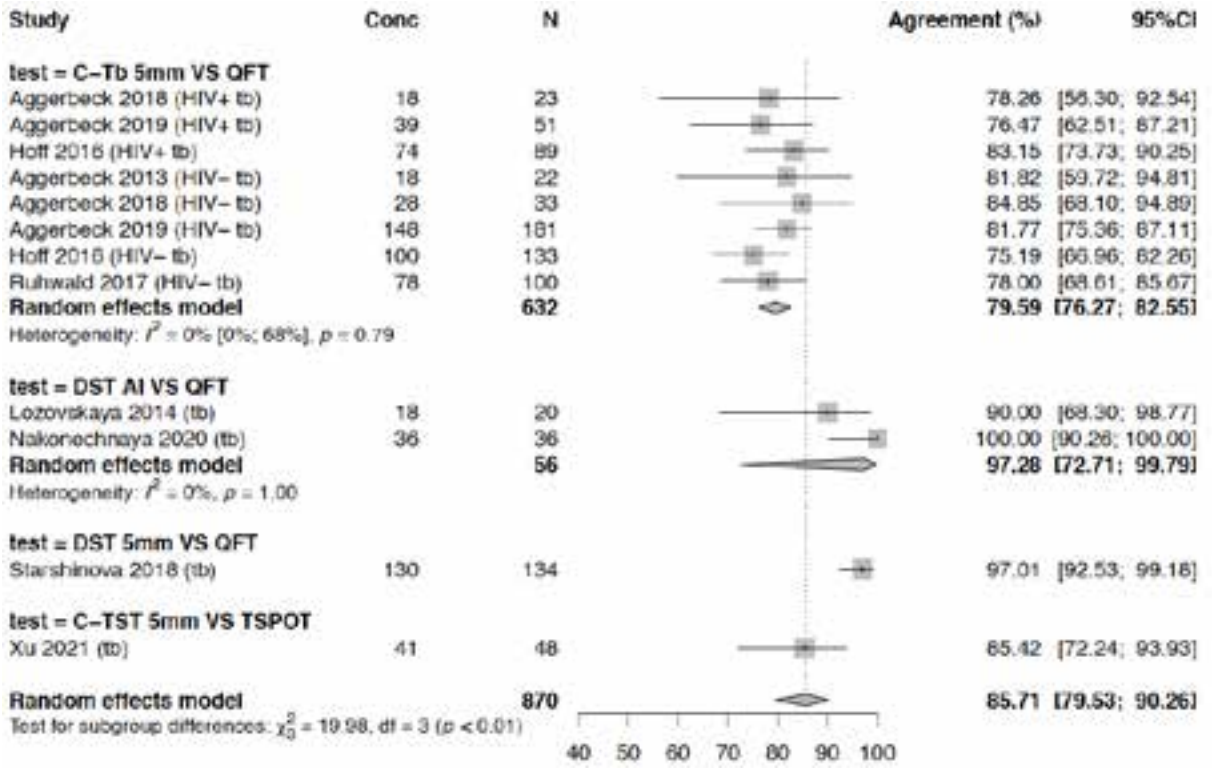
In participants without TB disease, agreement was high (≥90%) for Cy-Tb and Diaskintest – (any induration size) and Diaskintest 5 mm induration – compared with QFT (**Fig. 3.1.1.8**). Agreement was slightly lower at 85.5% (95% CI: 75.7–91.7%) for C-TST compared with T-Spot. In one study, which evaluated Diaskintest with induration of at least 7 mm compared with T-Spot, the agreement was considerably lower, at 60.9% (95% CI: 54.3–67.2%). Risk of bias was considered serious because the allocation of tests was not blinded in five studies; hence, certainty of the evidence was downgraded one level for risk of bias. Agreement ranged widely (from 61% to 97%) for various tests and studies, so the certainty of the evidence was downgraded one level for inconsistency. Consequently, certainty of the evidence for agreement between TBSTs and IGRAs was low.

Fig. 3.1.1.8 Agreement of TBSTs versus IGRAs in all studies including participants without active TB



In participants with TB disease, high agreement between TBSTs and IGRAs as the comparator (85.7%) was observed (**Fig. 3.1.1.9**). Some variability in agreement was seen between the different tests: 79.6% (95% CI: 76.3–82.6%) for Cy-Tb 5 mm compared with QFT; 97.3% (95% CI: 72.7–99.8%) for Diaskintest (any induration size) compared with QFT; and 97.0% (95% CI: 92.3– 98.9%) for DST 5 mm induration compared with QFT. Agreement was slightly lower at 85.4% (95% CI: 72.4–92.9%) for C-TST compared with T-Spot. Risk of bias was considered serious because, in four studies, the allocation of tests by arm was not blinded; hence, the certainty of the evidence was downgraded one level for risk of bias. The agreement ranged from 75% to 100% for various tests and studies, so certainty of the evidence was downgraded one level for inconsistency. The overall certainty of the evidence for agreement between TBSTs and IGRAs in people with TB disease was considered low.

Fig. 3.1.1.9 Agreement of TBSTs versus IGRAs in all studies including people with active TB



Safety

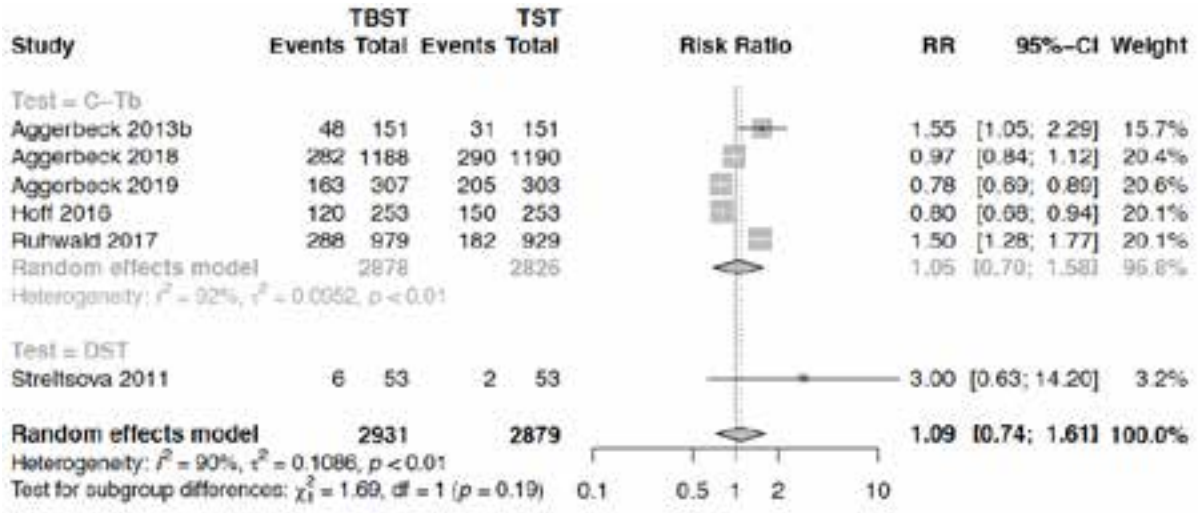
A systematic review of studies reporting the outcomes of interest, including local reactions – that is, injection site reactions (ISR) and systemic adverse events from TBSTs – was undertaken. The following databases were searched for studies from inception until 30 July 2021: Medline, Embase, e-library, the Chinese Biomedical Literature Database and the China National Knowledge Infrastructure Database. The test manufacturers were contacted for individual studies, and studies were identified through a public call for data by WHO. Longitudinal and case–control studies reporting adverse events of the index tests alone or compared with recognized comparator tests (e.g. QFT, T-Spot and the TST) in humans were included with no language restrictions. Screening of titles and abstracts as well as full-text articles and the assessment of quality were performed by two investigators in duplicate. A meta-analysis was conducted using a random-effects model, and studies that were considered to be clinically homogenous were pooled.

Overall, seven studies for Cy-Tb, five for C-TST and 11 for Diaskintest were identified. Characteristics of studies were as follows:

- Cy-Tb: clinical trials – three studies in South Africa and four in Europe. Most participants were adults; in studies in South Africa, 20–40% of participants were PLHIV. Five of seven studies included random allocation of Cy-Tb versus the TST into two arms and thus allowed comparison of ISR. All five studies were included in the pooled evidence assessment on any ISR. Only one study provided comparable data on systemic reactions. This study was also included in the pooled evidence assessment on systemic reactions.

- C-TST: all five studies were conducted in China and included only HIV-negative adults. All of them included non-random allocation of C-TST versus the TST into two arms; thus, no study evaluating C-TST was included in the pooled evidence assessment on any ISR. Also, no studies including any comparable data on systemic reactions were available.
- Diaskintest: cross-sectional studies using routinely collected data mostly in the Russian Federation, and one in Ukraine, including various populations (adults, children and adolescents – healthy, contacts of TB patients and with TB). Two studies on Diaskintest provided comparable data on ISR; however, one of them provided no information about the number of participants who experienced any ISR; thus, only one study on Diaskintest was included in the meta-analysis.

Fig. 3.1.1.10 Any injection site reactions



Proportion of PLHIV: Aggerbeck 2018 (*7*) (25%), Aggerbeck 2019 (*8*) (20%); Hoff 2016 (*10*) (39.5%). Other studies included HIV-negative individuals. Aggerbeck 2018 (*7*) included children aged under 5 years (20%) and aged 5–17 years (31%); Ruhwald 2017 (*9*) included children aged under 5 years (3.5%) and aged 5–17 years (8.8%). Other studies included adults. Hoff 2016 (*10*), Aggerbeck 2019 (*8*) and Streltsova 2011 (*11*) included people with TB only.

The pooled risk of any ISR due to Cy-Tb (n=2878, 5 studies) and Diaskintest (n=53, 1 study) presented in **Fig. 3.1.1.10** was not significantly different from the TST (risk ratio [RR] 1.09; 95% CI: 0.74–1.61). The risk of any systemic reaction was only analysable in one study (Cy-Tb) that allowed such comparison, and was not significantly different from the TST (RR 0.84; 95% CI: 0.60–1.10). The Diaskintest study was considered to have high risk of bias, while the overall certainty of evidence from the randomized controlled trials for any ISR was judged as high. For any systemic reactions, overall certainty of evidence was judged to be moderate because of the small sample size and wide CI.

Following the request from GDG members for the post-marketing surveillance data for Diaskintest, the following data were reported by the manufacturer: in 2019–2021, over a 55.7 mln Diaskintest tests were done, with 27 serious adverse effects and 30 non-serious adverse effects. Based on the totality of data, the GDG rated the certainty of evidence as high.

Based on the data presented at the GDG meeting, it was concluded that the safety profile of novel TBSTs is similar to that of the TST, and is associated with mostly mild ISR such as itching and pain. From the reviewed studies, there appears to be no safety signal that might affect the choice between specific TBSTs and the TST. However, the group also noted that this was not a full safety review covering product safety, animal or preclinical studies. Regulatory assessment for safety is needed before any of the TBST products are implemented.

Cost and cost–effectiveness analysis

Two reviews following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were carried out to look at costs and cost–effectiveness of:

- novel TBST, such as Diaskintest, C-TST and Cy-Tb (primary review); and
- TST and IGRA tests (secondary review).

The articles searched were those presenting economic evaluations of the diagnostic tests (costs and cost–effectiveness) using a health provider perspective and related to TB infection in humans. The articles reviewed were those written in English, Chinese or Russian languages, and published in Medline, OVID, Chinese Biomedical Literature, China National Knowledge Infrastructure and Russian e-library databases. Quality of studies was assessed using Drummond's checklist.

In addition, a Markov-chain model was developed for the purposes of the GDG meeting, to study the cost–effectiveness of TBSTs versus the currently available tests, the TST and IGRAs. When simulating a cohort of individuals transitioning among different states and steps along the TB cascade of care, the model took into consideration the following parameters:

- prevalence of TB infection in TB-negative individuals, percentage;
- people completing treatment after initiation following a positive TB infection result, percentage;
- people not initiating treatment after testing positive for TB infection, percentage;
- people interrupting treatment after initiation following a positive TB infection test result, percentage;
- progression from TB infection to active TB, probability;
- efficacy of TB infection treatment;
- active TB treatment coverage;
- recovery from active TB (treated + untreated);
- death from active TB (treated + untreated);
- probability of a true positive test result if the patient has TB infection (sensitivity); and
- probability of a true negative test result if the patient does not have TB infection (specificity).

Model parameters, unit costs and estimates of diagnostic test accuracy were sourced from the literature, including from the systematic reviews mentioned above. The manufacturers of novel TBSTs were also contacted to source costs of the new tests. However, only Generium, the manufacturer of Diaskintest, provided estimated test costs, including delivery costs, for different delivery volumes. Consequently, the modelling study focused on Diaskintest as the representative of the TBST class of tests.

The model was parameterized to three countries: Brazil, South Africa and the United Kingdom. Three testing strategies were considered in this analysis: Diaskintest (index); the TST; and

QuantiFERON-TB IGRAs, either Gold In-Tube or Gold Plus (comparator tests). Outcomes reported included unit cost (in US dollars)[5] per patient, incremental cost–effectiveness ratio (ICER) and incremental net benefit per quality-adjusted life year (QALY) gained. Unit costs considered in each country included test kit, staff time, laboratory and disposable costs. Costs were considered from a health system perspective and did not reflect patient or societal costs.

Given that only information on Diaskintest was available, a univariate sensitivity analysis on TBST unit costs and a comparison of the results of the three strategies was performed to identify possible maximum unit costs of new TBSTs, for the strategy to remain cost saving or cost-effective, but without specifying a particular type of TBST.

The conclusions were based on the predefined research questions outlined below.

How large are the resource requirements (costs)?

In the eight studies that assessed Diaskintest, most estimated a cost of $1.60 per test. One study evaluated the unit costs considering staff time, consumables and laboratory costs, resulting in a cost of $5.07. This study, using the same costing factors, also estimated the unit cost of C-TST as $9.96. The 29 studies on IGRAs or the TST (or both) estimated an average cost of $37.84 for the TST and $89.33 for IGRAs (accounting for different ingredients). The cost–effectiveness of the tests varied among and within risk groups, with no clear economic consensus around the cost–effectiveness of comparison tests.

What is the certainty of the evidence of resource requirements (costs)?

Based on Drummond's scores, the quality of studies that have assessed cost–effectiveness of C-TST and Diaskintest in this review was concerning; only one out of eight studies was of high quality. However, the quality of the studies that assessed cost–effectiveness of the TST and IGRAs was generally high.

Does the cost–effectiveness of the intervention favour the intervention or the comparison?

Based on the systematic review results, there was insufficient evidence regarding both the cost and cost–effectiveness of novel TBSTs. The quality of the studies was concerning according to the Drummond's checklist for economic evaluations. More high-quality studies are needed that consider different health settings and risk populations to estimate the cost–effectiveness and the likely economic impact of these tests.

Results of the Markov-chain model conducted for the purposes of the GDG meeting concluded that, in Brazil, Diaskintest is cost saving compared with the TST and IGRAs. Compared with the TST, Diaskintest is cost saving at $5.60, with an incremental gain of 0.02 QALYs per patient. Compared with IGRAs, Diaskintest is cost saving at $8.40, with an incremental gain of 0.01 QALYs. In South Africa, Diaskintest is more cost saving than the TST or IGRAs. Compared with the TST, Diaskintest is cost saving at $4.39, with an incremental gain of 0.02 QALYs, and compared with IGRAs, it is cost saving at $64.41, with an incremental gain of 0.01 QALYs. In the United Kingdom, Diaskintest is cost saving compared with the TST but not with IGRAs. Compared with the TST, Diaskintest is cost saving at $73.33, with an incremental gain of 0.04 QALYs; however, compared with IGRAs, Diaskintest showed an increase in cost of $15.80 but still an incremental gain of 0.03 QALYs.

In summary, the modelling and univariate sensitivity analysis results show that, in Brazil and South Africa, use of Diaskintest would potentially save costs per patient and result in greater health gains (QALYs per patient) compared with the TST and IGRAs. In the United Kingdom, Diaskintest results in health gains but is more expensive in terms of expected cost per patient than IGRAs. Our results also show that, in Brazil and South Africa, IGRAs are more costly to implement than the TST but would result in health gains. However, in the United Kingdom, IGRAs are cheaper to implement and are more cost-effective than the TST.

User perspective

User perspectives on the value, feasibility, usability and acceptability of diagnostic technologies are important in the implementation of such technologies. If the perspectives of laboratory personnel, clinicians, patients and TB programme personnel are not considered, the technologies risk being inaccessible to and underused by those for whom they are intended.

To address questions related to user perspective, the following activities were undertaken:

- Two systematic reviews, which synthesized the qualitative research evidence on end-user values and preferences for the use of specific TBSTs for TB infection, compared with existing tests (IGRAs and the TST). Study quality and confidence in the evidence were evaluated in accordance with the GRADE-CERQual.
- Twenty semi-structured interviews with a diverse range of clinicians, laboratory staff, programme officers and individuals living with TB infection (referred to as "consumers" throughout this report).
- A discrete choice experiment (DCE) survey, drawing from themes derived in systematic reviews and semi-structured interviews. DCE methodology was used to elicit stated values and preferences from participants (end-users) without directly asking them to state their preferred options.

Four studies were identified that met the inclusion criteria for both systematic reviews. From the review on specific TBST, only one data source was identified (from the Russian Federation), and that came from a WHO public call for data relating to the feasibility and acceptability of TBSTs. Participants were parents of children and adolescents with TB infection. From the review on current IGRAs and the TST, three peer-reviewed articles were found to meet the inclusion criteria; these three papers were from the Netherlands, South Africa and the United States of America (USA). Participants included a range of health professionals involved in TB care (Netherlands, South Africa and USA) and PLHIV (South Africa). The overall confidence in the quality of the evidence from the studies was low to moderate based on the GRADE-CERQual assessments, because the data lacked richness, with most studies reporting only summaries of participant quotes or limited direct quotes. All studies were conducted on specific subgroups (e.g. PLHIV, or parents of children and adolescents with TB infection).

For user interviews, 20 participants were recruited – 13 were TB health care providers (8 from low- and middle-income countries [LMIC]) and seven were people affected by TB (3 from LMIC). Health care providers included programme executives and decision-makers, public health practitioners and advocates, physicians, researchers and laboratory technicians, and a nurse.

For DCE, a total of 234 participants completed this activity (186 providers and 48 consumers). Overall, 59% of respondents were female and 56% were aged 36–55 years; the main countries in which respondents were based were India (26%), the USA (16%), South Africa (9%), Pakistan (8%) and Zimbabwe (7%).

The conclusions were based on the predefined research questions outlined below.

Is there important uncertainty about or variability in how much end-users value the main outcomes?

Qualitative data from the systematic reviews and end-user interviews, and quantitative data from the DCE indicated that health care consumers and providers had similar values and preferences in terms of TB infection tests. Key end-user values included test accuracy, convenience, positive patient experience, cost and resource requirements. In particular, end-users valued tests with high accuracy such as TBST and IGRAs (i.e. low false positive and false negative rates), because they reduce the risk of downstream consequences associated with false positive and false negative results (e.g. anxiety, and the need for additional testing or unnecessary treatment). End-users also preferred having a test that was convenient to administer and access. This included valuing tests that can be accessed in a community or primary care setting, that do not require follow-up visits to read test results, and that can be administered without the need for additional systems or infrastructure to be developed. These findings were initially identified from themes emerging from the systematic reviews and end-user interviews, and were confirmed by the DCE findings.

From the qualitative data from the reviews and interviews, all TB infection test options were found to have strengths and limitations in terms of convenience. End-users valued a positive consumer experience. This meant that tests with fewer psychological effects (e.g. anxiety, stigma and stress) and physical consequences (e.g. discomfort) were preferred. Tests that were more accurate tended to be associated with better consumer experience, although some aspects of consumer experience were worse in skin tests (e.g. stigma from the welt and discomfort) compared with non-skin-based tests. Low-cost tests were generally preferred due to greater accessibility in resource-limited contexts (e.g. TBST and the TST). Tests with lower resource requirements were preferred in resource-limited settings (e.g. TBST and the TST); however, this appeared to be less of a consideration in high-income countries. End-users showed a preference towards TB infection tests that used existing infrastructure in their health care setting. Data from the DCE confirmed that not requiring an in-person follow-up appointment and not requiring specialist staff or equipment to interpret or administer the test were important end-user preferences for TB testing.

What would be the impact on health equity?

Qualitative evidence from reviews and end-user interviews indicates that specific TBSTs are unlikely to create any new equity issues. Rather, TBSTs are likely to improve health equity through the provision of a more accurate, low-cost test for resource-limited settings where the TST is already in use. Moreover, their portability and low cost make them suited to use in large-scale screening programmes in vulnerable, hard-to-reach communities. However, it is possible that TBSTs may not affect health equity in low-resource settings that do not already use the TST, because there are barriers to accessing skin and other health care tests in these

settings, which would need to be addressed first, regardless of the type of TB test available. In terms of test accessibility, the data from the DCE found that consumers had a strong preference for testing in the community and primary care settings, compared with hospital locations; this finding could have health equity implications.

Is the intervention acceptable to key stakeholders?

Qualitative data from systematic reviews and end-user interviews suggest that TBSTs were perceived to have greater specificity and sensitivity than the TST. Having greater test accuracy was deemed desirable to avoid the negative consequences of false positives or negatives. However, TBSTs were expected to have many of the same limitations as skin tests in terms of patient experience (e.g. the need for a return visit, discomfort, a welt on the arm and stigma) compared with IGRAs. IGRAs were deemed the preferred test option in countries that already have IGRAs in use, because the required supporting infrastructure is already in place, and because TBSTs would have comparable accuracy and performance, thus would not add value. There were also broader concerns about skin tests because these tests were viewed as a dated, basic technology that is subject to human error and interpretation. Suggestions for improving the acceptability of TBSTs included careful communication during the implementation of this test, with endorsement by health care providers and organizations (e.g. WHO). Data from the DCE found strong and consistent preferences among both health care providers and consumers for tests that minimize false positive and false negative results. The DCE also found that consumers had a strong preference for testing in the community and primary care settings compared with hospital locations.

Is the intervention feasible to implement?

Findings from the qualitative evidence synthesis (reviews and end-user interviews) support the feasibility of use of TBSTs, but only in settings where the TST is already in use, and the required resourcing and training is already in place. TBST are likely to be low-cost, portable tests that can be well-suited for low-resource health care settings, which may not be able to support IGRAs owing to the greater cost and resources required to implement IGRAs. However, if health care settings already have the necessary infrastructure in place to implement IGRAs, then that is a more feasible test option than any skin tests because IGRAs do not require a return visit to read the result (a step where patients may be lost to follow-up). Results from the DCE found that not requiring an in-person follow-up appointment, or specialist staff or equipment to interpret or administer the test, were important preferences for TB testing that would influence feasibility. There was some suggestion that providers preferred more expensive tests (when offered a choice based on a hypothetical cost of $50 compared with $25), although test cost was the least important determinant of test choice.

Implementation considerations

Considerations for implementation were as follows:

- regulatory approval from national regulatory authorities or other relevant bodies is required before implementation of in vivo diagnostic tests;
- appropriate communication on the new class of tests is necessary, highlighting the difference between the TST and TBSTs;
- implementation of TBSTs requires a cold chain;

- well-trained skilled staff are needed to administer and interpret this class of tests;
- multiuse vials will require effective operational planning and batching; hence, single-use vials or vials with fewer doses to match daily needs are preferred;
- procurement and stock management aspects will have to be considered, as with implementing any new class of tests;
- because the reading of the TBST results requires a second patient visit, linkage to care requires reinforcement, to decrease loss to follow-up;
- global market availability and necessary volumes of the new class of tests must be considered; and
- measurement of the TBST reaction size and interpretation must be standardized.

Monitoring and evaluation

Factors that will require monitoring and evaluation are as follows:

- adverse event monitoring is a gap with the current TST use; thus, recording and reporting systems for results and adverse events need to be introduced when implementing the new tests; and
- there is a need to monitor the linkage between results of the new class of the tests and number of people placed on TPT.

Research priorities

Research priorities are as follows:

- specificity of Diaskintest and C-TST in populations with a low prevalence of TB infection, and direct head-to-head comparisons of all three TBST;
- assessing the barriers for implementation and patient access;
- additional accuracy studies on high-risk groups: children aged under 5 years, children (aged 5–10 years) and adolescents (aged 10–18 years), PLHIV, prisoners and migrants;
- studies evaluating the epidemiologic and economic impact of TBST use in the TB infection diagnosis and TPT cascade;
- longitudinal studies to assess the predictive value for active TB compared with current tests;
- economic studies (e.g. cost and cost–effectiveness of TBSTs under different scenarios); and
- studies evaluating the use of digital tools for reading of results, to avoid return patient visits.

3.2. TB skin tests and interferon gamma release assays for the diagnosis of TB infection

Testing for TB infection increases the certainty that individuals targeted for treatment will benefit from it. However, there is no gold-standard test to diagnose TB infection. Both currently available tests – the TST and IGRAs – are indirect and require a competent immune response to identify people infected with TB. A positive test result by either method is not by itself a reliable indicator of the risk of progression to active disease. This section discusses the evidence and the recommendations for TB infection testing.

Recommendation

19. Either a tuberculin skin test (TST) or interferon-gamma release assays (IGRAs) can be used to test for TB infection.

(Strong recommendation, very low certainty of the evidence)

Justification

A systematic review has informed the comparison of the predictive performance of IGRAs and the TST for identifying incident active TB in countries with a TB incidence of more than 100 per 100 000 population *(12)*. Only studies in which the TST was compared with IGRAs in the same population (i.e. "head-to-head" studies) were included. Relative risk ratios for TB for people who tested positive and those who tested negative with the TST and IGRAs were estimated.

Five prospective cohort studies were identified, with a total of 7769 participants. The pooled risk ratio estimate for the TST was 1.49 (95% CI: 0.79–2.80), and for IGRAs was 2.03 (95% CI: 1.18–3.50). Although the estimate for IGRAs was slightly higher than that for the TST, the 95% CIs for the estimates for the TST and IGRAs overlapped and were imprecise.

The GDG concluded that the comparison of the TST and IGRAs in the same population does not provide strong evidence that one test should be preferred over the other for predicting progression to active TB disease. The TST may require significantly fewer resources than IGRAs and may be more familiar to practitioners in resource-limited settings; however, recurrent global shortages and stock-outs of the TST reduce prospects for the scale-up of this test and for the programmatic management of TPT. The GDG also noted that equity and access could affect the choice and type of test used. The preferences of people to be tested and programmes depend on several factors, such as the requirement for an adequately equipped laboratory (e.g. for IGRAs) and possible additional costs for people being tested (e.g. for travel) and programmes (e.g. for infrastructure and testing). The GDG strongly recommended the two tests as equivalent options, with relatively similar advantages and disadvantages. The GDG stressed that the global shortage of the TST should be addressed urgently, and called for more investment into research on novel tests for TB infection with better predictive value. The GDG cautioned that imperfect performance of these tests can lead to false negative results, particularly in young children and immunocompromised individuals such as PLHIV with low CD4 counts. The GDG noted the importance of the tests to identify recent conversion from negative to positive, particularly among contacts of people with pulmonary TB, which is good practice when initiating TPT. Nevertheless, recent studies among health care workers in the USA tested serially for TB infection showed that conversions from negative to positive and reversions from positive to negative are more commonly identified with IGRAs than with the TST *(13)*. Thus, clinical judgement must still be used to interpret the results of serial TB infection tests.

The evidence reviewed and the recommendations given apply only to the use of the two commercially available IGRAs (QuantiFERON-TB Gold In-Tube and T-Spot.

Evidence base

PICO question

Could IGRA be used as an alternative to the TST, to identify individuals most at risk of progression from TB infection to active TB in high TB incidence settings?

Evidence on intervention effect

Five prospective cohort studies were identified, with a total of 7769 participants; four of the studies were newly identified. Three of the studies were conducted in South Africa and two in India *(14–18)*. The studies included PLHIV, pregnant women, adolescents, health care workers and household contacts. The pooled risk ratio estimate for the TST was 1.49 (95% CI: 0.79–2.80), and for IGRAs was 2.03 (95% CI: 1.18–3.50). Although the estimate for IGRAs was slightly higher than that for the TST, the 95% CIs for the estimates for the TST and IGRAs overlapped and were imprecise. Furthermore, there was limited evidence for the predictive utility of the tests in specific at-risk populations.

Cost–effectiveness

IGRA testing is more costly than the TST and requires appropriate laboratory services. TST testing is less costly and can be performed in the field, but it requires a cold chain, two health care visits and training in intradermal injection, reading and interpretation. The incremental cost–effectiveness of IGRAs and the TST appears to be influenced mainly by their accuracy.

User perspective

The preferences of people to be tested and programmes depend on several factors, such as the requirement for an adequately equipped laboratory (e.g. for IGRAs) and possible additional costs for people being tested (e.g. for travel) and programmes (e.g. for infrastructure and testing).

Implementation considerations

Where it is feasible, TB infection testing is desirable to identify individuals at highest risk for developing active TB. However, it is not required in PLHIV or in household contacts aged under 5 years. In HIV-negative household contacts aged 5 years and older, and in other risk groups, TB infection tests are recommended, but their unavailability should not be a barrier to treating people who are judged to be at higher risk. The GDG noted that the availability and affordability of the tests could determine which TB infection test is used. Other considerations include the structure of the health system, feasibility of implementation and infrastructure requirements.

Operational difficulties should be considered in deciding which test to use. For example, IGRAs requires phlebotomy, which can be difficult, particularly in young children; they also require laboratory infrastructure, technical expertise and expensive equipment, and their sensitivity is reduced in children aged under 2 years and PLHIV. However, only a single visit is required to do an IGRA test (although patients may have to make a second visit to receive the result). The TST requires a cold chain, two health care visits and training in intradermal injection, reading and interpretation. One other practical advantage of IGRAs over the TST is that IGRAs are notsusceptible to a "booster response", which makes a two-step approach necessary for the TST in situations where reactivity to the TST has waned since infection.

BCG vaccination plays a decisive role in reducing the specificity of the TST, although the GDG noted that the impact of BCG vaccination on the specificity of the TST depends on the strain of vaccine used, the age at which the vaccine is given and the number of doses administered. When BCG is given at birth, as is the case in most parts of the world, it has a variable, limited impact on TST specificity *(19)*.

The GDG agreed that a history of BCG vaccination has a limited effect on interpretation of TST results later in life; hence, BCG vaccination should not be a determining factor in selecting a test. Neither the TST nor IGRAs are to be used to diagnose active TB disease; also, they are not to be used for diagnostic work-up of adults suspected of having active TB.

Research priorities

There is a critical need for diagnostic tests with improved performance and predictive value for progression to active TB. In addition, the performance of TB infection tests should be evaluated in various risk groups, to assess reinfection and to understand how best to use available tools in each population (e.g. in combination, or sequential use of the TST and IGRAs).

Data synthesis was structured around the preset PICO question, as outlined above. See Web Annex H for additional information on evidence synthesis and analysis.

3.3. TB skin tests and interferon gamma release assays for the diagnosis of TB disease

Recommendation

20. Interferon-gamma release assays (IGRAs) (and the tuberculin skin test [TST]) should not be used in low- and middle-income countries for the diagnosis of pulmonary or extrapulmonary TB, or for the diagnostic work-up of adults (including people living with HIV) suspected of active TB in these settings

(strong recommendation)

The Guideline Development Group concluded that both the sensitivity and specificity of IGRAs in detecting active TB among individuals presumed of having TB were suboptimal and the quality of evidence was low. They also recommended that these tests not be used as a replacement for conventional microbiological diagnosis of pulmonary and extrapulmonary TB.

The Guideline Development Group noted that current evidence did not support the use of IGRAs or the TST as part of the diagnostic work-up of adults presumed of active TB, irrespective of HIV status. This recommendation placed a high value on avoiding the consequences of unnecessary treatment (owing to a high number of false positive results), given the low specificity of IGRAs and the TST in these settings.

Evidence base

A systematic, structured, evidence-based process for TB diagnostic policy generation was followed. The first step constituted systematic reviews and meta-analysis of available data (published and unpublished), using standard methods appropriate for diagnostic accuracy studies. The second step involved the convening of a GDG to evaluate the strength of the evidence base, evaluate the risks and benefits of using IGRAs in LMIC and identify gaps to be addressed in future research. Based on the Expert Group findings, the third and final step involved development of a WHO policy guidance, with eventual dissemination to WHO Member States for implementation.

The GRADE system, adopted by WHO for all policy and guideline development, was used by the GDG. Given the absence of studies evaluating patient-important outcomes among TB suspects randomized to treatment based on IGRA results, reviews were focused on the diagnostic accuracy of IGRAs versus the TST in detecting TB infection or TB disease. Recognizing that test results may be surrogates for patient-important outcomes, the GDG evaluated the accuracy of IGRAs while also drawing inferences on the likely impact of these tests on patient outcomes, as reflected by false negatives (i.e. cases of TB infection missed) or false positives.

Systematic reviews were undertaken following detailed protocols with predefined questions relevant to the individual topics. Summaries of methodologies followed for each topic are given in the relevant sections below.

PICO questions

What is the diagnostic accuracy of commercial IGRAs for pulmonary TB in adult pulmonary TB suspects and confirmed TB cases in LMIC as compared with microbiological (culture or smear-microscopy) or clinical diagnosis of pulmonary TB?

Hierarchy of reference standards

Studies evaluating the performance of IGRAs are hampered by the lack of a gold standard to distinguish the presence or absence of TB infection. Since diagnostic accuracy for TB infection could not be directly assessed, a hierarchy of reference standards was developed and agreed beforehand with the systematic reviewers, to evaluate the role of IGRAs, depending on the individual topic (i.e. not all systematic reviews necessarily used the hierarchy). Primary outcomes were predefined for each systematic review as relevant; for example, the predictive value of IGRAs for development of active TB, the sensitivity of IGRAs in individuals with culture- confirmed active TB (as a surrogate reference standard for TB infection), and the correlation between IGRA and TST results. In addition to primary outcomes, specific characteristics of IGRAs that could influence their overall utility were evaluated where relevant; for example, the proportion of indeterminate IGRA results (i.e. not able to be interpreted, either due to a high IFN-γ response in the negative control or a low IFN-γ response in the positive control), the impact of HIV-related immunosuppression (i.e. CD4+ cell count) on test performance where available and correlation of IGRA results with an exposure gradient (typically used in contact and outbreak investigations).

Studies search, selection and quality assessment

All studies evaluating IGRAs published up to the end of May 2010 were reviewed using predefined data search strings. In addition to database searches, bibliographies of reviews and guidelines were reviewed, citations of all included studies were screened, and experts in the field as well as IGRA manufacturers were contacted to identify additional studies (published, unpublished and ongoing). Pertinent information not reported in the original publications was requested from the primary authors of all studies included by the systematic reviewers.

Studies that evaluated the performance of currently available commercial IGRAs, published in all languages and in all LMIC, were reviewed by individual topic. Only studies evaluating IGRA performance in LMIC were included in this analysis. Excluded were studies that evaluated non-commercial (i.e. in-house) IGRAs, older generation IGRAs (i.e. PPD-based IGRAs) and IGRAs performed in specimens other than blood; studies that were focused on the effect of anti-TB treatment on the IGRA response; studies including fewer than 10 individuals; studies reporting insufficient data to determine diagnostic accuracy measures; and conference abstracts and letters without original data, and reviews.

Study quality was assessed by relevant standardized methods, depending on the topic. For primary outcomes focused on test accuracy, quality was appraised using a subset of relevant criteria from QUADAS, a validated tool for diagnostic accuracy studies. For studies of the predictive value of IGRAs, quality was appraised with a modified version of the Newcastle-Ottawa Scale (NOS) for longitudinal or cohort studies. Conflicts of interest are a known concern in TB diagnostic studies; therefore, the systematic reviews added a quality item about involvement of commercial test manufacturers in published studies; they also reported whether IGRA manufacturers had any involvement with the design or conduct of each study, including donation of test materials, provision of monetary support, work or financial relationships with study authors, and participation in data analysis.

Data synthesis and meta-analysis

A standardized overall approach was specified a priori for each systematic review, to account for significant heterogeneity in results expected between studies. First, data were synthesized separately for each commercial IGRA and by the World Bank country income classification (LMIC versus high-income countries) as a surrogate for TB incidence. Second, heterogeneity was visually assessed using forest plots, and the variation in study results attributable to heterogeneity was characterized (I-squared statistic) and statistically tested (chi-squared test). Third, pooled estimates were calculated using random-effects modelling, which provides more conservative estimates than fixed-effects modelling when heterogeneity is present. For each individual study, all outcomes for which data were available were assessed. First, forest plots were generated to display the individual study estimates and their 95% CIs. Pooled estimates were calculated when at least three studies were available in any subgroup, and individual study results were summarized when fewer than four studies were available. Standard statistical packages were used for analyses.

Use of IGRAs in the diagnosis of active TB

Studies included were those that evaluated the performance of the technologies of interest for the diagnosis of TB disease among adult (>15 years) with presumed TB or people with TB in LMIC.

The initial search yielded 789 citations. After full-text review of 185 papers evaluating IGRAs for the diagnosis of active TB, 22 were determined to meet eligibility criteria, covering 33 unique evaluations of one or more IGRAs (hereafter referred to as studies) in 19 published and three unpublished reports. Of the 33 studies, 10 (30%) were from low-income countries and 23 (70%) were from middle-income countries. Seventeen studies (52%) included PLHIV (n=1057), and 27 studies (82%) involved ambulatory subjects (outpatients as well as hospitalized patients. IGRAs were performed in people suspected of having active TB in 19 studies (58%) and in people with known active TB in 14 studies (42%). Because of the focus on diagnostic accuracy for active TB and the high prevalence of TB infection in high TB burden settings, IGRA specificity was estimated exclusively among studies enrolling TB suspects where the diagnostic work-up ultimately showed no evidence of active disease.

The results demonstrated the following in LMIC:

- The sensitivity of IGRAs in detecting active TB among people suspected of having TB ranged from 73% to 83% and specificity from 49% to 58%. Therefore, one in four patients, on average, with culture-confirmed active TB could be expected to be IGRA-negative in LMIC, with serious consequences for patients in terms of morbidity and mortality.
- There was no evidence that IGRAs have added value beyond conventional microbiological tests for the diagnosis of active TB. Among studies that enrolled TB suspects (i.e. patients with diagnostic uncertainty), both IGRAs demonstrated suboptimal "rule-out" values for TB disease.
- Even though data were limited, the sensitivity of both IGRAs was lower among PLHIV (about 60–70%), suggesting that nearly one in three PLHIV with active TB would be IGRA-negative.
- There was no consistent evidence that either of the two IGRAs was more sensitive than the TST for active TB diagnosis, although comparisons with pooled estimates of TST sensitivity were difficult to interpret owing to substantial heterogeneity.
- The few available head-to-head comparisons between QFT-GIT and T-Spot demonstrated higher sensitivity for the T-Spot platform, although this difference did not reach statistical significance.
- The specificity of both IGRAs for active TB was low, regardless of HIV status, and results suggested that one in two patients without active TB would be IGRA-positive, with adverse consequences for patients because of unnecessary therapy for TB and a missed differential diagnosis.
- Two unpublished reports reported no incremental or added value of IGRA test results combined with important baseline patient characteristics (e.g. demographics, symptoms or chest radiograph findings). Thus, these reports did not support a meaningful contribution of IGRAs for the diagnosis of active TB beyond readily available patient data and conventional tests.

- The systematic review focused on the use of IGRAs to diagnose active pulmonary TB, given that data for extrapulmonary TB were lacking; nevertheless, the GDG consensus was that recommendations for pulmonary TB could reasonably be extrapolated to extrapulmonary TB.
- Industry involvement was unknown in 18% of studies and acknowledged in 27% of studies, including donation of IGRA kits as well as work or financial relationships between authors and IGRA manufacturers.

Strengths and limitations of the evidence base

Strengths and limitations were as follows:

- Heterogeneity was substantial for the primary outcomes of sensitivity and specificity. Activities performed to minimize heterogeneity were empirical random-effects weighting, excluding studies contributing fewer than 10 eligible individuals, and separately synthesizing data for currently manufactured IGRAs.
- No standard criteria exist for defining high TB incidence countries, and the World Bank income classification is an imperfect surrogate for national TB incidence; nevertheless, results were fundamentally unchanged when restricted to countries with an arbitrarily chosen annual TB incidence of at least 50 per 100 000 population.
- It is possible that ongoing studies were missed, despite systematic searching. It is also possible that studies that found poor IGRA performance were less likely to be published. Given the lack of statistical methods to account for publication bias in diagnostic meta-analyses, it would be prudent to assume some degree of overestimation of estimates due to publication bias.
- The systematic review focused on test accuracy (i.e. sensitivity and specificity) and indirect assessment of patient impact (false positive and false negative results). None of the studies reviewed provided information on patient-important outcomes (i.e. showing that IGRAs used in a given situation resulted in a clinically relevant improvement in patient care or outcomes). In addition, no information was available on the values and preferences of patients.

Data synthesis was structured around the preset PICO question, as outlined above. Web Annex I provides additional information on evidence synthesis and analysis.

Operational aspects of the use of IGRAs

Operational aspects of the use of IGRAs were as follows:

- Cost of IGRAs was mentioned by four studies, which all stated that the assays are too expensive and that this is a limitation to their use.
- Only one study addressed reproducibility of T-Spot by assessing inter-observer agreement; it showed excellent correlation. No other study mentioned the issue of test reproducibility.
- Twelve studies reported on accepted transport times of samples to the laboratory, which were mainly less than 6 hours (i.e. within the limit accepted by the test manufacturers). One study accepted a transport time of 16 hours and another 24 hours. None reported on the impact of the transport times (i.e. delay between drawing the blood and initiating the IGRA test) and IGRA test results or performance.
- No study reported on time-to-result for IGRAs.

- Four studies reported on the impact of IGRAs on TB therapy. In two studies, IGRA results were reported to clinicians; one study did not discuss the consequences, and in the other study QFT-positive children and adolescents received preventive chemotherapy. The other two studies commented on the reduced number of patients that would require preventive therapy if IGRAs were part of the diagnostic algorithm.
- The following aspects related to the feasibility of IGRAs were highlighted:
  - blood amounts required may be an issue; however, tests were performed with less than 2 mL of blood (T-Spot) in some studies;
  - a strong interferon response in negative control tubes (high background results) in QFT may reflect the influence of other coincident diseases;
  - standardization and generation of automated, quantitative results should render IGRAs more objective than the TST; and
  - a well-equipped laboratory, expensive equipment and training are required for IGRA test performance, which may cause logistical problems.

Research priorities

Targeted further research to identify IGRAs with improved accuracy is strongly encouraged. Such research should be based on adequate study design, including quality principles such as representative suspect populations, prospective follow-up, and adequate and explicit blinding. It is also strongly recommended that proof-of-principle studies be followed by evidence produced from prospectively implemented and well-designed evaluation and demonstration studies, including assessment of patient impact.

3.4. References

1. Public announcement to TB in vitro diagnostics manufacturers, procurement agencies and national TB programmes on inclusion of WHO prequalification for TB in vitro diagnostics. Geneva: World Health Organization; 2021 (https://extranet.who.int/pqweb/sites/default/files/documents/210211_PublicAnnouncement_TB_%20in-vitro-diagnostics.pdf).
2. Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH 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. 2020;122:142–52 (https://pubmed.ncbi.nlm.nih.gov/32058069/).
3. Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH 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. 2020;122:129–41 (https://pubmed.ncbi.nlm.nih.gov/32060007/).
4. McMaster University. GRADEpro GDT [website]. 2020 (https://gradepro.org/).
5. Starshinova A, Zhuravlev V, Dovgaluk I, Panteleev A, Manina V, Zinchenko U et al. A comparison of intradermal test with recombinant tuberculosis allergen (diaskintest) with other immunologic tests in the diagnosis of tuberculosis infection. International Journal of Mycobacteriology. 2018;7(1):32.
6. Starshinova A, Zinchenko Y, Istomina E, Basantsova N, Filatov M, Belyaeva E et al. [Diagnosis of latent tuberculosis infection in personnel of various institutions and determination of the risk group for tuberculosis]. Bioprep Prev Diagnosis. Treat. 2019;19:178–84.
7. Aggerbeck H, Ruhwald M, Hoff ST, Borregaard B, Hellstrom E, Malahleha M et al. C-Tb skin test to diagnose Mycobacterium tuberculosis infection in children and HIV-infected adults: A phase 3 trial. PLoS One. 2018;13(9):e0204554 (https://www.ncbi.nlm.nih.gov/pubmed/30248152).

8. Aggerbeck H, Ruhwald M, Hoff S, Tingskov P, Hellstrom E, Malahleha M et al. Interaction between C-Tb and PPD given concomitantly in a split-body randomised controlled trial. Int J Tuberc Lung Dis. 2019;23(1):38–44.
9. Ruhwald M, Aggerbeck H, Gallardo RV, Hoff ST, Villate JI, Borregaard B et al. Safety and efficacy of the C-Tb skin test to diagnose Mycobacterium tuberculosis infection, compared with an interferon gamma release assay and the tuberculin skin test: a phase 3, double-blind, randomised, controlled trial. Lancet Respir Med. 2017;5(4):259–68 (https://www.ncbi.nlm.nih.gov/pubmed/28159608).
10. Hoff ST, Peter JG, Theron G, Pascoe M, Tingskov PN, Aggerbeck H et al. Sensitivity of C-Tb: a novel RD-1-specific skin test for the diagnosis of tuberculosis infection. Eur Respir J. 2016;47(3):919–28 (https://www.ncbi.nlm.nih.gov/pubmed/26677940).
11. Streltsova E, Ryzhkova O, Bespalova A. Comparative clinical studies of skin test Diaskintest and Mantoux test in patients with pulmonary tuberculosis. Astrakhan Medical Journal. 2011;6(1):265–70.
12. Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J et al. Predictive value of interferon-y release assays for incident active tuberculosis: a systematic review and meta- analysis. Lancet Infect Dis. 2012;12(1):45–55 (https://pubmed.ncbi.nlm.nih.gov/21846592/).
13. Dorman SE, Belknap R, Graviss EA, Reves R, Schluger N, Weinfurter P et al. Interferon-y release assays and tuberculin skin testing for diagnosis of latent tuberculosis infection in healthcare workers in the United States. Am J Respir Crit Care Med. 2014;189(1):77–87 (https://pubmed.ncbi.nlm.nih.gov/24299555/).
14. Rangaka MX, Wilkinson RJ, Boulle A, Glynn JR, Fielding K, van Cutsem G et al. Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. Lancet. 2014;384(9944):682–90 (https://pubmed.ncbi.nlm.nih.gov/24835842/).
15. Mahomed H, Hawkridge T, Verver S, Abrahams D, Geiter L, Hatherill M et al. The tuberculin skin test versus QuantiFERON TB Gold® in predicting tuberculosis disease in an adolescent cohort study in South Africa. PLoS One. 2011;6(3):e17984 (https://pubmed.ncbi.nlm.nih.gov/21479236/).
16. Mathad JS, Bhosale R, Balasubramanian U, Kanade S, Mave V, Suryavanshi N et al. Quantitative IFN-γ and IL-2 response associated with latent tuberculosis test discordance in HIV-infected pregnant women. Am J Respir Crit Care Med. 2016;193(12):1421–8 (https://pubmed.ncbi.nlm.nih.gov/26765255/).
17. McCarthy KM, Scott LE, Gous N, Tellie M, Venter WD, Stevens WS et al. High incidence of latent tuberculous infection among South African health workers: an urgent call for action. Int J Tuberc Lung Dis. 2015;19(6):647–53 (https://pubmed.ncbi.nlm.nih.gov/25946353/).
18. Sharma SK, Vashishtha R, Chauhan LS, Sreenivas V, Seth D. Comparison of TST and IGRA in diagnosis of latent tuberculosis infection in a high TB-burden setting. PLoS One. 2017;12(1):e0169539 (https://pubmed.ncbi.nlm.nih.gov/28060926/).
19. Zwerling A, Behr MA, Verma A, Brewer TF, Menzies D, Pai M. The BCG world atlas: a database of global BCG vaccination policies and practices. PLoS Med. 2011;8(3):e1001012 (https://pubmed.ncbi.nlm.nih.gov/21445325/).

Annex 1. Guideline development methods

Methods used to develop World Health Organization guidelines

To develop new or update existing guidelines for methods and tools to diagnose tuberculosis (TB), the World Health Organization (WHO) Global TB Programme commissions systematic reviews on the performance or use of the tool or method in question. A systematic review provides a summary of the current literature on diagnostic accuracy or user aspects, for the diagnosis of TB or the detection of anti-TB drug resistance in adults or children (or both) with signs and symptoms of TB.

The certainty of the evidence is assessed consistently for documented evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. GRADE produces an overall quality assessment (or certainty) of evidence and a framework for translating evidence into recommendations. The certainty of the evidence is rated as high, moderate, low or very low. These four categories imply a gradient of confidence in the estimates. Even if a diagnostic accuracy study is of observational design, it would initially be considered high-quality evidence in the GRADE approach *(1)*.

In addition, the WHO Global TB Programme commissions systematic reviews to collect evidence in the field of resource use (i.e. cost and cost–effectiveness), as well as end-user perspectives on particular diagnostic tests or interventions. This evidence-to-recommendation process will inform domains such as feasibility, accessibility, equity and end-user values.

If systematic review evidence is unavailable or is scarce, the potential subsequent effects can be modelled for both diagnostic accuracy as well as cost and cost–effectiveness. For instance, the prevalence of the disease in question, combined with the sensitivity and specificity of a certain test, can be used to estimate the number of false positives and false negatives in a population. Similarly, data on expenditures and cost–effectiveness ratios can be estimated and modelled, based on economical and epidemiological data. Finally, qualitative evidence on the end-user perspective of using a particular test may be generated through end-user interviews if data are scarce in the public domain.

Following a systematic review, the WHO Global TB Programme convenes a Guideline Development Group (GDG) meeting to review the collected evidence. The GDG is made up of external experts whose central task is to develop evidence-based recommendations. The GDG also performs the important task of finalizing the scope and key questions of the guideline in PICO (i.e. population, intervention, comparator and outcomes) format.

This group should be established early in the guideline development process, once the Steering Group has defined the guideline's general scope and target audience, and has begun drafting the key questions. The GDG should be composed of relevant technical experts; end-users, such as programme managers and health professionals, who will adopt, adapt and implement the guideline; representatives of groups most affected by the guideline's recommendations, such as service users and representatives of disadvantaged groups; experts in assessing evidence and developing guidelines informed by evidence; and other technical experts as required (e.g. a health economist or an expert on equity, human rights and gender).

Recommendations are developed based on consensus among GDG members, where possible. When it is not possible to reach consensus, a vote is taken. When a draft guideline is developed by a WHO steering committee, it is reviewed initially by GDG members and subsequently by an External Review Group (ERG). The ERG is made up of individuals interested in the subject, and may include the same categories of specialists as the GDG. When the ERG reviews the final guideline, its role is to identify any errors or missing data, and to comment on clarity, setting, specific issues and implications for implementation – not to change the recommendations formulated by the GDG *(2)*.

Formulation of the recommendations

Evidence is synthesized and presented in GRADE evidence tables. The evidence to decision (EtD) framework is used subsequently to facilitate consideration of the evidence and development of recommendations in a structured and transparent manner. Finally, recommendations are developed based on consensus among GDG members where possible. If it is not possible to reach consensus, then voting takes place. Decisions on the direction and strength of the recommendations are also made using the EtD framework.

Factors that influenced the direction and strength of a recommendation in this guideline were:

- priority of a problem;
- test accuracy;
- balance between desirable and undesirable effects;
- certainty of:
  - evidence of test accuracy;
  - evidence on direct benefits and harms from the test;
  - management guided by the test results;
  - link between test results and management;
- confidence in values and preferences and their variability;
- resource requirements;
- cost–effectiveness;
- equity;
- acceptability; and
- feasibility.

These factors are discussed below.

Priority of a problem

The GDG considers whether the overall consequences of a problem (e.g. increased morbidity, mortality and economic effects) are serious and urgent. The global situation is considered and available data reviewed. In most cases, the problem must be serious and urgent to be considered by a GDG.

Test accuracy

The pooled sensitivity and specificity presented in the GRADE evidence profile is assessed. Preferably and if available the review includes studies with both microbiological reference standards (culture) as well as composite reference standards (e.g. in children and in patients with extrapulmonary TB).

Balance between desirable and undesirable effects

Under this component, GDG members are asked to judge the anticipated benefits and harms from the test in question, including direct effects of the test (e.g. benefits such as faster diagnosis, and harms such as adverse effects from administration of the test). In addition, the possible subsequent effects of the test must be included; for instance, effects of treatment after a positive diagnosis (cure or decrease in mortality), and the effect of no treatment or further testing after a negative test result. Evidence, ideally retrieved from systematic reviews of randomized controlled trials (RCTs) of the test, should inform the GDG of these downstream effects. If evidence from RCTs is not available, diagnostic accuracy studies can be used. In the latter, true positive and true negative diagnosed cases are taken as benefits, whereas false positive and false negative cases are taken as harms.

Certainty of the evidence

Certainty of the evidence of test accuracy is judged scored on a scale from very low, via low and moderate, to high. Certainty of the evidence on direct benefits and harms from the test are assessed and scored in a similar way.

Certainty of management

For certainty of patient management being guided by the test results, the GDG focuses on whether the management would be any different, should it be guided by the test results.

For certainty of the link between test results and management, the panel assesses how quickly and effectively test results can transfer to management decisions.

Confidence in values and preferences and their variability

The value of the test to improve diagnosis and its impact on patient care is evaluated and scored with the help of evidence from qualitative research. The impact on notification and, moreover, the ability of the test to increase case notification is also evaluated and scored, taking into account the entire diagnostic cascade, including, for example, issues related to feasibility of implementation, rate of use, staff's confidence in test results and turnaround time of results.

Resource requirements

In relation to resource requirements, the following questions are answered:

- How large are the resource requirements for test implementation?
- What is the certainty of the evidence about resource requirements?
- Does the cost–effectiveness of the intervention favour the intervention or the comparison?

Cost–effectiveness

Available evidence on cost–effectiveness is evaluated and scored.

Equity

GDG members consider whether implementing the tool or method will positively or negatively affect access to health care (e.g. will it be possible to implement the test in distinct levels of health care or through self-administration, or are there other ways to make the tools or method available to all levels of the health care system).

Acceptability

In terms of acceptability, the panel considers whether the tool or method will be acceptable by all relevant stakeholders, such as health workers, health managers and patients.

Feasibility

The GDG considers how feasible it is to implement a tool or method in various settings. Aspects such as training and refresher training needs, hands-on time, biosafety requirements, time to results, service and maintenance, calibration, and effect on diagnostic algorithms are all taken into account in the final score.

For more details on the transition from evidence to recommendations, see **Web Annex 3**: *Evidence to decision tables*.

Reference for Annex 1

1. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008;336:1106–10. doi: https://doi.org/10.1136/bmj.39500.677199.AE.
2. Handbook for guideline development 2nd ed. Geneva: World Health Organization; 2014 (https://apps.who.int/iris/handle/10665/145714).

Annex 2. Conflict of interest assessment for Guideline Development Group and External Review Group members

Before being considered for group membership, each Guideline Development Group (GDG) and External Review Group candidate was required to submit a completed declaration of interest (DOI) form. In addition, a preliminary internet search was performed to identify any obvious public controversies or interests that may lead to compromising situations for the World Health Organization (WHO) and the expert concerned.

The candidate's curriculum vitae (CV) and DOI, and information retrieved from the internet, were examined by steering committee members to assess whether there were, or may be, actual or perceived conflicts of interest and, if so, whether a management plan was required. This evaluation process, and resultant management plans, were based on the *Guidelines for declaration of interests (WHO experts) (1)* and the *WHO handbook for guideline development (2nd edition) (2)*.

Both financial and non-financial interests were considered. A "significant" conflict of interest would include:

- "intellectual bias", where an individual may have repeatedly and publicly taken a position on an issue under review, which may affect the individual's objectivity and independence in the global policy development process;
- involvement in research or publication of materials related to issues under review; and
- a financial interest above US$ 5000.

Developers of any assay are never involved in the process of policy development; this is automatically considered a conflict of interest.

Once a determination was made that either no conflict of interest existed, or any conflict of interest could be appropriately managed, and a decision had been made to appoint the candidate, the name and a brief biography of each candidate were published on the WHO website for at least 14 days before the meeting, for public notice and comment.

DOI statements are summarized by the WHO steering committee at the start of the meeting. Selected individuals with intellectual or research involvement were invited as technical resource persons to provide technical input and answer technical questions. These individuals did not participate in the Grading of Recommendations Assessment, Development and Evaluation (GRADE) evaluation process and were excluded from the group discussions when recommendations were developed.

Table A.2.1. Summary of the declarations of interest statements for the GDG members: "Molecular assays intended as initial tests", 7–18 December 2020

GDG member	Interests declared	Conclusion
Holger Schünemann	None declared	No conflict of interest
Jeremiah Chakaya Muhwa	None declared	No conflict of interest
Denise Arakaki-Sanchez	None declared	No conflict of interest
David Branigan	None declared	No conflict of interest
Daniela Cirillo	Evaluation of an XDR test prototype. Project funded by Cepheid and FIND. Budget (Research Unit): US$ 14 296; 2018. Member of the Scientific Advisory Board (BIOMERIEUX). Budget (self): US$ 1000; 2020–2021. Evaluation of the blood stability for VIDAS (BIOMERIEUX). Budget (Research Unit): US$ 11 200; 2019.	Significant conflict of interest – excluded from the deliberations on low complexity automated NAATs
Celina Anna Maria Garfin	None declared	No conflict of interest
Petra de Haas	None declared	No conflict of interest
Patricia Hall	Receives funding from PEPFAR; uses PEPFAR funding to procure TB diagnostics tests and supplies for TB, HIV infant diagnostic and HIV viral load testing across multiple countries. Provides technical input into the PEPFAR Country Operational Guidance, including inputs on the appropriate procurement and use of TB and HIV instrumentation and test types. Oversees a global Cepheid GeneXpert-based proficiency testing program and provides technical assistance to PEPFAR-supported CDC country offices and partner ministries of health on the selection and implementation of TB and HIV diagnostic tests and testing resources.	Non-significant conflict of interest
Rumina Hasan	None declared	No conflict of interest

GDG member	Interests declared	Conclusion
Xia Hui	None declared	No conflict of interest
Farzana Ismail	Evaluation of an XDR test prototype (Cepheid). Budget (Research Unit): US$ 140 000; 2020. Bedaquiline post-marketing surveillance and emerging resistance (Janssen). Budget (Research Unit): US$ 300 000; ongoing. Latent TB infection in health care workers (Qiagen). Budget (Research Unit): consumables and personnel; ongoing. Sponsorship to the IUATLD conference participation, Den Haag (self); 2018.	Significant conflict of interest – excluded from the deliberations on low complexity automated NAATs
Katharina Kranzer	EDCTP: total budget about €3.2 million across 6 institutions; FIND: non-TB-related study (study on antimicrobial resistance), US$ 21 000; Cepheid: non-TB-related study (study on STIs), in-kind contributions of 3000 cartridges and the loan of an Xpert machine; Cepheid: TB-related study about 3 gene signature, in-kind contribution of 3500 cartridges and a loan of an Xpert machine; Roche: consumables for validations study in 2015–2016.	Significant conflict of interest – excluded from the deliberations on moderate complexity automated NAATs
Blessina Kumar	None declared	No conflict of interest
Nagalineswaran Kumarasamy	None declared	No conflict of interest
Lindiwe Mvusi	None declared	No conflict of interest
Viet Nhung Nguyen	None declared	No conflict of interest
Mark Nicol	Research grant funding received by the institution for studies of a broad range of novel diagnostics for TB, including NAATs (among others, BD MAX™, one of the tests of interest for the actual GDG). Gates Foundation; NIH; Wellcome Trust; FIND. Employer (University) received funding. Significant (several million US dollars). Ongoing funding from FIND, NIH; co-shares a pending patent for a novel method to extract and purify DNA from sputum samples (not related to or used by any commercial test for TB). Patent belongs to incumbent. No commercial value at present.	Significant conflict of interest – excluded from the deliberations on moderate complexity automated NAATs

GDG member	Interests declared	Conclusion
Leen Rigouts	Research support: the research unit received financial support through FIND for the coordination of and participation in a multicentre evaluation of the Genoscholar PZA-LPA. Part of that funding came from Nipro. Non-monetary renumeration: the research unit received the Genoscholar PZA-LPA kits to conduct evaluations of the test (multicentre study via FIND plus additional ongoing evaluation in the unit).	Significant conflict of interest – excluded from the deliberations on high complexity hybridization-based NAATs
Thomas Shinnick	As an independent consultant, received contracts and travel support from WHO, FIND and USAID for work related to laboratory strengthening and developing global guidance documents; ongoing.	Non-significant conflict of interest
Hojoon Sohn	None declared	No conflict of interest
Sabira Tahseen	None declared	No conflict of interest
Ezio Tavora dos Santos Filho	Has coordinated community advisory boards to the PROVE-IT study (TREAT-TB grant, Union/USAID) in Brazil (REDE-TB) from 2010 to 2015. Currently following up, as an interested party (not as a member of the study team, but as a community advisory board coordinator of other studies), the implementation of the Truenat validation study in Brazil, among other BRICS cooperation studies. Intends to follow up the study, settling establishing system of community advisory boards oversight and protocol analysis in Brazil and other partner countries.	Significant conflict of interest perceived for Molbio Truenat evaluation – excluded from a discussion on Molbio Truenat
Carrie Tudor	Employment (starting January 2015) in the International Council of Nurses, whose TB project received funding from the Eli Lilly Foundation – Lilly MDR-TB Partnership. Funding received was approx US$ 1 million from 2013 to 2019. Current funding period for 2019 is approx. US$ 100 000.	Non-significant conflict of interest
Diana Vakhrusheva	None declared	No conflict of interest

GDG member	Interests declared	Conclusion
Elisabetta Walters	Recipient of grants and a scholarship for doctoral research which included work on GeneXpert MTB/RIF for the diagnosis of intrathoracic TB in children. South African Medical Research Council grant (2012–2015) and scholarship (2015–2018). South African National Research Foundation. FIND. TBTC (CDC); (Research Unit): ZAR 4.2 million; US$ 90 000; 2012–2018.	Non-significant conflict of interest

AIDS: acquired immunodeficiency syndrome; BRICS: Brazil, Russia, India, China and South Africa; CDC: Centers for Disease Control and Prevention; DNA: deoxyribonucleic acid; EDCTP: European and Developing Countries Clinical Trials Partnership; FIND: Foundation for Innovative New Diagnostics; GDG: Guideline Development Group; HIV: human immunodeficiency virus; IUATLD: International Union Against Tuberculosis and Lung Disease; MDR-TB: multidrug-resistant TB; NAAT: nucleic acid amplification test; NIH: National Institutes of Health; PEPFAR: United States President's Emergency Plan for AIDS Relief; STI: sexually transmitted infection; TB: tuberculosis; TBTC: Tuberculosis Trials Consortium; USAID: United States Agency for International Development; WHO: World Health Organization; XDR: extensively drug-resistant.

Table A.2.2. Summary of the declarations of interest statements for the ERG members: "Molecular assays intended as initial tests", 7–18 December 2020

ERG member	Interests declared	Conclusion
Lucilaine Ferrazoli	None declared	No conflict of interest
Alaine Umubyeyi Nyaruhirira	None declared	No conflict of interest
Elisa Tagliani	Involved in the WHO Expert Review Panel for Diagnostics Round 15; specifically, in the evaluation of the BD MAX platform; €5940. A unit at San Raffaele has received funding from an EDCTP grant (TB-CAPT) to FIND. Incumbent has collaborated with FIND in a multicentre clinical study to assess the performance of the Xpert MTB/XDR assay.	Non-significant conflict of interest for External reviewer. Management of potential conflict of interest by interpretation of the comments in the context of conflict of interest
Francis Varaine	None declared	No conflict of interest
Danila Zimenkov	None declared	No conflict of interest

EDCTP: European and Developing Countries Clinical Trials Partnership; ERG: External Review Group; FIND: Foundation for Innovative New Diagnostics; WHO: World Health Organization.

Table A.2.3. Summary of the declarations of interest statements for the GDG: "Targeted next-generation sequencing" 2–5 May 2023

GDG member	Interests declared	Conclusion
Nimalan Arinaminpathy	Employment: Imperial College London. Consulting: Global Fund, WHO, WHO Regional Office for South-East Asia, Clinton Health Access Initiative, Copenhagen Consensus, Stop TB Partnership, USAID. Research support: Gates Foundation, US CDC, USAID, United Kingdom Medical Research Council. Member of various expert advisory bodies, the most relevant to this work being the regional Green Light Committee for the WHO South-East Asia Region, and the TB Strategic and Technical Advisory Group (STAG-TB) for WHO.	Not significant conflict of interest
David Branigan	None declared	No conflict of interest
Daniela Cirillo	Research support including the following: IMI Unite4TB, including coordinating microbiology workplan for clinical trials, monitoring of trials sites and retesting (ongoing); EUCAST, coordinating work for standard protocol for different TB drugs involving reference laboratories (ongoing); TB Alliance, unrestricted grant multipartner to test mic for pretomanid (ended 2020). No studies evaluating targeted NGS, no contracts paid by companies related to targeted NGS.	Not significant conflict of interest
Petra de Haas	Research support including the following: funding for the evaluation of MinION (Oxford Nanopore Technologies), a 5-year KNCV project funded by a national postcode lottery, evaluating implementation and use of targeted NGS for diagnosis of multiple infectious diseases across three countries.	Not significant conflict of interest
Patricia Hall-Eidson	Employed by the US CDC, as a funding organization to support TB research and oversee implementation of WHO TB policies and recommendations globally, including use of NGS for surveillance purposes.	Not significant conflict of interest
Rumina Hasan	None declared	No conflict of interest
Sirinapha Jittimanee	None declared	No conflict of interest
Kobto Koura	None declared	No conflict of interest
Blessina Kumar	None declared	No conflict of interest

GDG member	Interests declared	Conclusion
Nicole Menezes de Souza	None declared	No conflict of interest
Jeremiah Chakaya Muhwa	None declared	No conflict of interest
Norbert Ndjeka	None declared	No conflict of interest
Mark Nicol	Research support: Employer (University of Cape Town, South Africa) has received grant funding to conduct studies of a range of novel diagnostics for TB in children (including Xpert, LAM, RNAseq, and others). No funding for or studies of targeted NGS technologies.	Not significant conflict of interest
Thomas Shinnick	Work as a consultant, through which contracts and travel support have been received from WHO, FIND and USAID for work related to laboratory strengthening and developing global guidance documents.	Not significant conflict of interest
Hojoon Sohn	None declared	No conflict of interest
Sabira Tahseen	None declared	No conflict of interest
Ezio Tavora dos Santos	As an advocate and researcher working in community engagement in research and improvement of policies, the results of this work can potentially benefit the community with which he works.	Not significant conflict of interest
Nguyen Viet Nhung	None declared	No conflict of interest
Elisabetta Walters	Employment: Stellenbosch University, South Africa, no studies assessing or using targeted NGS technologies.	Not significant conflict of interest
Yanlin Zhao	None declared	No conflict of interest

EUCAST: European Committee on Antimicrobial Susceptibility Testing; FIND: Foundation for Innovative New Diagnostics; GDG: Guideline Development Group; Global Fund: Global Fund to Fight AIDS, Tuberculosis and Malaria; IMI: Innovative Medicines Initiative; LAM: lipoarabinomannan antigen test; NGS: next-generation sequencing; TB: tuberculosis; United Kingdom: United Kingdom of Great Britain and Northern Ireland; USAID: United States Agency for International Development; US CDC: United States Centers for Disease Control and Prevention; WHO: World Health Organization.

Table A.2.4. Summary of the declarations of interest statements for the ERG: "Targeted Next-Generation Sequencing" 2–5 May 2023

ERG member	Interests declared	Conclusion
Nagalineswaran Kumarasamy	None	No conflict of interest
Katharina Kranzer	Employment at LSHTM; research support from EDCTP.	Not significant conflict of interest
Farzana Ismail	Research support: Research Unit was included as one of the sites for the FIND Multi-centre Clinical Trial to Assess the Performance of Culture-free, End-to-end Targeted NGS Solutions for Diagnosis of Drug Resistant TB.	Not significant conflict of interest
Mitarai Satoshi	None	No conflict of interest
John Metcalfe	Research support from NIH; patent application entitled "Methods for Producing Circular Deoxyribonucleic Acids".	Not significant conflict of interest

EDCTP: European and Developing Countries Clinical Trials Partnership; ERG: External Review Group; FIND: Foundation for Innovative New Diagnostics; LSHTM: London School of Hygiene & Tropical Medicine; NGS: next-generation sequencing; NIH: United States National Institutes of Health; TB: tuberculosis.

Table A.2.4. Summary of the declarations of interest statements for the GDG: "Low complexity nucleic acid amplification testing for detection of TB and resistance to rifampicin" 6–10 May 2024 NEW

GDG member(s)	Interests declared	Conclusion
David Branigan	None declared	No conflict of interest
Jeremaya Chakaya Muhva	None declared	No conflict of interest
Chamreun Sok Choub	None declared	No conflict of interest
Katherine Fielding	None declared	No conflict of interest
Rumina Hasan	None declared	No conflict of interest
Kobto Gislain Koura	None declared	No conflict of interest
Andrei Maryandyshev	None declared	No conflict of interest
Norbert Ndjeka	None declared	No conflict of interest
Thomas Shinnick	None declared	No conflict of interest
Hojoon Sohn	None declared	No conflict of interest
Sabira Tahseen	None declared	No conflict of interest

GDG member(s)	Interests declared	Conclusion
Timothy Walker	None declared	No conflict of interest
Ou Xichao	None declared	No conflict of interest
Daniela Cirillo	Validation of the standard method for determination of MIC (BD/Jannsen), for research unit, ceased in 2023	Not significant conflict of interest
Keertan Dheda	Research grant to evaluate transcriptomic signature for BioMerieux (Awarded to UCT Lung Institute); Honoraria for being on speakers bureau or providing training Otsuka TB Drug; Honoraria for speakers on bureau (African Society of Laboratory Medicine) Cepheid – US$ 1200.	Not significant conflict of interest
Patricia Hall-Eidson	All travel costs associated with meeting attendance were supported by the US CDC. Employed and salaried by the US President's Emergency Plan for AIDS Relief, a global health program that recommends, procures, and provides virtual and in-country technical assistance on TB low complexity automated nucleic acid amplification tests (and other TB assays) in low, moderate, and high burden countries to aid in TB case finding efforts among pediatric, adolescent, and adult persons living with HIV.	Not significant conflict of interest
Sirinapha Jittimanie	Consultancy for GFATM country coordination mechanism in Thailand	Not significant conflict of interest
Katharina Kranzer	Employment at LSHTM; Pretomanid consultancy for WHO; Research grant from EDCTP: ERASE-TB. In-kind contribution from Cepheid, SD-Biosensor for ERASE-TB.	Not significant conflict of interest
Shaheed Vally Omar	Clinical Evaluation of the Xpert MTB/RIF assay using the GeneXpert OMNI; Clinical Evaluation of the Xpert® MTB/XDR Assay; Head-to-Head Evaluation of the CAT-A Enzyme Xpert MTB/RIF Ultra test vs. the Current on Market Xpert MTB/RIF Ultra V2 test, Cepheid, USD 66,390, for research unit, ceased in 2023.	Not significant conflict of interest

References for Annex 2

1. Declaration of interests [website]. Geneva: World Health Organization; 2023 (https://www.who.int/about/ethics/declarations-of-interest).
2. Handbook for guideline development 2nd ed. Geneva: World Health Organization; 2014 (https://apps.who.int/iris/handle/10665/145714).

Annex 3. GDG members expertise, region, gender

Table A.3.1. Guideline development group members: "Targeted next-generation sequencing" 2–5 May 2023

GDG member	Expertise	WHO region	Gender
Nimalan Arinaminpathy	TB epidemiological moeling	European Region	M
David Branigan	Patient rights; Community care; TB detection and diagnosis	Region of the Americas	M
Daniela Cirillo	TB laboratory diagnosis	European Region	F
Petra de Haas	TB laboratory diagnosis	European Region	F
Patricia Hall-Eidson	TB laboratory diagnosis	Region of the Americas	F
Rumina Hasan	TB laboratory diagnosis	Eastern Mediterranean Region	F
Sirinapha Jittimanee	TB nursing care	Western Pacific Region	F
Kobto Koura	TB treatment	European Region	M
Blessina Kumar	Patient rights; Community care	South-East Asia Region	F
Nicole Menezes de Souza	TB laboratory diagnosis	Region of the Americas	F
Jeremiah Chakaya Muhwa	TB detection and diagnosis/ TB treatment	African Region	M
Norbert Ndjeka	TB treatment	African Region	M
Mark Nicol	TB diagnostics research	Western Pacific Region	M
Thomas Shinnick	TB laboratory diagnosis	Region of the Americas	M
Hojoon Sohn	Health Economics	Western Pacific Region	M

GDG member	Expertise	WHO region	Gender
Sabira Tahseen	TB laboratory diagnosis	Eastern Mediterranean Region	M
Ezio Tavora dos Santos	Patient rights; Community care; TB detection and diagnosis	Region of the Americas	M
Nguyen Viet Nhung	TB program management	Western Pacific Region	M
Elisabetta Walters	Pediatric TB diagnosis and treatment	African Region	F
Yanlin Zhao	TB program management	Western Pacific Region	M

GDG: Guideline Development Group; WHO: World Health Organization.

Table A.3.2. Summary of the declarations of interest statements for the GDG: "Low complexity nucleic acid amplification testing for detection of TB and resistance to rifampicin" 6–10 May 2024 NEW

GDG member	Expertise	WHO Region	Gender
David Branigan	Patient advocacy and rights; Community care; TB detection and diagnosis	Region of the Americas	M
Jeremaya Chakaya Muhva	TB detection and diagnosis/ TB treatment	African Region	M
Chamreun Sok Choub	Patient advocacy and rights	Western Pacific Region	M
Katherine Fielding	TB epidemiology and data science	European Region	F
Rumina Hasan	TB laboratory diagnosis	Eastern Mediterranean Region	F
Kobto Gislain Koura	TB treatment	European Region	M
Andrei Maryandyshev	TB clinical management and treatment	European Region	M
Norbert Ndjeka	TB program management	African Region	M
Thomas Shinnick	TB laboratory diagnosis	Region of the Americas	M
Hojoon Sohn	Health Economics	Western Pacific Region	M

GDG member	Expertise	WHO Region	Gender
Sabira Tahseen	TB laboratory diagnosis	Eastern Mediterranean Region	M
Timothy Walker	TB laboratory diangosis and treatment	European Region	M
Ou Xichao	TB laboratory diagnosis	Western Pacific Region	M
Daniela Cirillo	TB laboratory diagnosis	European Region	F
Keertan Dheda	TB treatment	African Region	M
Patricia Hall-Eidson	TB laboratory diagnosis	Region of the Americas	F
Sirinapha Jittimanie	Patient advocacy and rights; Community care; Health program management (Nursing)	Southeast Asia Region	F
Katharina Kranzer	TB epidemiology and treatment	European Region	F
Shaheed Vally Omar	TB laboratory diagnosis	African Region	M

For further information, please contact:

Global Programme on Tuberculosis & Lung Health
World Health Organization
20, Avenue Appia CH-1211 Geneva 27 Switzerland
Web site: www.who.int/tb