WHO consolidated guidelines on tuberculosis

Module 3: Diagnosis

Tests for tuberculosis infection

World Health Organization

WHO consolidated guidelines on tuberculosis

Module 3: Diagnosis

Tests for tuberculosis infection

World Health Organization

WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for tuberculosis infection

ISBN 978-92-4-005608-4 (electronic version)

ISBN 978-92-4-005609-1 (print version)

Suggested citation. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for tuberculosis infection. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

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

Design by Inís Communication

Contents

Web annexes

Acknowledgements

The recommendations included in these consolidated policy guidelines on tuberculosis (TB) are the result of a collaborative effort among professionals from a range of specialities. The World Health Organization (WHO) is grateful for their time and support.

TB antigen-based skin tests (TBSTs) for the diagnosis of TB infection

GDG meeting on use of TBSTs for the diagnosis of TB infection, Geneva, Switzerland, 31 January–3 February 2022.

Guideline Development Group

Helen Ayles, London School of Hygiene & Tropical Medicine (LSHTM)/Zambart, United Kingdom of Great Britain and Northern Ireland (United Kingdom); **David Branigan**, Treatment Action Group, New York, United States of America (USA); **Jeremiah Chakaya Muhwa**, Co-chair, International Union Against Tuberculosis and Lung Disease, Nairobi, Kenya; **Daniela Cirillo**, San Raffaele Tuberculosis Supranational Reference Laboratory (SRL), Milan, Italy; **Frank Cobelens**, Amsterdam Institute of Global Health and Development, Netherlands; **Anand Date**, Centers for Disease Control and Prevention (CDC), Atlanta, USA; **Petra de Haas**, KNCV Tuberculosis Foundation, The Hague, Netherlands; **Rumina Hasan**, Department of Pathology and Microbiology, Aga Khan University, SRL Karachi, Pakistan; **Farzana Ismail**, Centre for Tuberculosis, National Institute for Communicable Diseases/National Health Laboratory Service and WHO SRL, Johannesburg, South Africa; **Katharina Kranzer**, LSHTM, Harare, Zimbabwe; **Afranio Kritski**, University of Rio de Janeiro, Brazil; **Blessina Kumar**, Global Coalition of TB Activists, India; **Nagalineswaran Kumarasamy**, Voluntary Health Services Infectious Diseases Medical Centre (VHS IDMC), Chennai, India; **Andrei Maryandyshev**, Nordic University, Arkhangelsk, Russian Federation; **Alberto Matteelli**, University of Brescia, Italy; **Satoshi Mitarai**, Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association (RIT/JATA), Japan; **Lindiwe Mvusi**, national TB programme (NTP), Pretoria, South Africa; **Mark Nicol**, University of Western Australia, Perth, Australia; **Leen Rigouts**, Prince Leopold Institute of Tropical Medicine, Brussels, Belgium; **Thomas Shinnick**, Independent Consultant, Atlanta, USA; **Hojoon Sohn**, Seoul National University College of Medicine, Seoul, Republic of Korea; **Sabira Tahseen**, NTP, Ministry of National Health Services Regulations and Coordination, Government of Pakistan, Islamabad, Pakistan; **Ezio Távora dos Santos Filho**, Civil Society Task Force Member, Rio de Janeiro, Brazil; **Carrie Tudor**, TB Project Director, International Council of Nurses, Durban, South Africa; **Marieke van der Werf**, European Centre for Disease Prevention and Control, Sweden; and **Yanlin Zhao**, National Center for Tuberculosis Control and Prevention, China CDC, Beijing, China.

Methodologist

Lawrence Mbuagbaw, McMaster University, Canada.

External Review Group

Francis Drobniewski, Imperial College London, United Kingdom; **Sergei Skornyakov**, Ural Research Institute for Phthisiopulmonology, National Medical Research Center for Phthisiopulmonology and Infectious Diseases of the Ministry of Health of Russian Federation, Yekaterinburg, Russian Federation; **Francis Varaine**, Médecins Sans Frontières, Paris, France; and **Elisabetta Walters**, Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Stellenbosch University, Cape Town, South Africa.

Systematic review team

Molebogeng (Lele) Rangaka, University College London, United Kingdom; **Yohhei Hamada**, University College London, United Kingdom; **Lara Goscé**, University College London, United Kingdom; **Amy Chan**, University of Auckland, New Zealand; **Victoria Liu**, LSHTM, United Kingdom; **Ting Ting Wang**, University College London, United Kingdom; and **Liliya Ziganshina**, Cochrane Russia, Russian Federation.

Observers

Draurio Cravo Neto Barreira, Unitaid, Switzerland; **Fatim Cham-Jallow**, Global Fund to Fight AIDS, Tuberculosis and Malaria, Switzerland; **Brian Kaiser**, Global Drug Facility, Stop TB Partnership, Switzerland; **Amy Piatek**, United States Agency for International Development (USAID), USA; **Morten Ruhwald**, Foundation for Innovative New Diagnostics (FIND), Switzerland; and **Wayne Van Gemert**, TB Diagnostics Market Strategies, Stop TB Partnership, Switzerland.

WHO steering committee

Work on these guidelines was overseen by **Nazir Ismail** and **Alexei Korobitsyn**, with input from **Dennis Falzon**, **Avinash Kanchar**, **Linh Nguyen**, **Lice Gonzalez-Angulo**, **Annabel Baddeley**, **Carl-Michael Nathanson** (all WHO Global TB Programme [WHO/GTB]), **Mercedes Gonzalez** (Science division, Research for Health department (SCI/RFH) and **Soudeh Ehsani** (WHO Regional Office for Europe) under the overall coordination of **Matteo Zignol** (WHO/GTB) and the direction of **Tereza Kasaeva**, Director of WHO/GTB.

Funding

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

Tuberculin skin test (TST) and interferon-gamma release assays (IGRAs) for the diagnosis of TB infection

GDG meeting on use of TST and IGRAs in tuberculosis control in low- and middle-income settings Geneva, Switzerland, 25 April–10 May 2017.

The production of the *WHO consolidated guidelines on tuberculosis Module 1: Prevention – tuberculosis preventive treatment* was coordinated and written by **Dennis Falzon**, **Avinash Kanchar** and **Matteo Zignol** under the overall direction of **Tereza Kasaeva**, Director of WHO/GTB. WHO/GTB gratefully acknowledges the contribution of all experts involved in the production of these guidelines.

Guideline Development Group

Nadia Ismail Abu Sabrah, consultant TB physician and epidemiologist, Jordan; **Mohammed Al Lawati**, consultant physician, Oman; **Si Thu Aung**, NTP, Department of Public Health, Ministry of Health, Myanmar; **Rolando Cedillos**, Service of Infectology and Integrated Programme for STI/HIV/AIDS, El Salvador; **Richard Chaisson**, Center for TB Research, Johns Hopkins University, USA; **Padmapriyadarsini Chandrasekaran**, National Institute for Research in Tuberculosis, India; **Lucy Chesire**, TB Advocacy Consortium, Kenya; **Betina Durovni**, Federal University of Rio de Janeiro, Brazil; **Diana Gibb**, Medical Research Council, United Kingdom; **Stephen Graham**, Centre for International Child Health, University of Melbourne, Australia; **Sally Hargreaves**, Imperial College London, United Kingdom; **Diane Havlir**, University of California, USA; **Nguyen van Hung**, National Lung Hospital, Viet Nam; **Alberto Matteelli**, Co-chair, University of Brescia, WHO Collaborative Centre for TB/HIV and TB Elimination, Italy; **Lindiwe Mvusi**, National Department of Health, South Africa; **Marieke van der Werf**, European Centre for Disease Prevention and Control, Sweden; **Wim Vandevelde**, Global TB Community Advisory Board, South Africa; and **Irina Vasilyeva**, Ministry of Health, Russian Federation.

Methodologist

Nandi Siegfried, Medical Research Council / University of Cape Town, South Africa.

External Review Group

Connie Erkens, KNCV Tuberculosis Foundation, Netherlands; **Steve Graham**, Center for International Child Health University of Melbourne, Australia; **Giovanni B. Migliori**, Istituti Clinici Scientifici Maugeri IRCCS and WHO Collaborating Centre for Tuberculosis and Lung Diseases, Italy; **Rohit Sarin**, National Institute of TB and Respiratory Diseases, India; **James Seddon**, Imperial College, United Kingdom; **Alena Skrahina**, Republican Scientific and Practical Centre for Pulmonology and TB, Belarus; and **Carrie Tudor**, International Council of Nurses, South Africa.

WHO steering committee

Work on these guidelines was overseen by **Annabel Baddeley**, **Annemieke Brands**, **Dennis Falzon**, **Carmen Figueroa**, **Medea Gegia**, **Christopher Gilpin**, **Philippe Glaziou**, **Avinash Kanchar** and **Matteo Zignol** from WHO/GTB; **Françoise Renaud** and **Satvinder Singh** from the WHO Department of HIV/AIDS; **Andreas Reis** from the Department of Information, Evidence and Research; and **Lorenzo Moja** from the Department of Essential Medicines and Health Products. **Wilson Were** from the WHO Department of Maternal, Newborn, Child and Adolescent Health also reviewed the guidelines.

Funding

Funding for the guidelines update in 2019–2020 derived from WHO grants provided by USAID and the Russian Federation.

Use of interferon-gamma release assays (IGRAs) for the diagnosis of TB disease

Expert Group meeting on use of interferon-gamma release assays (IGRAs) in tuberculosis control in low- and middle-income settings Geneva, Switzerland, 19–20 July 2010.

Guideline Development Group

Richard A. Adegbola, Global Health, Bill & Melinda Gates Foundation, USA; **Catharina Boehme**, FIND, Switzerland; **Daniela Cirillo**, San Raffaele SRL, Italy; **Peter Godfrey-Faussett**, LSHTM, United Kingdom; **Anneke C. Hesseling**, Stellenbosch University, South Africa; **Phillip Hill**, University of Otago, New Zealand; **Oluwamayowa Joel**, Communication for Development Centre, Nigeria; **Suman Laal**, New York University, USA; **Philip LoBue**, CDC, USA; **Rick O'Brien**, FIND, Switzerland; **Holger Schünemann**, McMaster University, Canada; and **Lakhbir Singh Chauhan**, Ministry of Health and Family Welfare, India.

Methodologist

Holger Schünemann, McMaster University, Canada.

Systematic review team/Observers

Adithya Cattamanchi, San Francisco General Hospital Pulmonary Division, USA; **Anand Date**, CDC, USA; **Anne Detjen**, North America Office, International Union Against Tuberculosis and Lung Disease, USA; **David Dowdy**, University of California San Francisco, USA; **Richard (Dick) Menzies**, Montreal Chest Institute, Canada; **John Metcalfe**, University of California San Francisco, USA; **Madhukar Pai**, McGill University, Canada; **Molebogeng (Lele) Rangaka**, University College London, United Kingdom; **Karen R. Steingart**, Curry International Tuberculosis Center, University of California, San Francisco, USA; and **Alice Zwerling**, Ottawa University, Canada.

WHO steering committee

Work on these guidelines was overseen by **Chris Gilpin** and **Jean Iragena** with input from **Regina Kulier**, **Christian Lienhardt**, **Fuad Mirzayev**, **Mario Raviglione**, **Karin Weyer** and **Matteo Zignol** (all WHO Stop TB Department), **Luis Cuevas**, **Jane Cunningham**, **Andy Ramsay** and **Soumya Swaminathan** (all WHO Special Programme for Research and Training in Tropical Diseases [TDR]).

Funding

Funding support from the United Nations Children's Fund, the United Nations Development Programme, the World Bank, WHO/TDR and TREAT TB/The Union. The views of the funding agencies have not influenced the development and content of these guidelines.

Abbreviations and acronyms

AIDS	acquired immunodeficiency syndrome
BCG	bacille Calmette-Guérin
CFP-10	culture filtrate protein 10
CI	confidence interval
DCE	discrete choice experiment
DOI	declaration of interest
DST	drug susceptibility testing
ELISA	enzyme-linked immunosorbent assay
ELISPOT	enzyme-linked immunospot
ERG	External Review Group
ESAT-6	early secretory antigenic target 6 kDa protein
GDG	Guideline Development Group
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HIV	human immunodeficiency virus
IFN-γ	interferon-gamma
IGRA	interferon-gamma release assay
ISR	injection site reaction
LMIC	low- and middle-income countries
Mtb	*Mycobacterium tuberculosis*
NTM	nontuberculous mycobacteria
NTP	national tuberculosis programme
PICO	population, intervention, comparator and outcome
PLHIV	people living with HIV
PPD	purified protein derivative
PPD-S	purified protein derivative–standard
QALY	quality-adjusted life year
QFT-G	QIAGEN QuantiFERON®-TB Gold
QFT-GIT	QIAGEN QuantiFERON-TB Gold In-Tube
RR	risk ratio
TB	tuberculosis
TBST	*Mycobacterium tuberculosis* antigen-based skin test
TPT	TB preventive treatment
T-Spot	Oxford Immunotec T-SPOT®.*TB*
TST	tuberculin skin test

United Kingdom	United Kingdom of Great Britain and Northern Ireland
USA	United States of America
WHO	World Health Organization
WHO/GTB	World Health Organization Global TB Programme

Executive summary

Tuberculosis (TB) infection is a state that is characterized by persistent immune response to stimulation by *Mycobacterium tuberculosis* (*Mtb*) antigens with no evidence of clinically manifest TB disease.[1] It is estimated that about a quarter of the world's population is infected with *Mtb*. Testing for TB infection increases the probability that individuals who are the target for TB preventive treatment (TPT) will benefit from such treatment. However, there is no gold-standard test to diagnose TB infection. The two currently available classes of tests – interferon-gamma release assays (IGRAs) and the tuberculin skin test (TST) – 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.

In 2011, WHO issued recommendations on the use of IGRAs for the diagnosis of TB infection. The following technologies were included in the evaluation:

- TST;
- QIAGEN QuantiFERON®-TB Gold (QFT-G);
- QIAGEN QuantiFERON-TB Gold In-Tube (QFT-GIT); and
- Oxford Immunotec T-SPOT®.*TB* (T-Spot) assays.

In 2018, WHO updated the recommendations stipulating that the TST or IGRAs (or both) can be used for TB infection. This recommendation was included in the *WHO consolidated guidelines on tuberculosis Module 1: Prevention – tuberculosis preventive treatment*.[1]

In 2021, the WHO recommendations were extended for the following technologies [2]:

- Beijing Wantai's TB-IGRA;
- QIAGEN QuantiFERON-TB Gold Plus.

Newer *Mtb* antigen-based skin tests (TBSTs) based on specific *Mtb* antigens have been developed, using the early secretory antigenic target 6 kDa protein (ESAT-6) and culture filtrate protein 10 (CFP-10) antigens; these tests combine the simpler skin-test platform with the specificity of IGRAs. Emerging evidence suggests that, compared with IGRAs, TBSTs may have similar specificity and provide more reliable results in children, adolescents and in people living with HIV (PLHIV). However, the evidence has not been systematically reviewed.

In 2022, WHO issued recommendations on the use of TBSTs for the diagnosis of TB infection. The following technologies were included in the evaluation:

---

1 WHO consolidated guidelines on tuberculosis Module 1: Prevention – tuberculosis preventive treatment. Geneva: World Health Organization; 2020 (https://www.who.int/publications/i/item/9789240001503).

2 Use of alternative interferon-gamma release assays for the diagnosis of TB infection: WHO policy statement. Geneva: World Health Organization; 2022 (https://www.who.int/publications/i/item/9789240042346).

- Cy-Tb (Serum Institute of India, India);
- Diaskintest® (Generium, Russian Federation); and
- C-TST (formerly known as ESAT6-CFP10 test, Anhui Zhifei Longcom, China).

Objectives of the current guidelines

The objectives of the current guidelines are to:

- assess the available data on the diagnostic accuracy (sensitivity and specificity) of the TST, IGRAs and TBSTs for the diagnosis of TB infection;
- assess the available data on the concordance of the TST, IGRAs and TBSTs;
- assess the available data related to the impact of the TST, IGRAs and TBSTs on patient-important outcomes: efficacy of TPT based on diagnostic test results, predictive value for progression to TB disease, correlation with exposure gradient and proportion started on TPT (if available);
- present a review of the published qualitative data on feasibility, accessibility, equity and end-user values related to TST, IGRAs and TBST implementation;
- review the published economic data on affordability, cost and cost–effectiveness of TST, IGRAs and TBST implementation; and
- determine questions for future research and issues to be addressed by WHO in subsequent policy recommendations.

Target audience

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

Methods

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 on respective tests for TB infection (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 and identify gaps to be addressed in future research. 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 people with presumed TB randomized to treatment based on TB infection results, reviews were focused on the diagnostic accuracy of respective TB infection tests (TBSTs, IGRAs and TST) in detecting TB infection or TB disease. Recognizing that test results may be surrogates for patient-important outcomes, the GDGs evaluated the accuracy of TB infection tests 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.

In 2018, 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. 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. The data on cost and cost-effectiveness as well as qualitative evidence were assessed where possible.

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.

Recommendations

- *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
- Either the 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
- Interferon-gamma release assays (IGRAs) (and the tuberculin skin test [TST]) should not be used in low- and middle-income countries (LMIC) 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

Justification

Based on available evidence, in 2022 the WHO GDG panel concluded that the diagnostic accuracy of TBSTs is similar to that of IGRAs and greater than that of the TST. The GDG panel expressed concerns about the certainty (quality) of evidence in many areas and the lack of longitudinal studies that include impact on people affected by important outcomes of TB. The risk of bias was primarily from non-blinded studies, and the quantity and quality of evidence varied among the different tests. For two of the three tests (Diaskintest and C-TST) evaluated during the GDG meeting, evidence on specificity was generated in high TB burden settings; therefore, additional analysis considered the concordance in specificity with existing WHO-recommended IGRAs. All three evaluated TBSTs have the potential to be used for the detection of TB infection and are recommended.

In 2018 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. 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.

In 2011, the GDG 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. Furthermore, the GDG 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.

The current recommendations are based on evaluation of data for all classes of TB infection diagnostic tests that were included in the respective evaluations. The findings cannot be extrapolated to other brand-specific tests. Also, any new in-class technologies will need to be specifically evaluated by WHO, in line with updated WHO procedures, to determine procurement eligibility for in vitro diagnostics for TB. [3]

Dissemination and evaluation

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. Global TB report is collecting and reporting data on TB infection tests for all country-members used in respective year, which include information on all WHO recommended TB infection tests use, being direct indicators of current guidelines uptake. In addition, data on TB patient contacts investigation and TB preventive treatment, included in the same report, are indirect indicators of the current guidelines uptake.

[3] 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).

1. Use of *Mycobacterium tuberculosis* antigen-based skin tests for the diagnosis of TB infection NEW

1.1. Background

Tuberculosis (TB) infection is a state that is characterized by persistent immune response to stimulation by *Mycobacterium tuberculosis* (*Mtb*) antigens with no evidence of clinically manifest TB disease *(1)*. It is estimated that about a quarter of the world's population is infected with *Mtb*. Testing for TB infection can identify individuals who would benefit the most from TB preventive treatment (TPT). However, there is no gold-standard test to diagnose TB infection. The two classes of tests that are currently available – the tuberculin skin test (TST) and interferon-gamma release assays (IGRAs) – are indirect tests for immune sensitization, and they require an 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.

In 2011, the World Health Organization (WHO) issued recommendations on the use of IGRAs for the diagnosis of TB infection, including the blood-based QIAGEN QuantiFERON®-TB Gold (QFT-G), QIAGEN QuantiFERON-TB Gold In-Tube (QFT-GIT) and Oxford Immunotec T-SPOT®. *TB* (T-Spot) assays. 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.

Newer *Mtb* antigen-based skin tests (TBSTs) based on specific antigens have 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 has not been systematically reviewed.

The WHO assessment process for TB diagnostics has evolved into a mechanism that focuses on the evaluation of classes of TB diagnostic technologies rather than of specific products. Some of the important elements to guide implementation are diagnostic accuracy, the epidemiological and geographical setting, operational aspects (turnaround times, throughput, existing infrastructure and specimen referral networks), economic aspects and qualitative aspects on acceptability, equity, end-user values and preferences.

The TBST class is defined as in vivo skin tests for the detection of TB infection that use *Mtb*-specific antigens (ESAT-6 and CFP-10).

In 2021, WHO commissioned a systematic review of published and unpublished data on this new class of tests for TB infection not previously reviewed 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).

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.

1.2. Recommendation

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

1.2.1. Justification

Based on available evidence, the WHO GDG panel concluded that the diagnostic accuracy of TBSTs is similar to that of IGRAs and greater than that of the TST. The GDG panel expressed concerns about the certainty (quality) of evidence in many areas and the lack of longitudinal studies that include impact on people affected by important outcomes of TB. The risk of bias was primarily from non-blinded studies, and the quantity and quality of evidence varied among the different tests. For two of the three tests (Diaskintest and C-TST) evaluated during the GDG meeting, evidence on specificity was generated in high TB burden settings; therefore, additional analysis considered the concordance in specificity with existing WHO-recommended IGRAs.

No safety concerns were identified for the class of tests; however, evaluation and approval by the competent regulatory agencies for the individual products are essential before introduction of these in vivo tests. Data on cost and cost–effectiveness were limited. Cost–effectiveness modelling calibrated to three countries (Brazil, South Africa and the United Kingdom of Great Britain and Northern Ireland [United Kingdom]) was commissioned to inform the GDG meeting. It was found that in Brazil and South Africa, use of TBSTs would be cost saving compared with both the TST and IGRAs; in the United Kingdom, it would be still cost saving compared with the TST but only cost-effective compared with IGRAs.

Qualitative evidence indicates that 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. It was suggested that TBSTs were perceived to have greater accuracy than the TST and were considered desirable because they avoid the negative consequences of false positive results. Finally, qualitative evidence supports the feasibility of using TBSTs in settings where the TST is already in use, because the required resourcing and training are already in place.

All three evaluated TBSTs have the potential to be used for the detection of TB infection and are recommended.

1.2.2. Subpopulations

Although the data were limited, based on the available evidence, the GDG members supported extrapolation of the recommendation for the following populations:

- PLHIV;
- children and adolescents aged under 18 years; and
- people who have been vaccinated with BCG.

For all the above-mentioned subpopulations, TBSTs are recommended.

Testing for TB infection is not a precondition for initiating TPT among and children under 5 years of age who are household contacts of people with active TB. Nevertheless, testing for TB infection increases the certainty that individuals will benefit from TPT.

1.3. Test descriptions

The following tests 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).

All these tests use intradermal injection of ESAT-6 and CFP-10 antigens that are specific to *Mtb* and stimulate T-cell release of IFN-γ. The effects of the products are based on a cellular immune response to *Mtb*-specific antigens. When administered intracutaneously in people with TB infection, tests included in this class induce a specific skin reaction, which is a manifestation of delayed-type hypersensitivity. The immune response is measured after 48–72 hours as induration in millimetres.

Cy-Tb was manufactured by the Statens Serum Institut (SSI), Denmark, as a solution of two recombinant proteins of ESAT-6 and CFP-10 (in a 1:1 ratio), produced by genetically modified *Lactobacillus lactis*. One single test dose of 0.1 mL contains 0.05 µg of rd ESAT-6 and 0.05 µg of rCFP-10. In 2019, SSI entered into a partnership with the Serum Institute of India, which is licensed to produce and distribute the test.

Diaskintest is a recombinant protein that is produced by genetically modified *Escherichia coli* BL21 (DE3)/pCFP-ESAT, diluted with sterile isotonic phosphate buffer solution, with a preservative (phenol). One dose (0.1 mL) of the product contains 0.2 µg of ESAT-6 and CFP-10 recombinant protein and auxiliary ingredients: disodium phosphate dihydrate, sodium chloride, potassium dihydrogen phosphate, polysorbate 80, phenol and water for injections. The product is approved by the Russian National Medicines Regulatory Authority.

C-TST uses a recombinant fusion protein of ESAT-6 and CFP-10, and is manufactured by Anhui Zhifei Longcom Biopharmaceutical Co. Ltd. The product is approved by the Chinese National Medicines Regulatory Authority. The active ingredient in this test is an ESAT-6 and CFP-10 fusion recombinant protein expressed in genetically modified *E. coli.* Each test dose of 0.1 mL contains 5 units (U) of recombinant *Mtb* fusion protein and auxiliary ingredients: disodium hydrogen phosphate, potassium dihydrogen phosphate, sodium chloride, phenol and polysorbate 80.

1.4. 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 testtests (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 overarching 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?**

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

BCG: bacille Calmette-Guérin; CXR: chest X-ray; HIV: human immunodeficiency virus; IGRA: interferon-gamma release assay; *Mtb*: *Mycobacterium tuberculosis*; PICO: population, intervention, comparator and outcome; PLHIV: people living with HIV; TB: tuberculosis; TBST: *M*tb antigen-based skin test; TNF: tumour necrosis factor; TPT: TB preventive treatment; TST: tuberculin skin test.

a 100/100 000 population.

b For estimation of specificity, the ideal population is one with very low likelihood of prior exposure to Mtb.

c TB disease is used as a proxy diagnosis for 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:

[4] End-users are health care providers, laboratory technicians and managers, programme staff, community workers, people being offered the test and family.

- ***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*

1.4.1. Diagnostic accuracy

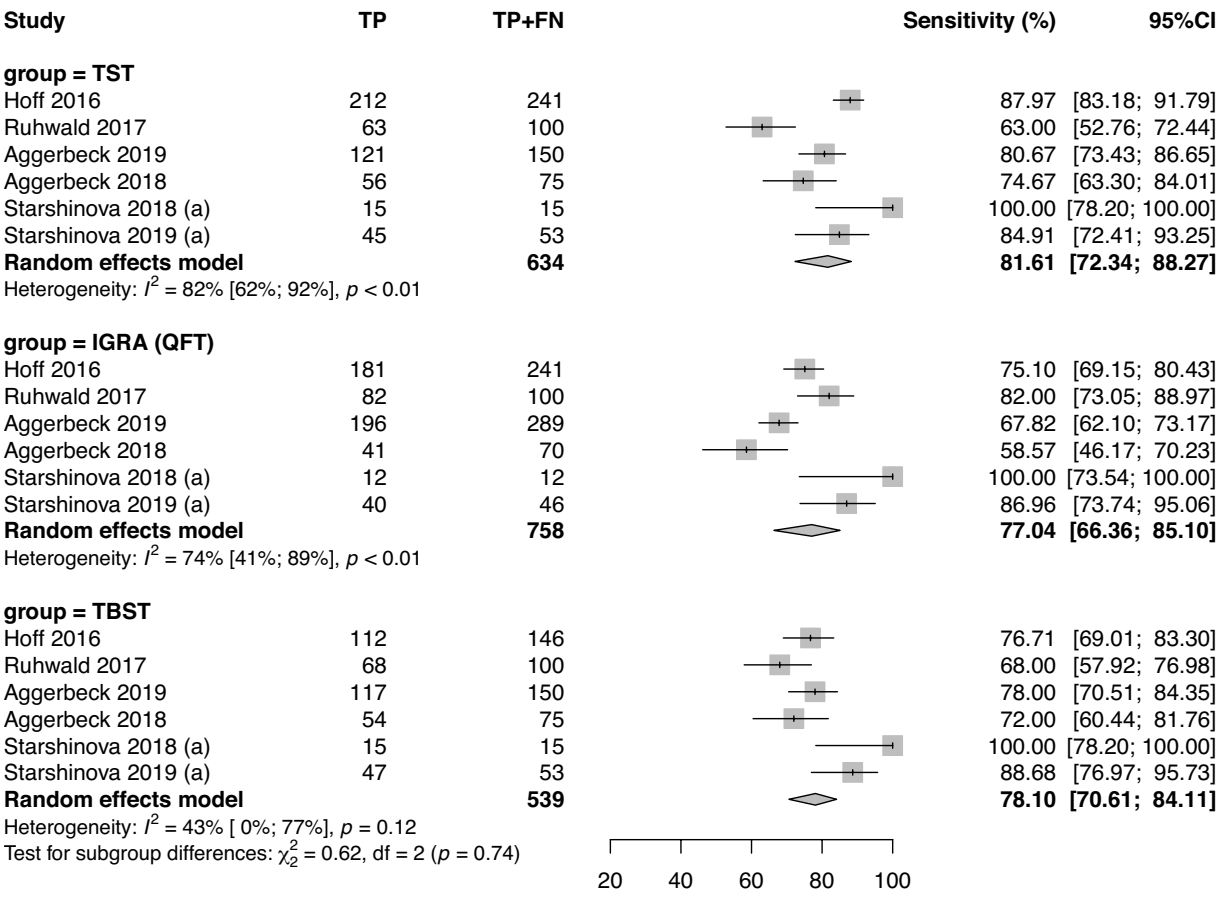
Diagnostic accuracy studies evaluating sensitivity, specificity and concordance (agreement) of TBSTs were identified. There were no identified studies on the efficacy of TPT based on diagnostic test results, on the predictive value for progression to TB disease or on the proportion started on TPT.

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. 1–4 (all of which are from Web Annex A).

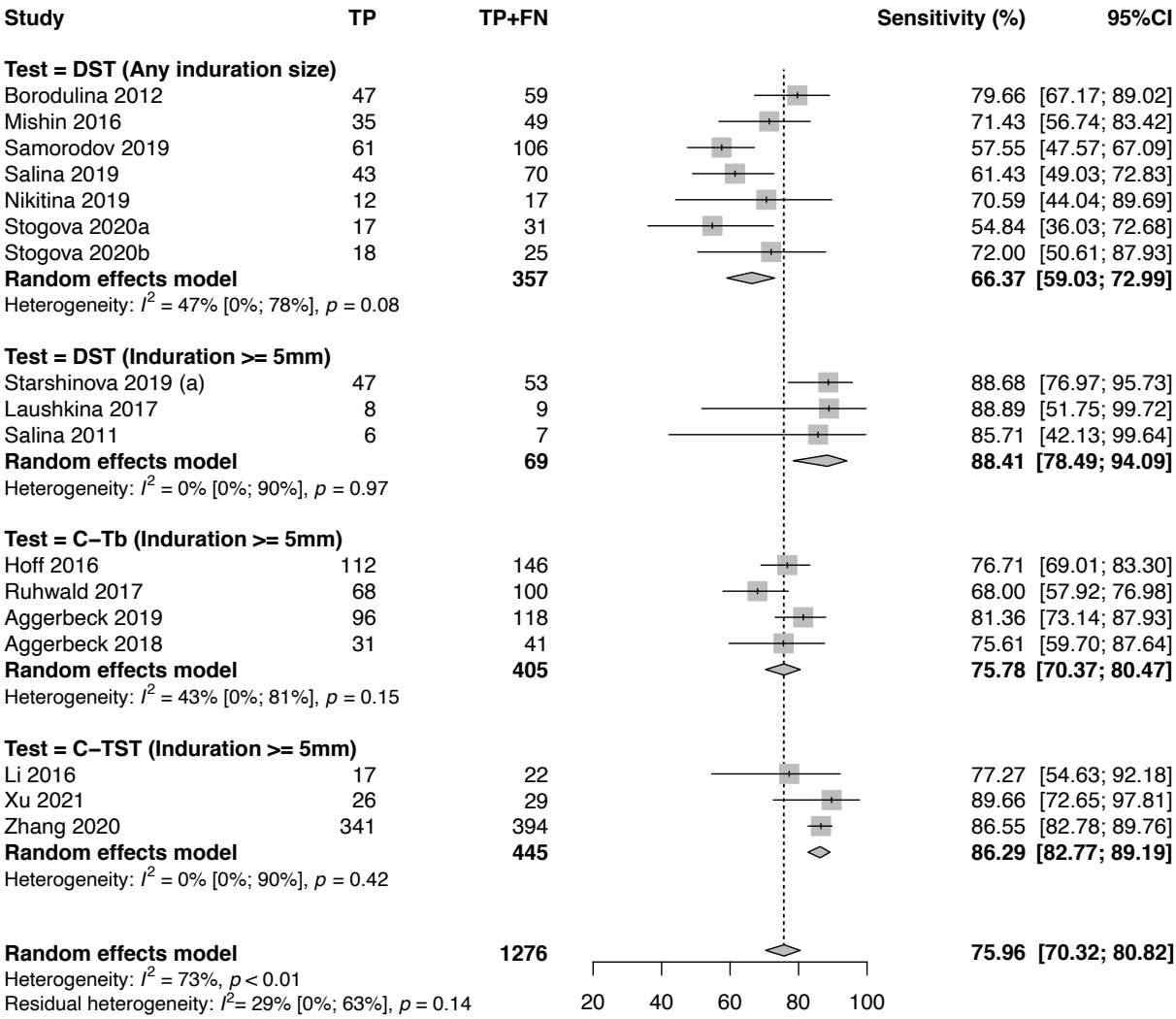
Fig. 1. Sensitivity of TBSTs in head-to-head studies



CI: confidence interval; FN: false negative; IGRA: interferon-gamma release assay; QFT: QIAGEN QuantiFERON; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TP: true positive; TST: tuberculin skin test.

The pooled sensitivity against the microbiological reference standard for TB disease in six head-to-head studies (Fig. 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. 2. Sensitivity of TBSTs in all studies in individuals with HIV-negative or unknown status



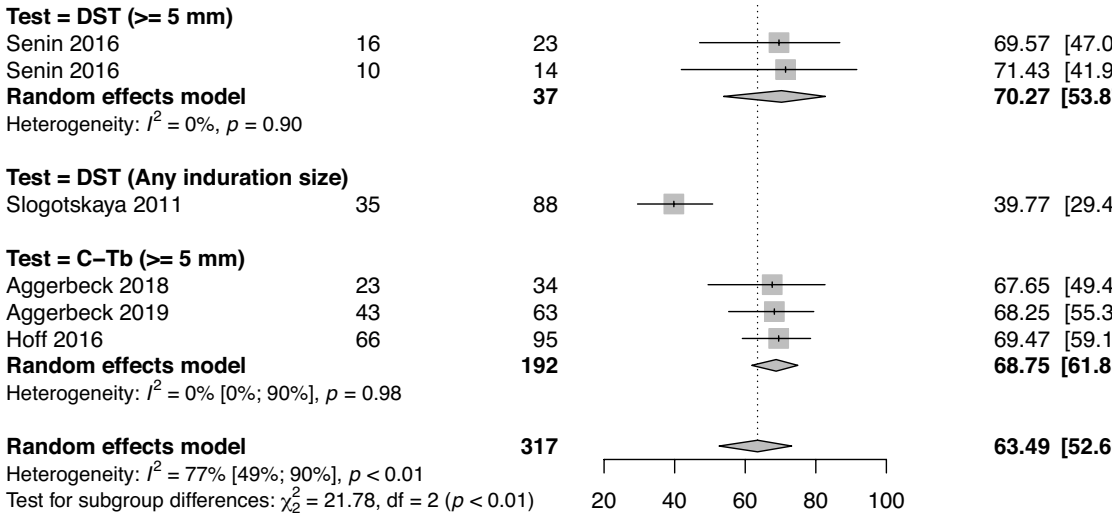
CI: confidence interval; DST: drug susceptibility testing; FN: false negative; HIV: human immunodeficiency virus; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TP: true positive.

The pooled sensitivity in 17 studies presented in Fig. 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.

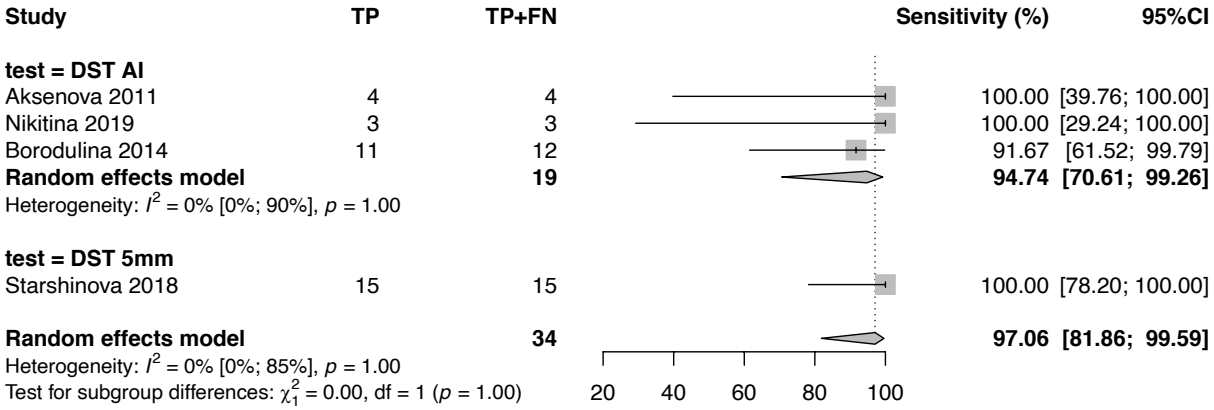
Fig. 3. Sensitivity of TBSTs in PLHIV



CI: confidence interval; DST: drug susceptibility testing; FN: false negative; HIV: human immunodeficiency virus; PLHIV: people living with HIV; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TP: true positive.

Only studies on Diaskintest and Cy-Tb were included in the analysis presented in Fig. 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. 4. Sensitivity of TBSTs in children and adolescents



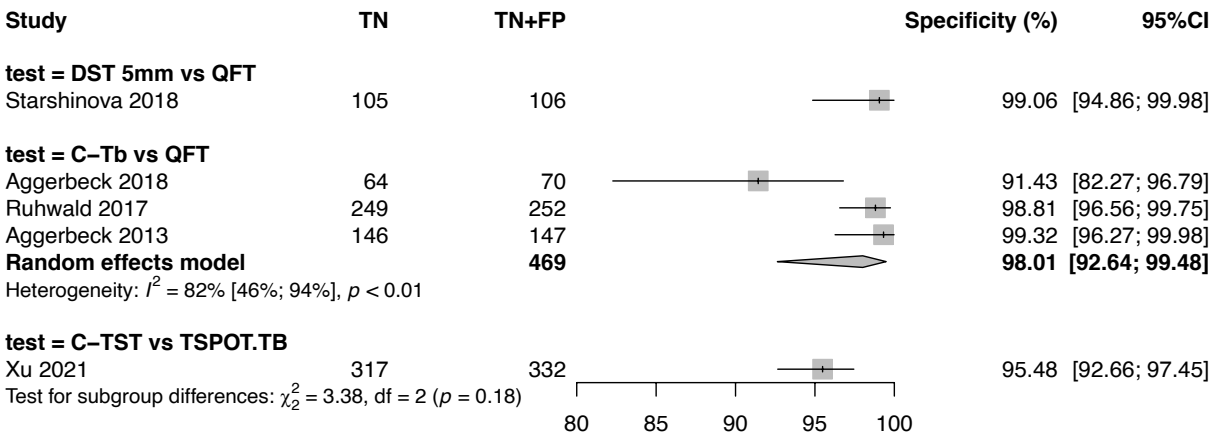
AI: any induration size; CI: confidence interval; DST: drug susceptibility testing; FN: false negative; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TP: true positive.

Sensitivity of TBSTs among children and adolescents is shown in Fig. 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. 5. Specificity in healthy individuals with negative IGRA results

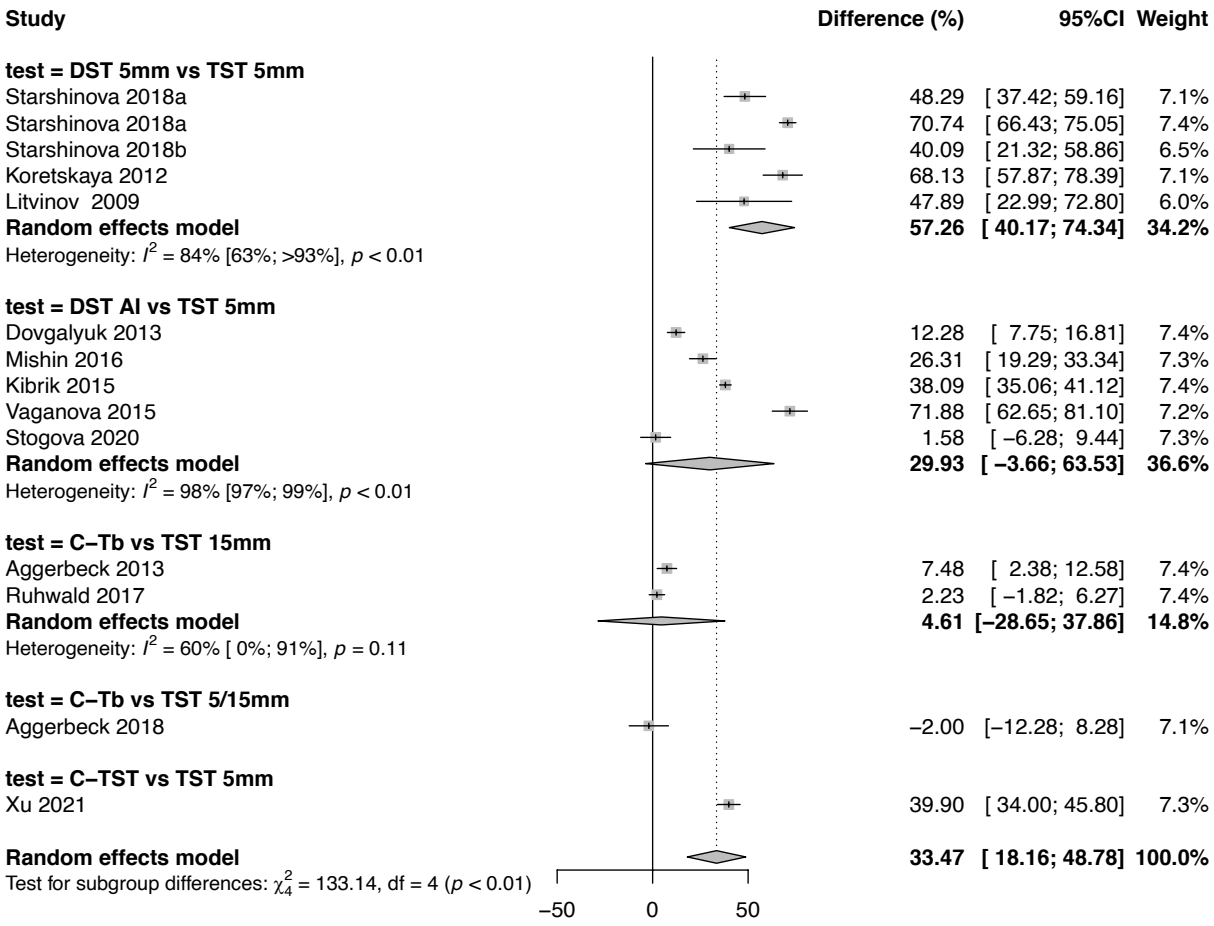


CI: confidence interval; DST: drug susceptibility testing; FN: false negative; IGRA: interferon-gamma release assay; QFT: QIAGEN QuantiFERON; TN: true negative; TST: tuberculin skin test.

The specificity assessed in the five studies presented in Fig. 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. 6. Difference in specificity – TBSTs versus the TST

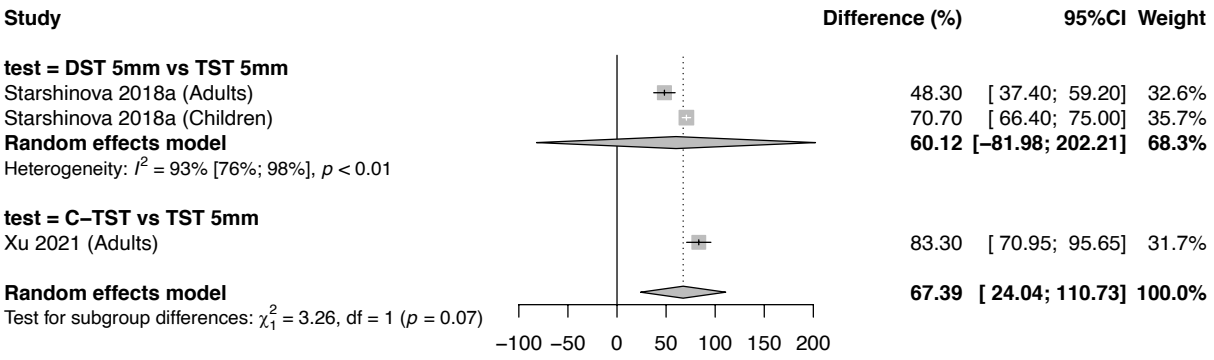


AI: any induration size; CI: confidence interval; DST: drug susceptibility testing; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TST: tuberculin skin test.

Values above "0" indicate higher specificity for TBSTs, while values below "0" indicate higher specificity for the TST.

The overall pooled difference in specificity in 14 studies (Fig. 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. 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. 7. Difference in specificity – TBSTs versus the TST in BCG-vaccinated population



BCG: bacille Calmette-Guérin; CI: confidence interval; DST: drug susceptibility testing; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TST: tuberculin skin test.

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.

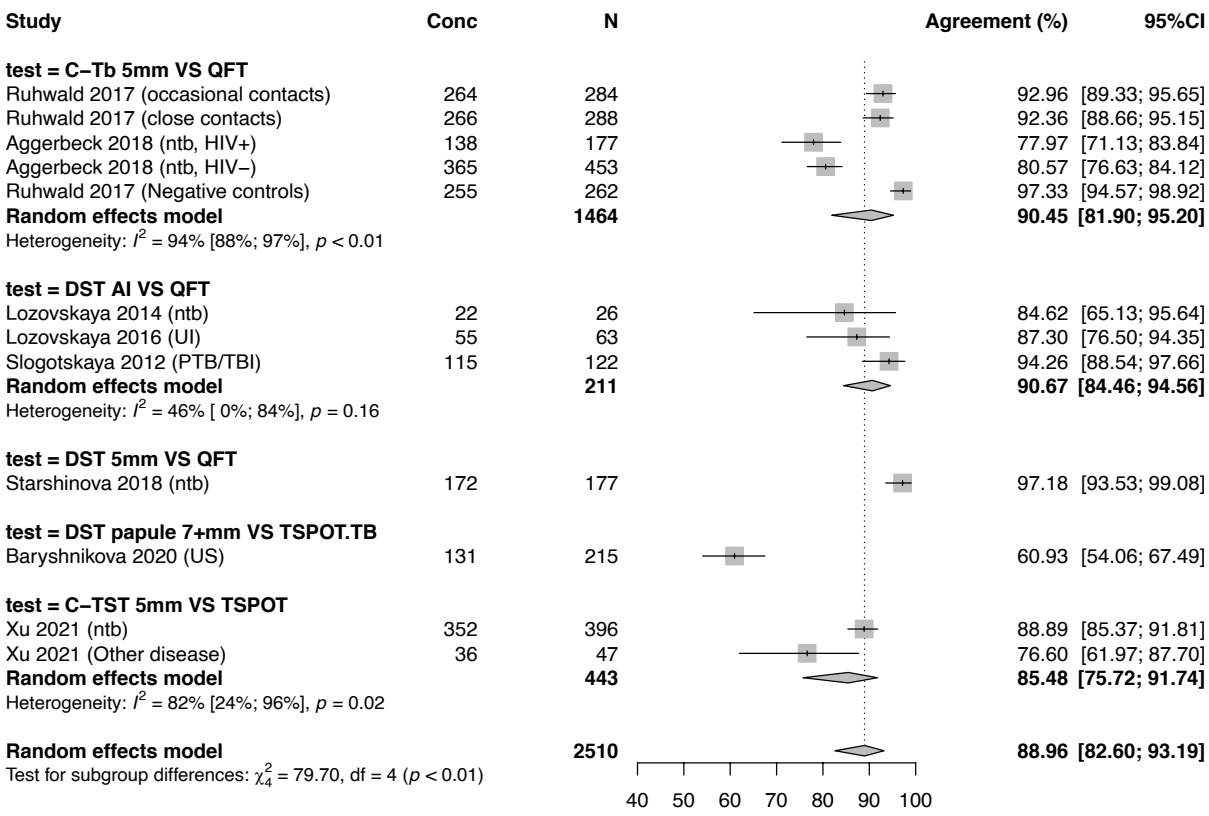
The pooled difference in specificity in six studies comparing TBSTs and IGRAs was low, at 2.3% (95% CI: –1.6–6.2%), meaning that TBSTs were similar to IGRAs in terms of specificity. More details can be found in Web Annex A.

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. 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.

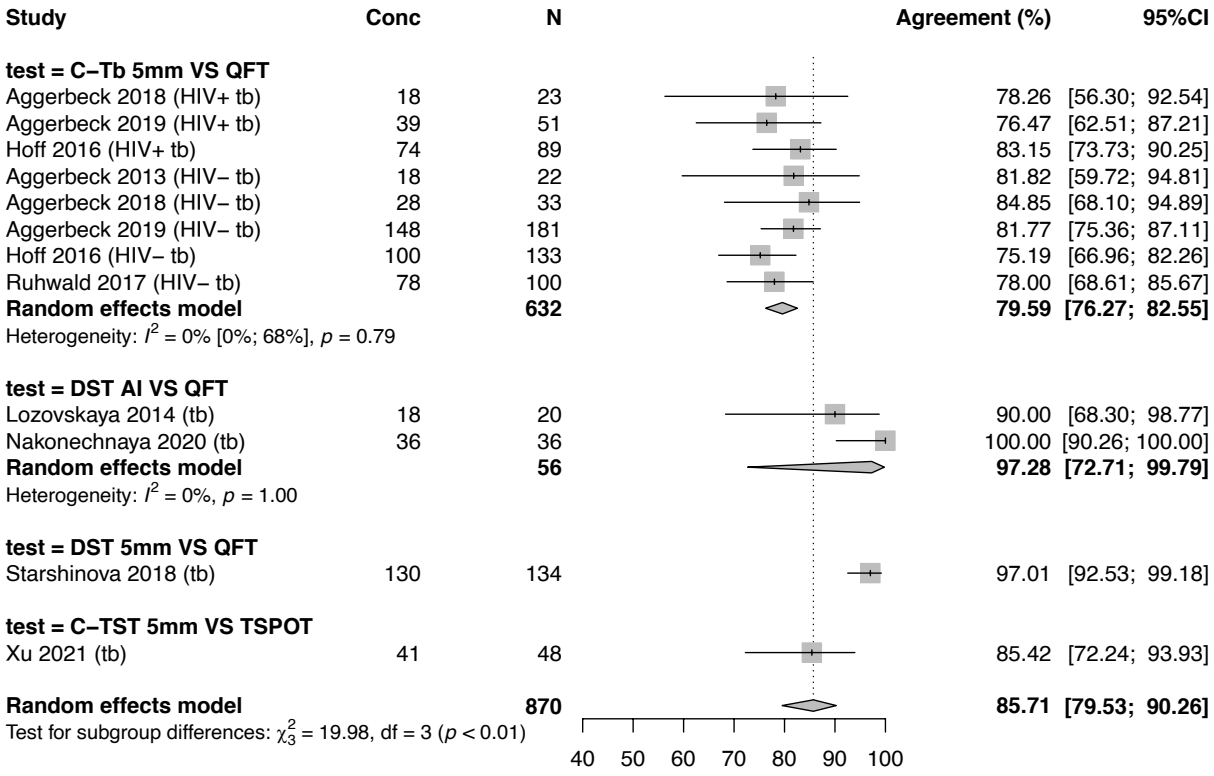
Fig. 8. Agreement of TBSTs versus IGRAs in all studies including participants without active TB



AI: any induration size; CI: confidence interval; DST: drug susceptibility testing; HIV: human immunodeficiency virus; IGRA: interferon-gamma release assay; QFT: QIAGEN QuantiFERON; TB: tuberculosis; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TST: tuberculin skin test.

In participants with TB disease, high agreement between TBSTs and IGRAs as the comparator (85.7%) was observed (Fig. 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. 9. Agreement of TBSTs versus IGRAs in all studies including people with active TB



AI: any induration size; CI: confidence interval; DST: drug susceptibility testing; HIV: human immunodeficiency virus; IGRA: interferon-gamma release assay; QFT: QIAGEN QuantiFERON; TB: tuberculosis; TBST: *Mycobacterium tuberculosis* antigen-based skin test; TST: tuberculin skin test.

1.4.2. 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.

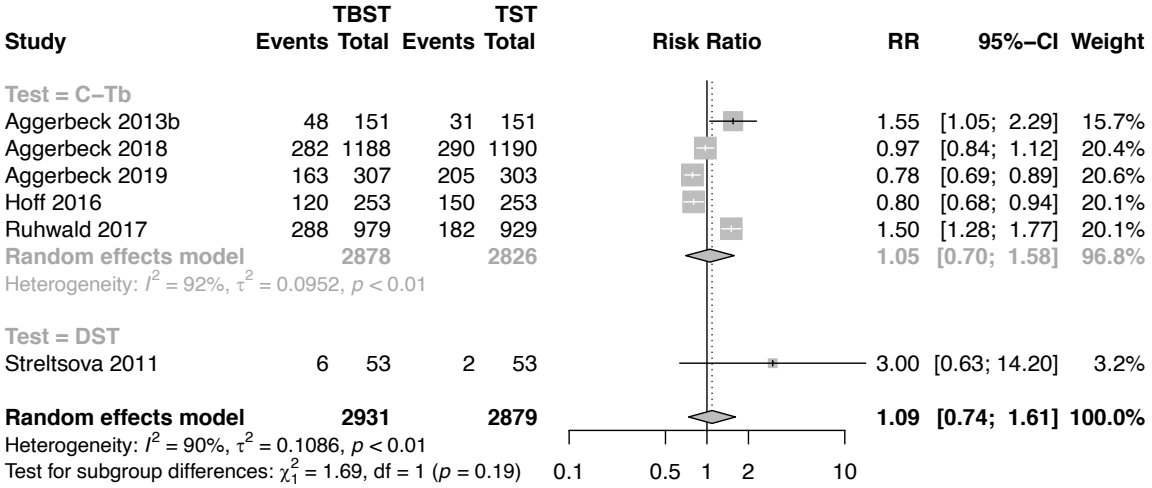
Characteristics of studies were as follows:

- 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. 10. Any injection site reactions



CI: confidence interval; DST: drug susceptibility testing; HIV: human immunodeficiency virus; PLHIV: people living with HIV; RR: risk ratio; TB: tuberculosis; TBSTs: *Mycobacterium tuberculosis* antigen-based skin tests; TST: tuberculin skin test.

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. 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.

1.4.3. 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

[5] Throughout this text, "$" signifies US dollars.

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.

1.4.4. 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.

1.5. 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.

1.6. 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.

1.7. 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.

2. Use of the TST and IGRAs for the diagnosis of TB infection

2.1. Background

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.

2.2. Recommendation

> 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*

2.2.1. 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.

2.3. Test descriptions

The following technologies were included in the evaluation:

- TST (RT23 PPD or PPD-S);
- QuantiFERON-TB Gold In-Tube (QFT-GIT, QIAGEN, Australia); and
- T-SPOT.*TB* (T-Spot, Oxford Immunotec, United Kingdom).

The original tuberculin material used by Mantoux in his first studies of tuberculin reactions was a heterogeneous mix of substances from killed *Mtb*. This so-called old tuberculin was replaced in the 1960s by a standardized preparation of purified protein, derived (hence the term "PPD") from *Mtb*. Florence Seibert produced a single standard lot of this material, termed PPD-S; subsequently, all newly produced tuberculin material has been produced using the same methods, and tested against PPD-S, measuring induration in sensitized guinea pigs.

PPD-S contains a mix of antigens, including some that are specific to *Mtb*, but also many that are found in NTM and BCG. Hence, false positive reactions to PPD-S have been described in people with NTM disease, in those sensitized to NTM antigens, or in those who have received BCG vaccination (particularly if they received BCG more than once, or after infancy).

All of the assays work on the principle that the T-cells of an individual who has acquired TB infection will respond to re-stimulation with *Mtb*-specific antigens by secreting IFN-γ. The QuantiFERON-TB Gold and the newer version QuantiFERON-TB Gold In-Tube are whole-blood based enzyme-linked immunosorbent assays (ELISAs) that use whole blood and measures the amount of IFN-γ produced in response to specific *Mtb* antigens (QFT-G: ESAT-6 and CFP-10; QFT-GIT: ESAT-6, CFP-10 and TB7.7). In contrast, the enzyme-linked immunospot (ELISPOT)-based T-Spot measures the number of peripheral mononuclear cells that produce INF-g after stimulation with ESAT-6 and CFP-10.

2.4. Evidence base

2.4.1. 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?

2.4.2. 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.

2.4.3. 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.

2.4.4. 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).

2.5. 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 not

susceptible 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.

2.6. 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. Use of the TST and IGRAs for the diagnosis of TB disease

3.1. Background

As explained in Chapter 1, TB infection is a state that is characterized by persistent immune response to stimulation by *Mtb* antigens with no evidence of clinically manifest TB disease *(1)*. Initially, the TST was the only tool available for TB infection detection.

The identification of genes in the *Mtb* genome that are absent from *M. bovis* BCG and most NTM has supported the development of more specific and sensitive tests for the detection of *Mtb*. The *M. bovis* BCG has 16-gene deletions, including the region of difference 1 (RD-1) that encodes for ESAT-6 and CFP-10, both of which are strong targets of the cellular immune response in patients with *Mtb* infection. In such people, sensitized memory or effector T-cells produce IFN-γ in response to these antigens, allowing a biological basis for T-cell-based tests such as IGRAs.

In 2011, WHO 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 for TB infection.

Among the WHO-recommended tests (i.e. the TST and IGRAs), the TST has some disadvantages, in that it has relatively low specificity in those with recent BCG vaccination and immunosuppressed individuals (e.g. PLHIV), requires two clinic visits and is only valid if the results are read within the suggested time frame. In contrast, IGRAs measure T-cell release of IFN-γ following stimulation by ESAT-6 and CFP-10 (antigens that are specific to *Mtb*). Unlike the TST, IGRAs are not affected by prior BCG vaccination or by infection with NTM (with a few exceptions), but IGRA platforms are more expensive to run, requiring specialized facilities and trained personnel. Thus, the TST is the most commonly used test for TB infection globally; however, recent global shortages of the TST have underscored the need for alternatives.

3.2. Recommendation

> 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.

3.3. Test descriptions

The following technologies were included in the evaluation:

- QuantiFERON-TB Gold (QFT-G, QIAGEN, Australia);
- QuantiFERON-TB Gold In-Tube (QFT-GIT, QIAGEN, Australia); and
- T-SPOT.*TB* (T-Spot, Oxford Immunotec, United Kingdom).

All of these assays work on the principle that the T-cells of an individual who has acquired TB infection will respond to re-stimulation with antigens specific to *Mtb* by secreting IFN-γ. The QuantiFERON-TB Gold, and the newer version QuantiFERON-TB Gold In-Tube, are whole-blood-based ELISAs that measure the amount of IFN-γ produced in response to specific *Mtb* antigens (QFT-G: ESAT-6 and CFP-10; QFT-GIT: ESAT-6, CFP-10 and TB7.7). In contrast, the ELISPOT-based T-Spot measures the number of peripheral mononuclear cells that produce INF-g after stimulation with ESAT-6 and CFP-10.

3.4. 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,[6] 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.

3.4.1. 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?

3.4.2. 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).

3.4.3. 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 IGRAs

[6] See www.gradeworkinggroup.org.

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.

3.4.4. 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.

3.4.5. 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.

3.4.6. 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.

3.5. 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.

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. Summary of changes between the 2011–2020 guidance and the 2022 update

2022	2020	2011	Changes
1. *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*			New; incorporated in the 2022 update as Recommendation 1.
2. Either the 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*	2. Either the 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*		None; incorporated in the 2022 update as Recommendation 2.
		4. IGRAs should not replace the TST in low- and middle-income countries for the diagnosis of latent TB infection in children and adolescents, nor for the diagnostic work-up of children and adolescents (irrespective of HIV status) suspected of active TB in these settings. *Strong recommendation*	2022 Recommendation 2 has superseded the 2011 recommendation.
		5. IGRAs should not replace the TST in low- and middle-income countries for the diagnosis of latent TB infection in individuals living with HIV infection. *Strong recommendation*	2022 Recommendation 2 has superseded the 2011 recommendation.

2022	2020	2011	Changes
		6. IGRAs should not be used in health care worker screening programmes in low- and middle-income countries. *Strong recommendation*	2022 Recommendation 2 has superseded the 2011 recommendation.
		7. IGRAs should not replace the TST in low- and middle-income countries for the screening of latent TB infection in adult and paediatric contacts, or in outbreak investigations. *Strong recommendation*	2022 Recommendation 2 has superseded the 2011 recommendation.
		8. Neither IGRAs nor the TST should be used in low- and middle-income countries for the identification of individuals at risk of developing active TB. *Strong recommendation*	2022 Recommendation 2 has superseded the 2011 recommendation.
3. 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*		3. IGRAs (and the 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 HIV-positive individuals) presumed of active TB in these settings. *Strong recommendation*	None; incorporated in the 2022 update as Recommendation 3.

HIV: human immunodeficiency virus; IGRA: interferon-gamma release assay; TB: tuberculosis; TST: tuberculin skin test.

Annex 2. GDG processes and decision-making

For every GDG meeting PICO questions were drafted by the WHO steering group and were presented to the respective GDG for discussion and modification. The WHO steering group has been making an initial list of patient-important outcomes, including desirable and undesirable effects and has solicited feedback from the Guideline Development Group to identify any other outcomes. The outcomes were further rated, according to the procedure described in *WHO handbook for guideline development (2nd edition)*[7]

The GDG meetings in 2010 and 2018 were conducted as face-to-face meeting, whereas in 2022 in the format of half day sessions for 10 days as a virtual meeting considering the situation with COVID-19 and related travel restrictions. A descriptive analysis of qualitative data, the economic analyses were presented during the introductory webinars along with the process for assessing the evidence for each of the PICO questions. In every meeting decisions were based on consensus (preferred option) or else by voting (with simple majority rule) only if consensus is not achieved. Concerns and separate opinions by members if any were noted and included in the final Evidence-to-decision tables. The guideline narrative were undergoing several iterations (managed by the WHO steering group).

In an online setting, the meeting participants were required to include their names as identifiers, so as various types of meeting participants, i.e. steering committee members, GDG members, systematic reviewers can be easily identified. Participation in discussion were prioritized for GDG members. Steering committee members and consultants with specialized technical expertise were invited by chair-persons to provide feedback when necessary. Observers participation were limited to the feedback after all above-mentioned categories of the meeting participants. All three last categories of the meeting participants were excluded from recommendation deliberation and voting.

Draft WHO policy guidance based on the consensus recommendations will subsequently be prepared by the WHO steering group and reviewed by both the Guideline Development Group and the External Review Group, before finalization.

[7] Handbook for Guideline Development 2nd Ed. Geneva: World Health Organization; 2014 (https://www.who.int/publications/i/item/9789241548960, accessed 1 June 2018)

Annex 3. Conflict of interest assessment for Guideline Development Group and External Review Group members

Before being considered for group membership, each candidate for the Guideline Development Group (GDG) and External Review Group (ERG) 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 might lead to compromising situations for the World Health Organization (WHO) and the expert concerned.

The candidate's curriculum vitae and DOI, and information retrieved from the internet, were examined by WHO steering committee members to assess whether there were, or might 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 – such involvement is automatically considered a conflict of interest.

Once a determination had been made that either no conflict of interest existed or that 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 were 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. The DOI findings are summarized in Table A1.

1. TB antigen-based skin tests for the diagnosis of TB infection

Table A1. Conflict of interest summary for TB antigen-based skin tests for the diagnosis of TB infection

GDG member	Interests declared	Conclusion
Helen Ayles	As part of an EDCTP-funded award, related to TB diagnostics, has received support from: • QIAGEN (technical assistance with training and a slightly reduced price on QFT-Plus test kits); the company has not participated in data management and analysis; and • Delft – training on the use of the digital X-ray facilities, as well as cost waivers to the CXR/CAD equipment (about US$ 10 000).	Conflict of interest not significant
David Branigan	None declared	No conflict of interest
Jeremiah Chakaya Muhwa	None declared	No conflict of interest
Daniela Cirillo	Participation in the 2020 Biomérieux advisory board (EUR 1000). San-Rafaele research unit has participated in: • the Biomérieux evaluation of blood stability for VIDAS, 2019 (EUR 11 200); and • the evaluation of XDR cartridge prototype, Cepheid / FIND (EUR 14 295).	Conflict of interest not significant
Frank Cobelens	As project coordinator, has managed a grant aimed at evaluating assays for prediction of incident TB among exposed household contacts. Total grant value US$ 3.3 million of which US$ 293 000 was for his research unit. Technologies involved in the project evaluation are not in the scope of the relevant GDG meeting.	Conflict of interest not significant
Anand Date	None declared	No conflict of interest
Petra de Haas	None declared	No conflict of interest

GDG member	Interests declared	Conclusion
Rumina Hasan	None declared	No conflict of interest
Farzana Ismail	None declared	No conflict of interest
Katharina Kranzer	Contributed to a 2-day advisory board meeting about IGRAs (QIAGEN) and their use in children – both for TB diagnostics and latent TB infection (EUR 1500) in May 2021. As part of EDCTP-funded project has received 2100 IGRA SD biosensor tests for free, for a diagnostic study in TB household contacts.	Conflict of interest not significant
Afranio Kritski	None declared	No conflict of interest
Blessina Kumar	None declared	No conflict of interest
Nagalineswaran Kumarasamy	None declared	No conflict of interest
Andrei Maryandyshev	None declared	No conflict of interest
Alberto Matteelli	None declared	No conflict of interest
Satoshi Mitarai	None declared	No conflict of interest
Lindiwe Mvusi	None declared	No conflict of interest
Mark Nicol	None declared	No conflict of interest
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.	Conflict of interest not significant
Hojoon Sohn	None declared	No conflict of interest
Sabira Tahseen	None declared	No conflict of interest
Ezio Távora dos Santos Filho	None declared	No conflict of interest
Carrie Tudor	None declared	No conflict of interest
Marieke van der Werf	None declared	No conflict of interest
Zhao Yanlin	None declared	No conflict of interest
ERG member	**Interests declared**	**Conclusion**
Francis Drobniewski	None declared	No conflict of interest
Francis Varaine	None declared	No conflict of interest

GDG member	Interests declared	Conclusion
Elisabetta Walters	None declared	No conflict of interest
Sergei Skornyakov	None declared	No conflict of interest

CAD: computer-aided detection; CXR: chest X-ray; EDCTP: European and Developing Countries Clinical Trials Partnership; ERG: External Review Group; FIND: Foundation for Innovative New Diagnostics; GDG: Guideline Development Group; IGRA: interferon-gamma release assay; TB: tuberculosis; USAID: United States Agency for International Development; WHO: World Health Organization.

2. Tuberculin skin test (TST) and interferon-gamma release assays (IGRAs) for the diagnosis of TB infection

The members of the GDG and ERG completed a WHO DOI form. All declarations were evaluated by the WHO steering committee for any financial conflict of interest that might warrant exclusion from membership or from certain discussions of the GDG. The completed forms were summarized and presented to all GDG members at the first meeting, at which point the members were asked to update their declarations. Intellectual conflict of interest was not considered a motive for exclusion from the GDG, because expertise on the topic was considered an important criterion for selection and the diversity and representation in the GDG was large enough to balance any individual member's intellectual interest. The biographies of the GDG members were made public alongside the background document outlining the 2019 update on 1 July 2019, ahead of the GDG meetings.

GDG

The following GDG members declared no interests that could conflict with the objectives of the guidelines: Mohammed Al Lawati, Rolando A. Cedillos, Diana Gibb, Yohhei Hamada, Nasehi Mahshid, Alberto Matteelli, Lindiwe Mvusi, Kuldeep Singh Sachdeva and Irina Vasilyeva.

The following GDG members declared interests that were judged not to conflict with the objectives of the meeting:

- Helen Ayles declared a research grant received by her institution from European and Developing Countries Clinical Trials Partnership (EDCTP) plus in-kind support for a project in which she is principal investigator (QFT test kits at a subsidized price from QIAGEN and support to try a new, simplified version of the QFT test). Delft diagnostics provides in-kind support in the order of US$ 100 000, to subsidize the cost of using its digital chest radiography and computer-aided detection. Helen is a member of the Technical Review Panel for the Global Fund to Fight AIDS, Tuberculosis and Malaria, which promotes adherence to the normative guidance of WHO.
- Padmapriyadarsini Chandrasekaran declared research grants received by her employer, the National Institute for Research in TB in Chennai, India, and collaboration, sponsorships and

other funding from the United States Agency for International Development (USAID) under a model directly observed treatment short course (DOTS) project. The study is now complete.

- Anthony D. Harries is a Senior Advisor at The Union, Paris, and was the lead author on a paper titled *Challenges and opportunities to prevent tuberculosis in people living with HIV in low-income countries (3)*. With other Union colleagues, he wrote a review titled *Treatment for latent tuberculosis infection in low- and middle-income countries: progress and challenges with implementation and scale-up (4)*, as part of unpaid work. The paper, which was under review at the time of the GDG meeting, deals largely with the programmatic requirements to scale up the WHO treatment recommendations for TB infection.
- Alexander Kay declared a research grant received by his organization from the United Nations Office for Project Services (UNOPS) and the Stop TB Partnership for a "TB Reach" project that is designed to increase the uptake of preventive therapy in children and adolescents exposed to TB (US$ 400 000). This work with the national TB programme (NTP) of Eswatini aims to enhance access to preventive therapy among household contacts, and includes nurse-led community-based TB screening and preventive therapy. A shorter preventive therapy regimen (3RH) is used for children and adolescents, and isoniazid in adults and children and adolescents living with HIV; however, no specific regimen is privileged.
- Nandi Siegfried declared consultation work with WHO.
- Ezio Távora dos Santos Filho declared delivering a talk at the Regional International AIDS Society conference in April 2018 in Mexico City on the need for advanced tools for TB infection treatment, without endorsing any particular study. He also declared that, as a TB advocate, he participated in many discussions with the Global TB Community Advisory Board and the Brazilian National TB Community Advisory Board (CAB) on the implementation of new TB infection treatment methods. The Brazilian TB CAB is now raising awareness of 3HP and TB infection. The affected communities may benefit directly from new guidelines.
- Marieke van der Werf declared the interest of her employer in the topic and that the European Centre for Disease Prevention and Control is working on TB infection.
- Wim Vandevelde declared travel support from the Stop TB Partnership for participation as a speaker in the TB/HIV symposium at the International AIDS Society conference in 2019 in Mexico City and that symposium will cover TB infection.

ERG

The following ERG members declared no interests that could conflict with the objectives of the guidelines: Stephen Graham, Giovanni B. Migliori, Rohit Sarin and Alena Skrahina.

The following ERG members declared interests that were judged not to conflict with the policy of WHO or the objectives of the meeting:

- Connie Erkens declared that a study she was involved in on TB infection screening in migrants to the Netherlands received 1800 QuantiFERON-TB Gold Plus tests from the manufacturer QIAGEN (value EUR 27 000) in 2016–2018. QIAGEN had a role in the study design, implementation, data collection and analysis, and in the decision to publish or in the preparation of the report.
- James Seddon declared that he is employed by Imperial College, London, United Kingdom of Great Britain and Northern Ireland (United Kingdom) to carry out research on childhood

TB; some of these studies involve the investigation of treatment of TB infection. He also has a grant from Global Trials Scheme (WT/MRC/DFID/HIHR) to carry out a trial of preventive therapy for children and adolescents exposed to multidrug-resistant TB (MDR-TB). He also has a fellowship from the Medical Research Council (MRC) to carry out studies evaluating correlates of risk in MDR-TB exposed children and adolescents. He is also involved in the TB-CHAMP trial (an MDR-TB prevention trial in children aged under 5 years) and has a personal fellowship to examine correlates of risk, using the trial as a research platform. His PhD examined MDR-TB preventive therapy in an observational cohort of children treated in Cape Town. He collaborates with the Unitaid grant CaP TB, which includes household contact tracing and preventive therapy. He has been involved in an application as a co-investigator to the National Institutes of Health (NIH) for an International Research Career Development Award (US$ 43 000) to look at implementation of 3HR in South Africa. He is collaborating on a modelling exercise (funded by the TB Modelling and Analysis Consortium [TB-MAC]) to look at implementation of different types of TB infection treatment, use of the TST and decisions about whom to treat in terms of risk–benefit and cost–effectiveness; also, he has collaborated on modelling the impact of household contact activities on childhood TB burden. The amounts were not disclosed. He is co-investigator in an NIH R01 study that has been submitted (Implementation of child contact management interventions to prevent TB in children and adolescents in Cape Town, South Africa); budget US$ 2.8 million. He has also written reviews on TB infection (including treatment).

- Carrie Tudor declared that she is employed with the International Council of Nurses, which received US$ 1 million from the Eli Lilly Foundation MDR-TB partnership to train nurses between 2013 and 2019.

Evidence reviewers

The evidence reviewers provided the estimates for the evidence summaries but did not participate in formulating the recommendations for policy. The following reviewer declared interests that were judged not to conflict with the policy of WHO or the objectives of the meeting:

- Lynne M. Mofenson declared that in 2018 she served as a consultant for WHO on the use of dolutegravir in pregnancy.

3. Use of interferon-gamma release assays (IGRAs) for the diagnosis of TB disease

Individuals were selected to be members of the Expert Group to represent and balance important perspectives for the process of formulating recommendations. The Expert Group therefore included technical experts, end-users, patient representatives and evidence synthesis methodologists. Interchange by Expert Group meeting participants was restricted to those who attended the Expert Group meeting in person, both for the discussion and follow-up dialogue.

Expert Group members were asked to submit completed DOI forms. These were reviewed by the WHO legal department before the Expert Group meeting. DOI statements were summarized by the co-chair (Karin Weyer, Stop TB Department) of the Expert Group meeting at the start of the meeting.

P. Hill and R. O'Brien declared conflicts of interest that were deemed to be insignificant: P. Hill declared receipt of kits from Cellestis (now QIAGEN) and Oxford Immunotec for research projects, and R. O'Brien declared Foundation for Innovative New Diagnostics (FIND) support to academia to develop a point-of-care serodiagnostic test, including the FIND biomarker discovery project.

Selected individuals with intellectual or research involvement (or both) in the use of IGRAs in low- and middle-income settings were invited as observers to provide technical input and answer technical questions. P. Godfrey-Fausett declared a research grant for the investigation of the use of the QuantiFERON-TB Gold In-Tube assay in South Africa and Zambia, and M. Pai declared conduct of research studies on IGRAs. These individuals did not participate in the GRADE evaluation process and were excluded from the Expert Group discussions when recommendations were developed. They were also not involved in the development of the final Expert Group meeting report, nor in preparation of the Strategic and Technical Advisory Group for Tuberculosis (STAG-TB) documentation or preparation of the final WHO policy statement.

The systematic reviewers (A. Cattamanchi, A. Date, A. Detjen, D. Dowdy, R. Menzies, J. Metcalfe, M. Pai, M. Rangaka, K. Steingart and A. Zwerling) were deemed to have a conflict of interest, and consequently were observers to the meeting, providing technical clarifications on the findings of the systematic reviews. They did not participate in the GRADE evaluation process, did not contribute to the meeting discussions where recommendations were developed, and did not provide comments on the final WHO policy statement.

References for Annex 3

1. Declarations of interest [website]. Geneva: World Health Organization; 2022 (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).
3. Harries AD, Schwoebel V, Monedero-Recuero I, Aung TK, Chadha S, Chiang CY et al. Challenges and opportunities to prevent tuberculosis in people living with HIV in low-income countries. Int J Tuberc Lung Dis. 2019;23(2):241–51 (https://www.ncbi.nlm.nih.gov/pubmed/30808459).
4. Harries AD, Kumar AMV, Satyanarayana S, Takarinda KC, Timire C, Dlodlo RA. Treatment for latent tuberculosis infection in low- and middle-income countries: progress and challenges with implementation and scale-up. Expert Rev Respir Med. 2020;14(2):195–208 (https://www.ncbi.nlm.nih.gov/pubmed/31760848).

For further information, please contact:

Global Tuberculosis Programme
World Health Organization
20, Avenue Appia CH-1211 Geneva 27 Switzerland
Web site: www.who.int/tb

9789240056084