Consensus Meeting Report

Development of a Target Product Profile (TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease

2017

World Health Organization

FIND
Because diagnosis matters
WHO collaborating centre for the evaluation of new diagnostic technologies

Consensus Meeting Report

Development of a Target Product Profile (TPP)
and a framework for evaluation for a test
for predicting progression from tuberculosis infection
to active disease

2017

World Health Organization

FIND
Because diagnosis matters
WHO collaborating centre for the evaluation
of new diagnostic technologies

WHO/HTM/TB/2017.18

© World Health Organization 2017

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

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

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization.

Suggested citation. Consensus meeting report: development of a Target Product Profile (TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease. Geneva: World Health Organization; 2017 (WHO/HTM/TB/2017.18). Licence: CC BY-NC-SA 3.0 IGO.

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

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see http://www.who.int/about/licensing.

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

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

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

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

Printed in Switzerland

Contents

Acknowledgements ..... **iv**

Participants in Technical Expert Consultation ..... **iv**

Abbreviations ..... **vi**

1. Introduction **1**

2. Scope **2**

3. Expert Consultation Process **3**

4. The evolving concept of LTBI diagnosis **4**

5. Target Product Profile: Test predicting progression from tuberculosis infection to active disease **8**

6. Framework for the evaluation for tests that predict progression from tuberculosis infection to active disease **10**

6.1 Clinical evaluation studies 10

6.2 Health impact studies 14

6.3 Summary of clinical and health impact studies 16

Annex 1: Target Product Profile: Test predicting progression from tuberculosis infection to active disease **17**

Acknowledgements

This document was prepared by Alberto Matteelli (University of Brescia), Sandra Kik (KNCV), Samuel Schumacher (FIND) and Alessandra Varga (FIND), with input from Christopher Gilpin and Alexei Korobitsyn (Global TB Programme, World Health Organization) on the basis of a New Diagnostic Working Group consensus building process used to develop target product profiles (TPP) for tests that can predict progression from latent TB infection to active TB disease. The TPP was drafted by Prathiba Seshadri and Claudia Denkinger (FIND) and refined by Samuel Schumacher (FIND) with the input from the New Diagnostic Working Group Taskforce on LTBI and stakeholders in the field.

Participants in Technical Expert Consultation

New Diagnostic Working Group Taskforce on LTBI Members

Catharina Boehme (FIND Geneva, Switzerland); Gavin Churchyard (Aurum Institute for Health Research, Johannesburg, South Africa); Frank Cobelens (AIGHD, Amsterdam, The Netherlands); Daniela Maria Cirillo (NDWG Co-Chair, San Raffaele Research Institute, Milan, Italy); Christopher Gilpin (Global TB Programme, World Health Organization, Geneva, Switzerland); Delia Goletti (National Institute for Infectious Diseases, Rome, Italy); Christian Lienhardt (Global TB Programme, World Health Organization, Geneva, Switzerland); Alberto Matteelli (Coordinator, NDWG TF on LTBI, University of Brescia, Brescia, Italy); Dick Menzies (remotely - McGill University, Montreal, Canada); Molebogeng Rangaka (University College London, London, United Kingdom); Samuel Schumacher (FIND, Geneva, Switzerland); Alessandra Varga (Secretariat NDWG, FIND, Geneva, Switzerland)

Technical Expert Members

Helen Ayles (LSHTM/Zambart, London, United Kingdom); Pauline Beattie (EDCTP, The Netherlands); Grania Brigden (The Union, Paris, France); Hanif Esmail (University of Oxford, Oxford, United Kingdom); David Lewinsohn (Chair Working Group on New Vaccines, University of Portland, Portland, USA); Thomas Scriba (remotely- University of Cape Town, Cape Town, South Africa); Marieke van der Werf (ECDC, Stockholm, Sweden); Kevin Winthrop (remotely - Oregon HS, Portland, USA); Jean-Pierre Zellweger (Swiss Lung Association, Bern, Switzerland)

Community representatives

Ruvandhi Nathavitharana (TB Proof, South Africa); Khairunisa Suleiman (Global TB Community Advisory Board Diagnostics Workgroup, Nairobi, Kenya)

Country representatives

Norbert Ndjeka (Drug-Resistant TB, TB & HIV National Department of Health, Pretoria, South Africa); Rohit Sarin (National Institute of TB & Respiratory Diseases, New Delhi, India); Irina Vasilieya (Ministry of Health, Moscow, Russian Federation); Nguyen Viet Nhung (National TB Programme Manager, Director of National Lung Hospital, Hanoi, Vietnam)

Donor Agencies

René Becker-Burgos (Global Fund, Geneva, Switzerland); Thomas Forissier (Bill & Melinda Gates Foundation, Seattle, USA)

Diagnostic test developers (Observers)

Alexandra Asbach-Nitzsche (Lophius, Germany); Jeff Boyle (Qiagen, USA); Anke Coblenz (Abbott, USA); William Cruikshank (Oxford Immunotec, United Kingdom); Philippe Jacon (Cepheid, France); Masae Kawamura (Qiagen, USA); Oksana Markova (Generium, Russian Federation); Chris Novak (Roche, Switzerland); Morten Ruhwald (Staten Serum Institute, Denmark); Andrae Vinson (BD, USA)

WHO Secretariat (WHO Global TB Programme)

Haileyesus Getahun, Christopher Gilpin, Yohhei Hamada, Alexei Korobitsyn, Karin Weyer

Target Product Profile – External Review Group

Sunday Agbochenu Aboje (HIV/AIDS Division, Federal Ministry of Health, Nigeria); Ibrahim Abubaker (Institute for Global Health, University College London, London, United Kingdom); Mohammed Rheda Al Lawati (Consultant Physician, Oman); Peter Henrik Andersen (Statens Serum Institut, Denmark); Judith Bruchfeld (Karolinska Institute and Karolinska University Hospital, Sweden); Rolando Cedillos (Servicio de Infectología y Programa de Atención Integral en ITS/ VIH/SIDA Hospital Nacional Rosales, El Salvador); Richard Chaisson (Center For Tuberculosis Research, John Hopkins University, USA); Lucy Chesire (TB Advocacy Consortium, Kenya); Ray Cho (Department of Microbiology, Yonsei University College of Medicine, Republic of Korea); Gavin Churchyard (The AURUM Institute, South Africa); Daniela Maria Cirillo (WHO Collaborating Centre and TB Supranational Reference Laboratory, San Raffaele Scientific Institute, Italy); Thierry Comolet (Direction Générale de la Santé, lutte contre la tuberculose, Ministère du travail, de l'emploi et de la santé, France); Anand Date (Centers for Diseases Control and Prevention, USA); Gerard de Vries (KNCV Tuberculosis Foundation, The Netherlands); Claudia Denkinger (Foundation for Innovative New Diagnostics (FIND), Switzerland); Betina Durovni (Federal University of Rio de Janeiro, Brazil); Serge Eholie (Treichville Hospital, University of Abidjan, Côte D'Ivoire); Negussie Eyerusalem (Federal Ministry of Health, Ethiopia); Mike Frick (Treatment Action Group, USA); Samarn Futakul (Director of the Bureau of AIDS TB and STIs, Department of Disease Control, Ministry of Public Health, Thailand); Mina Gaga (Athens Chest Hospital, Greece); Un-Yeong Go (Korea Centers for Disease Control and Prevention, Republic of Korea); Stephen Graham (Center for International Child Health, University of Melbourne, Australia); Endang Budi Hastuti (National HIV AIDS and STI Programme, Ministry of Health, Indonesia); Einar Heldal (National TB Register, National Screening Services, Norway); Adeeba Kamarulzaman (University of Malaya, Malaysia); Phillipe LoBue (Centers for Disease Control and Prevention, USA); Richard Menzies (Montreal Chest Institute MUHC and McGill University, Canada); Giovanni Battista Migliori (WHO Collaborating Centre for TB and Lung Diseases, Maugeri Care and Research Institute, Italy); Beatrice Mutayoba (National Tuberculosis and Leprosy Programme, Ministry of Health, United Republic of Tanzania); Lindiwe Mvusi (National Department of Health, South Africa); Ivan Solovic (National Institute for TB, Lung Diseases and Thoracic Surgery, Vysne Hagy, Slovakia); Giovanni Sotgiu (Clinical Epidemiology and Medical Statistics Unit, Department of Biomedical Sciences, University of Sassari, Italy); Tim Sterling (Vanderbilt University School of Medicine, USA); Alistair Story (TB Screening Service Find & Treat, United Kingdom); Marieke van der Werf (European Centre for Disease Prevention and Control (ECDC), Sweden); Wim Vandevelde (Global TB Community Advisory Board, South Africa); Nguyen Van Hung (National TB Reference Laboratory, National Lung Hospital, Viet Nam); Kitty van Weezenbeek (KNCV Tuberculosis Foundation, The Hague, Netherlands); Tuula Vasankari (Finnish Lung Association, Finland); Constantia Voniatis (Clinical Laboratories, Ministry of Health, Cyprus); Maryse Wanlin (Belgian Lung and Tuberculosis Association (BELTA), Belgium); Brita Askeland Winje (Department of Infectious Disease Epidemiology, Norwegian Institute of Public Health, Norway); Takashi Yoshiyama (Research Institute of Tuberculosis, Japan).

Abbreviations

BCG	Bacille Calmette–Guérin
FIND	Foundation for Innovative New Diagnostics
HEOR	health economics and outcomes research
HI	health informatics
HIV	human immunodeficiency virus
IGRA	interferon-gamma release assays
IPT	intermittent preventive treatment
ITT	incipient TB test
LTBI	latent TB infection
MDR-TB	multidrug-resistant tuberculosis
MTB	*M. tuberculosis*
NDWG	New Diagnostics Working Group
NGO	Nongovernmental organization
NPV	negative predictive value
PIT	persistent infection test
PLHIV	Persons living with HIV
PPV	positive predictive value
PT	preventive treatment
TB	tuberculosis
TBI	TB infection
TEG	Technical Expert Group
TPP	target product profile
TST	Mantoux tuberculin skin test
WHO	World Health Organization

1. Introduction

The targets of the WHO End-TB Strategy will not be achieved without addressing diagnosis and treatment of latent TB infection (LTBI)[1]. It is essential to develop newer diagnostic tests with significantly increased predictive value for the development of active disease among those who are infected than the currently available tests for LTBI[2]. Of equal importance is establishing a consensus on the terminology and definitions dealing with LTBI.

Preventive treatment of persons at risk is among key components of the first pillar of the WHO End TB strategy 2016-2035[3]. One forth[4] to one third[5] of the world's population is infected with *M. tuberculosis* (MTB). Infected individuals are at risk of endogenous reactivation of the same strain and progression to active tuberculosis (TB) disease. The lifetime risk of developing TB among infected individuals is between 5 and 15 per cent with the highest risk in the first two years after infection[6,7].

While current diagnostic tests for infection (tuberculin skin test - TST/ Interferon Gamma Release Assays – IGRAs) show that an individual has been exposed to MTB, they poorly predict whether an individual will progress to active TB in the future[8]. This translates into a high number of individuals who would need to be treated in order to prevent one case of active TB and as such is a barrier to further scale-up of the programmatic management of LTBI.

Diagnostic tests that are highly predictive of development of the disease in the near future are urgently needed. An ideal test of progression would likely differentiate patients in the various stages from infection to active TB, and it may detect the presence or absence of incipient TB (defined as the prolonged asymptomatic phase of early disease during which pathology evolves, prior to clinical presentation as active disease).

1 Getahun H, Matteelli A, Abubakar I, Hauer B,Pontali E, Migliori GB. Advancing global programmatic management of latent tuberculosis infection for at risk populations. Eur Respir J 2016;47(5):1327–30.
2 Guidelines on the management of latent tuberculosis infection. Geneva: World Health Organization; 2015 (WHO/HTM/TB/2015.01; http://apps.who.int/iris/bitstream/10665/136471/1/9789241548908_eng.pdf?ua=1&ua=1, accessed 18 July 2017).
3 Implementing the END-TB Strategy: the essentials. Geneva:World Health Organization; 2015 (WHO/HTM/TB/2015.31; http://www.who.int/tb/publications/2015/end_tb_essential.pdf?ua=1, accessed 18 July 2017).
4 Houben RM, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. PLoS Med 2016;13(10):e1002152.
5 Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA 1999;282(7):677–86.
6 Getahun H, Matteelli A, Chaisson RE, Raviglione M. Latent Mycobacterium tuberculosis infection. N Engl J Med 2015;372(22):2127–35.
7 Trauer, J. M., Moyo, N., Tay, E.-L., Dale, K., Ragonnet, R., McBryde, E. S., & Denholm, J. T. (2016). Risk of Active Tuberculosis in the Five Years Following Infection . . . 15%? Chest, 149(2), 516–525. http://doi.org/10.1016/j.chest.2015.11.017
8 Matteelli A, Sulis G, Capone S, D'Ambrosio L, Migliori GB, Getahun H. Tuberculosis elimination and the challenge of latent tuberculosis. Presse Med. 2017 Mar;46(2 Pt 2):e13-e21.

2. Scope

To support the development of novel tests for predicting the risk of progression from latent infection to active disease, guidance is needed to inform test manufacturers, researchers and research funders regarding the nature and significance of LTBI and the relevant implications for the development of new diagnostic technologies. The document presents a Target Product Profile (TPP) for a test of progression of LTBI that defines key specifications, such as intended use, performance and operational characteristics, and pricing; along with a framework for the evaluation of tests that predict progression to active TB disease using standard study designs and evaluation protocols.

3. Expert Consultation Process

In May 2015, an Expert Consultation was convened by the World Health Organization (WHO) Geneva on behalf of the New Diagnostics Working Group, Stop TB Partnership (NDWG) and FIND to identify the operational and performance characteristics of tests that could predict progression from latent TB infection to active TB disease. Members of the Expert Consultation identified the following two objectives: i) develop a target product profile (TPP) for a test of progression to provide a framework for test development; and ii) develop guidance on the type of studies that would be needed to assess the performance of a test of progression to generate evidence suitable for evaluation by WHO.

A NDWG Task Force on LTBI was subsequently established and was convened at an Expert Consultation at the San Raffaele Scientific institute, Milan, Italy in July 2016. The purpose of the NDWG LTBI taskforce meeting was to develop consensus on new definitions of LTBI and to review the minimal and optimal performance characteristics of relevant diagnostics described in an advanced draft of the TPP for a test of progression of LTBI. Preliminary guidance on suggested study designs to assess the performance of tests of progression was also presented to the taskforce on LTBI.

On February 8, 2017 the Global TB Programme at WHO convened a final Expert Consultation on behalf of the NDWG in Geneva, Switzerland to reach consensus on the two documents in a face-to-face stakeholder meeting. Participants were selected to ensure a broad representation of all stakeholders and beneficiaries, including representatives of the NDWG taskforce on LTBI, as well as experts from high and low TB and HIV burden countries, funding agencies, test developers, community representatives, scientific associations, industry, education, and the non-profit sector. The methodology to reach the overarching goal of achieving final consensus in the face-to-face stakeholder meeting incorporated guided discussions on available draft documents and on-line survey results.

4. The evolving concept of LTBI diagnosis

Current tests for latent TB infection (LTBI), the tuberculin skin test (TST) and interferon gamma release assays (IGRAs) provide evidence of an immune memory response to *Mycobacterium tuberculosis* (MTB) rather than confirming the presence of viable organisms. The capacity of these tests to predict incident tuberculosis is very low, so that a high number of individuals need to receive treatment in order to prevent one case of active disease. In one meta-analysis, the pooled positive predictive value (PPV) of the TST to predict active TB disease occurring within two years was 1·5% and the number needed to treat (NNT) in order to prevent one TB case through preventive therapy was 67.3[9]; IGRAs performed slightly better, with a PPV of 2·7% and a NNT of 37.3.

To facilitate programmatic scale-up of LTBI diagnosis and treatment, new diagnostic tools are needed that are unaffected by prior BCG vaccination or exposure to nontuberculous bacteria, and can achieve a much higher PPV for predicting incident TB. A recent paper discussed whether such a test could potentially be developed, based on the latest understanding of the nature of MTB latency and the relevant implications for diagnosis[10].

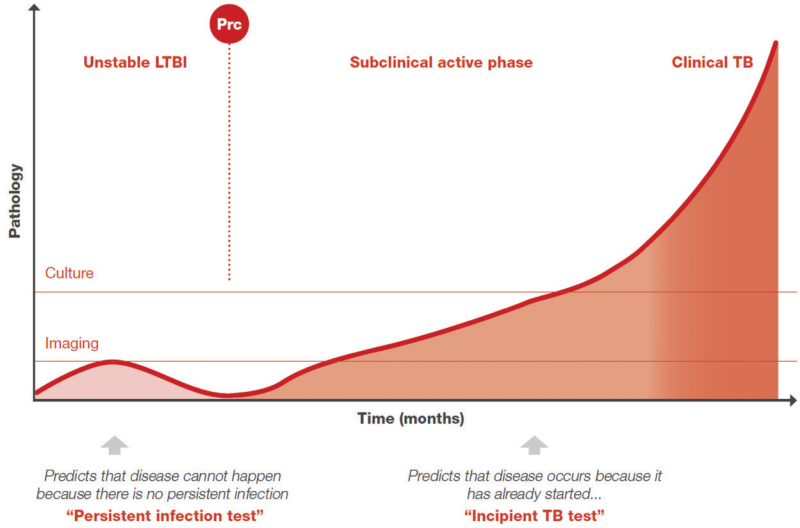
It is now widely recognized that a clear distinction between active disease (a symptomatic and potentially infectious state with evidence of pathology resulting from ineffective control of bacillary replication) and latent tuberculosis infection (an asymptomatic state in which bacillary replication is controlled) does not exist[11]. Recent research postulates the existence of a spectrum from spontaneous clearance to quiescent infection and disease. Patients position on this spectrum will be defined by their capacity to control bacillary replication[12] (Figure 1).

Following infection, there may be a critical period where the fate of infection is determined by predisposing factors (including HIV, malnutrition, diabetes, alcoholism and young age) influencing this outcome. In a small proportion, the primary infection may be progressive; in those that control primary infection, a proportion may eliminate TB or exert highly effective control and be at very low risk of reactivation. In the third group, control may be unstable, waxing and waning in response to a variety of precipitating factors (Prc) with reactivation of TB most likely to occur in this group. The conditions that currently identify at-risk populations have low relative risk for active disease development, and are unlikely to be sufficient drivers of the transition towards disease[13]. Possible precipitating factors include HIV infection, treatment with tumour necrosis factor-α antagonists, malnutrition, vitamin D deficiency and viral infection. However, other unidentified factors may remain that trigger reactivation or rapid progression to disease through failure of host defenses.

The postulate that, prior to clinical presentation with active disease, there might be a prolonged asymptomatic phase of early disease during which pathology evolves is now widely accepted. This state identifies incipient tuberculosis. Data from community surveys suggest that bacilli might be shed in the sputum for approximately a year before clinical presentation[14]. Incipient tuberculosis might involve periods of healing

9 Diel R, Loddenkemper R, Nienhaus A. Predictive value of interferon-γ release assays and tuberculin skin testing for progression from latent TB infection to disease state: a meta-analysis. Chest 2012; 142: 63–75.
10 Cobelens F, Kik S, Esmail H, Cirillo DM, Lienhardt C, Matteelli A. From latent to patent: rethinking prediction of tuberculosis. Lancet Respir Med. 2017 Apr;5(4):243-244.
11 Esmail H, Barry CE, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Philos Trans R Soc Lond B Biol Sci 2014; 369: 20130437.
12 Esmail H, Lai RP, Lesosky M, et al. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[(18)F]fluoro-D-glucose positron emission and computed tomography. Nat Med 2016; 22: 1090–93
13 Dheda K, Barry CE, Maartens G. Tuberculosis. Lancet 2016; 387: 1211–26.
14 Wood R, Middelkoop K, Myer L, et al. Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control. Am J Respir Crit Care Med 2007; 175: 87–93.

Figure 1. The postulated spectrum of TB infection and the progression to active TB disease (adapted from Esmail et al. 2014)



Precipitating factors (Prc) may lead to progression of disease. Prior to presentation these individuals may pass through a subclinical phase of active infection which may last months; during this phase *M. tuberculosis* may be isolated by culture or pathology may be visible through imaging prior to symptomatic presentation.

and disease regression as evidenced by radiographic and pathological findings of inactive fibrotic scarring[15] and some individuals with incipient tuberculosis might not progress to active disease for 12 months or longer.

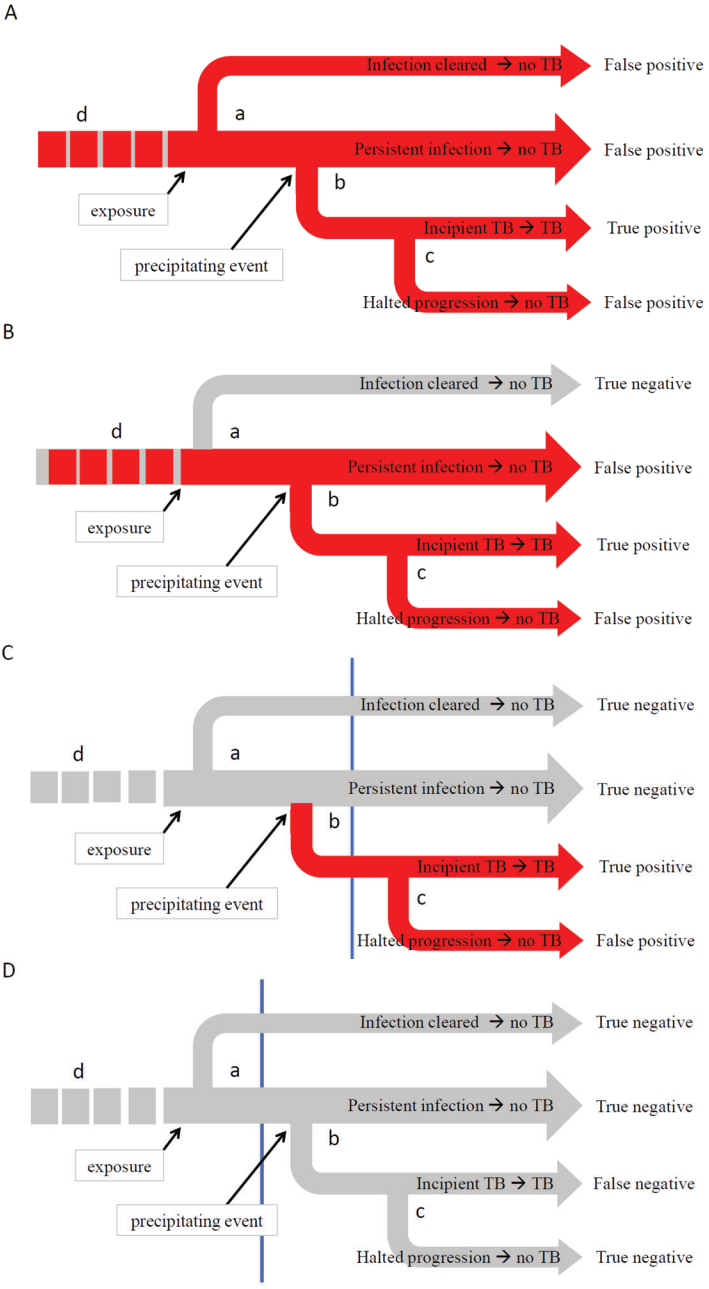
Based on these assumptions, diagnostic tests for the identification of latent tuberculosis infection should be conceptually categorised as persistent infection tests (PIT) versus incipient tuberculosis tests (ITT). Figure 2 gives a graphic representation of the theoretical performance of PIT and ITT. The distinction of these two categories of LTBI tests is important, as their performance, use, and design requirements differ, affecting the preparation of Target Product Profiles.

An **immune memory response** (Figure 2A) remains positive after infection regardless of spontaneous clearance. A test of **persistent infection** (Figure 2B) upon infection will turn negative if the infection is spontaneously cleared but will otherwise remain positive. A test of **incipient TB done after the precipitating event** (Figure 2C) will be positive if progression to TB disease has started regardless of whether progression is spontaneously halted. A test of **incipient TB done before the precipitating event** (Figure 2D) will be negative even though progression to TB disease will subsequently occur. For each test the positive predictive value (PPV) is the ratio [number of true positives]/[number of true positives + number of false positives].

PITs would probably measure persistent antigenic stimulation. As persistent infection is a necessary condition for active TB, PITs have high sensitivity for tuberculosis disease developing in the near future. However, their PPV is low to moderate and population-dependent. It is lower if more infected individuals remain with persistent infection over time or have acquired their infection remotely rather than recently. Expectedly, PPV is lower in high-incidence populations than in low-incidence populations.

15 Opie EL, Aronson JD. Tubercle bacilli in latent tuberculous lesions and in lung tissue without tuberculous lesions. Arch Pathol Lab Med 1927; 4: 1.

Figure 2 (A,B,C,D): Schematic of test results for immune memory response for a test of persistent infection and for a test of incipient TB as predictor of progression to tuberculosis disease



Schematic of test results for immune memory response (A), for a test of persistent infection (B) and for a test of incipient TB (C-D) as predictor of progression to tuberculosis disease. The red colour denotes a positive test. TB: tuberculosis disease, i.e. symptomatic disease with evidence of pathology. Exposure: moment at which individual is exposed to *Mycobacterium tuberculosis*. Precipitating event: event that results in failure of host control of infection. True positive: the test is positive and TB disease occurs. True negative: the test is negative and no TB disease occurs. False positive: the test is positive but no TB disease occurs. False negative: the test is negative but TB disease does occur. Lower case letters denote probabilities: a – probability that infection spontaneously cleared; b – probability that incipient TB occurs; c – probability that progression from incipient TB to TB disease is spontaneously halted; d – probability that infection had occurred upon previous exposure. Vertical blue line: moment when test is done (panels C and D only).

This has indeed been observed for IGRAs[8,16]. IGRAs likely belong to PIT rather than ITT; however, they probably cannot discriminate infections that have been cleared. PITs act very well as rule-out tests: whereas a positive result might not be very informative, a negative result provides confidence that the individual is unlikely to develop tuberculosis disease in the near future.

ITTs would probably detect mycobacterial replication or the resulting inflammatory response. Provided that analytical performance is adequate, the specificity and PPV of an ITT will be high, population-independent, and determined primarily by the probability that asymptomatic progression is halted spontaneously. Timing is crucial for ITTs: sensitivity greatly varies if the test is applied before or after the precipitating event occurred. Sensitivity and specificity (and thus PPV) of an ITT are higher, the closer the test is performed to the point of clinical presentation of tuberculosis.

Recently, a 16-transcript blood signature published by Zak and colleagues responded to all the above expectations, suggesting that this could be the first ITT ever described[17]. ITTs should be considered rule-in tests: a negative result provides limited information but a positive result indicates that TB will probably develop.

ITTs may not perform equally well for all disease states (e.g. localised disease compared with, pulmonary or disseminated TB) or in all patient groups (e.g. HIV positive individuals compared with HIV negative individuals, or among adults compared with children). This may be because the biological processes in the context of host biomarkers (e.g. RNA signature) that precede disease presentation may differ between these group or the extent to which this is detectable in a particular sample (e.g. blood) may differ.

The changing paradigm of latent tuberculosis infection as a spectrum leading to disease progression implies that two complementary types of test with different purposes are needed.

PITs would be used as rule-out tests in individuals at high risk of developing severe TB irrespective of when they were infected, such as those with HIV infection or starting anti-tumour necrosis factor-α treatment. IGRAs are very good examples of PIT. An improved PIT would be non-reactive whenever infection was cleared. For example, such PIT would turn negative after effective treatment for MTB infection. Improved PITs would be important for clinical use.

Conversely, ITTs would best be used as rule-in tests for screening of those who have been recently exposed to MTB, such as contacts of infectious tuberculosis patients. ITTs might need to be repeated to increase sensitivity. They should therefore be inexpensive and easy to perform, and ideally have a semi-quantitative readout reflecting the bacterial burden to allow informed decisions about preventive versus full-course treatment. ITTs would potentially be important new tools in public health, allowing scale-up of contact tracing strategies and mass test-and-treat campaigns in high-transmission settings that could have substantial impact on tuberculosis incidence.

16 Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J, et al. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. Lancet Infect Dis. 2012;12:45–55

17 Zak DE, Penn-Nicholson A, Scriba TJ, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet 2016; 387: 2312–22.

5. Target Product Profile: Test predicting progression from tuberculosis infection to active disease

An ideal test of progression would detect the presence of incipient TB. The test could possibly rely on identification of a mycobacterial product or host response marker that is identified in individuals further in the spectrum towards active TB. This may be particularly challenging as active TB is itself an eclectic disease largely dependent on host response e.g. primary pulmonary vs. disseminated or miliary TB.

In addition, for a test to have impact in high-burden settings, it would likely need to be repeated periodically to detect patients shortly after they have acquired an infection in order to prevent progression to disease. Therefore, the test should use an easily accessible sample and be suited for use in a primary or secondary healthcare facility by health care personnel with minimal training. The test should have higher positive predictive value for progression of infection to active TB than current tests and high negative predictive value for active TB, which may be mutually exclusive. Alternatively, a two-step process involving a highly sensitive screening test for infection followed if positive by a biomarker test to assess progression risk may be employed. A test with a lower positive predictive value may be acceptable in the setting of less-complex and less toxic regimens for the treatment of infection, but will still be sub-optimal owing to the risk of subjecting a low-risk individual to potential drug toxicity.

It may be challenging to develop an affordable test with all the above-mentioned characteristics. However, increasing use in developed nations and saving costs on treatment of infection/monitoring may help reduce test costs and costs to the health care system overall in the future.

Technical Expert Group Consensus for the Target Product profile

To facilitate consensus building for the development of a robust TPP, a Delphi-like methodology was adopted and involved two on-line surveys conducted by the NDWG to gather input from stakeholders to refine and improve the draft TPP and inform follow-up activities.

The first on-line survey was conducted in May 2016 and targeted the TB community at large. Ten of the 31 items in the TPP for a test of progression were selected for evaluation by survey participants, based on their scientific and implementation relevance. Participants in the survey included representatives of academia, multilateral and international agencies, NGOs, civil society and community representatives, endemic countries, and test developers, in addition to about 400 members of the NDWG.

A second survey was conducted in January 2017 and targeted more specifically participants invited to participate in the Technical Expert Group (TEG) consultation of 8 February 2017, with the aim of identifying areas of disagreement to help frame discussions during the TEG. Based on the responses from these two surveys four main areas were identified for further discussion during the TEG hosted by WHO on 8 February 2017. These were (1) the goal of test and/or the intended, (2) the target population, (3) performance characteristics and (4) instrumentation and a number of minor other discussion points.

Goal of the test

Several survey participants had noted that it would be unrealistic to expect that assays would be able to rule out active TB while at the same time predict progression from infection to disease. The same view was shared by the TEG consultation participants who agreed that the ability to rule out active disease should not be an optimal characteristic. The Technical Expert Consensus was that an optimal test would provide a quantitative result that correlates with the risk of progression and thus give an indication where on the spectrum of TB a patient may lie, which could aid in decisions about further workup and treatment.

Target population

There was discussion as to whether the target population should be broadened to go beyond individuals at increased baseline risk of infection or progression. However, the TEG consensus was that testing the general population in a low-risk setting would generate a high number of false-positive test results and would thus likely carry an unfavorable risk-benefit profile for individuals and be costly and inefficient for health systems. It was noted that an exception may be for individuals in settings with high levels of ongoing transmission where even individuals without typical risk factors could be considered as the target population.

Performance characteristics

The TEG participants agreed that setting performance targets is challenging. The consensus was that any performance targets needed to be balanced between aspirational and achievable targets. Ambitious targets that motivate further research and development to find the best possible solution need to be considered against what is realistically achievable with one-off testing to predict an event in the future. The group noted that repeat testing may enable improving both sensitivity and specificity.

Instrumentation

Several survey participants had noted that it would be unrealistic to expect an instrument-free solution. TEG consensus was that a robust and affordable point-of-care (POC) device would be optimal, while larger instrumentation suitable for centralized testing would meet the minimal criterion.

Other discussion points

It was noted that sputum should not be considered an optimal specimen type, due to the difficulty in obtaining sputum samples in particular from children and persons living with HIV. As a result this was removed as an optimal characteristic. It was noted that breath should be added as an option for an optimal specimen type. To enable inclusion of imaging-based solutions, the phrase “biomarker-based” was removed from the description. The minimal number of training days was reduced to 1-3, since participants agreed that this was sufficient even for more complex technologies. There was some discussion about the cost of instrumentation and assays but the TEG consensus proposed that the optimal requirement for cost of equipment should ideally be less than 500USD and as a minimum requirement the maximum price should not exceed 5000USD.

The consensus TPP is provided in Annex 1.

6. Framework for the evaluation for tests that predict progression from tuberculosis infection to active disease

Since 2008, WHO follows the GRADE process for evidence synthesis and evaluation when developing new guidelines and policy recommendation[18]. An evaluation framework has been developed and presents a standardised approach to generate performance data of an ITT. The purpose of the framework is to guide test manufacturers, researchers and research funders about the study designs that are required to generate evidence suitable for WHO evaluation and subsequent development of policy guidance. Prior to the evaluation of any new test in field, early analytical studies should be conducted to assess its reproducibility, robustness and variability under different conditions. The design standards and requirements for such early evaluations are outside the scope of this document.

As described above, an ITT will have the characteristics described in the Box 1 below.

To generate admissible evidence for a WHO evidence assessment of a novel ITT, two key research questions need to be addressed. Firstly, the predictive ability of the test should be assessed in clinical evaluation studies that include the intended target population, although individuals should not receive preventive therapy. These studies are intended to generate evidence solely on test performance in the absence of any additional intervention. Secondly, public health impact studies are necessary to evaluate the ITT under routine programmatic conditions and to assess the potential impact of the test on patient-important or health system-important outcomes. These studies should compare the programmatic results of a strategy where the new test is applied with the alternative that is currently in place, which can either be an alternative test-and-treat strategy (e.g. TST or IGRA testing) or no alternative test in settings or populations where LTBI testing is not (yet) being applied.

BOX. Characteristics of an incipient TB test

- To be negative in individuals never exposed to TB, including individuals who may be symptomatic for other (respiratory) illnesses but who have an alternative diagnosis.
- To be negative in individuals who are infected with MTB but who have no incipient TB. They might have a persistent TB infection, have a positive LTBI test (TST or IGRA) but do not develop TB disease within the next 2 years.
- To be negative in individuals who have been treated for LTBI.
- To be positive in individuals who develop TB within a short period after the test was done (e.g. 2 years), and who do not have any indication of re-exposure after the test was performed.
- To be positive in individuals with symptomatic TB disease.
- To be negative in individuals who completed TB treatment and are considered cured.

6.1 Clinical evaluation studies

Clinical evaluation studies should be used to determine the ability of the test to predict TB disease. Therefore, certain study designs used previously for the evaluation of IGRAs are non-informative in this respect, such as comparisons with IGRA or TST as the 'reference standard' or analyses of test results along a *Mycobacterium tuberculosis* exposure gradient.

18 Handbook for Guideline Development 2nd Ed. Geneva: World Health Organization; 2014 (http://www.who.int/publications/guidelines/handbook_2nd_ed.pdf?ua=1, accessed 1 June 2017)

To assess the predictive ability of the test, studies should evaluate the performance of the test in the intended target population, in settings where diagnostics such as culture or Xpert MTB/RIF® (Xpert) are available to confirm or exclude incident TB among those tested.

Research questions

The research questions to inform the predictive ability of an ITT are:

1. What is the accuracy (sensitivity and specificity) of the test to predict incident active TB within a pre-specified period?
2. What is the positive and negative predictive value of the test for incident active TB within a pre-specified period, and what are the corresponding number needed to screen to find a single positive test (NNS) and number needed to treat to prevent one incident TB case (NNT)?
3. What is the relative risk (RR) of a positive compared to a negative test for incident active TB within a pre-specified period?
4. What is the incident rate (IR) of TB after a positive and negative test, and what is the corresponding incidence rate ratio (IRR)?

Study design and population

Study designs should be longitudinal (prospective) studies in which a cohort of individuals at risk is tested at baseline and followed and evaluated for a specified duration (e.g. 2 years) for the occurrence of TB disease.

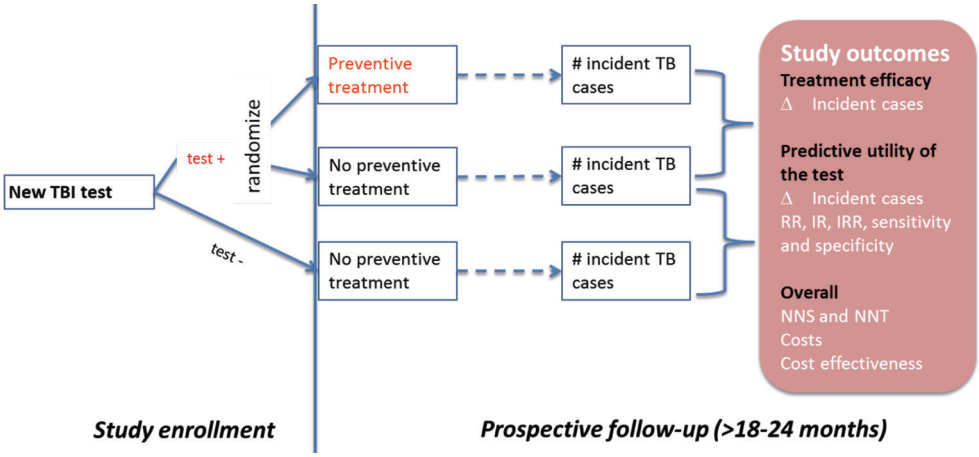
An alternative design is a case-control study nested within an existing cohort study. All individuals should be tested with the ITT at baseline. Incident TB cases should be captured through robust registries and a random subset of those who have not been registered with TB at the end of the study period should be contacted to confirm that they remained TB free. This study design is less costly because of its retrospective nature, but individuals are more easily lost to follow-up, leading to potential selection bias. For reasons of efficiency the study would ideally enrolls individuals with recent TB exposure (e.g. household contacts of infectious TB patients) or individuals with TB exposure (not necessarily recent) who are at a relatively high risk of progression to active TB disease but are currently not recommended for preventive therapy (PT) according to national guidelines[19]. Eventually, the test should be evaluated in a number of patient groups to ensure that performance is consistent in all risk groups and all disease presentations.

Clinical evaluation studies of an ITT pose a number of design challenges. Where TB incidence is low, it may be challenging to find and enroll sufficient numbers of eligible individuals with a history of recent exposure to infectious TB patients. Individuals who are eligible to receive preventive treatment (e.g. HIV-infected individuals and children) according to WHO guidelines cannot be included in the study without introducing ethical dilemmas and bias. Therefore, these studies should only enroll individuals not routinely recommended for preventive treatment. One option is to randomize individuals with a positive ITT who are according to national guidelines not recommended for PT, to either PT or placebo, as is done in the CORTIS study in South Africa. In this trial individuals with a positive RNA signature will be randomized to receive a course of 3 months isoniazid and rifapentine or no PT[20]. All individuals irrespective of their RNA signature will be followed-up, which allows determining its predictive ability for incident active TB (Figure 3).

19 Guidelines on the management of latent tuberculosis infection. Geneva: World Health Organization; 2015 (WHO/HTM/TB/2015.01 http://apps.who.int/iris/bitstream/10665/136471/1/9789241548908_eng.pdf?ua=1&ua=1, accessed 18 July 2017).

20 ClinicalTrials.gov. The Correlate of Risk Targeted Intervention Study (CORTIS). 2016 12-07-2017]; https://clinicaltrials.gov/ct2/show/NCT02735590?term=tuberculosis+cortis&rank=1, accessed 18 July 2017).

Figure 3. Example of study design for the clinical evaluation of a novel ITT test



Note: study design is based on the CORTIS trial (https://clinicaltrials.gov/ct2/show/NCT02735590?term=CORTIS+tuberculosis&rank=1).
Abbreviations: D=difference, IR=incidence rate, IRR=incidence rate ratio, NNS=number of individuals needed to screen to find a positive test, NNT=number of individuals needed to treat to prevent one incident TB case, RR=risk ratio, TBI=tuberculosis infection.

Enrolling HIV-negative adult contacts of infectious TB patients living in countries where they are not indicated for PT may pose ethical problems as this may be a reflection of resource constraints rather than of standard of care. A careful assessment and weighing of the potential benefits and harms of participating in research of ITTs, irrespective of existing country policies for LTBI testing and treatment, will therefore be essential. Moreover, in these studies one should avoid enrolling individuals who are at repeated risk of TB exposure, such as health care workers exposed to TB patients, since re-infection during the study period may bias the test accuracy estimates. On the other hand, study populations may include individuals who have other common bacterial or viral infections than TB.

Other challenges are the low disease progression rates. Even in subpopulations that carry an increased risk for breakdown to disease, the cumulative TB incidence usually do not exceed 5% over a period of 2 years[21,22]. Studies therefore require large sample sizes to ensure that sufficient events (i.e. incident TB cases) are observed during follow-up.

Finally, TB re-infections may occur during the study period after the ITT result was obtained. The rate of re-infection will be higher with higher TB incidence in the population in which the study is conducted. Re-infection may lead to misclassification bias in the accuracy estimates of the novel test depending on the re-infection rate and the length of follow-up. Since the re-infection rate may be modified by partial immunity due to existing LTBI and differ between those tested positive and those tested negative, the magnitude and direction of this bias (under- or overestimation of the predictive values of the ITT) will be difficult to predict. One way to minimize the risk of misclassification bias is to shorten the follow-up period in studies

21 Mathad, J.S., et al., Quantitative IFN-gamma and IL-2 Response Associated with Latent Tuberculosis Test Discordance in HIV-infected Pregnant Women. Am J Respir Crit Care Med, 2016. 193(12): p. 1421-8.
22 Rangaka, M.X., et al., Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. Lancet, 2014. 384(9944): p. 682-90.

conducted in high-incidence settings or repeat the ITT during the study period and assess its predictive ability for different lengths of follow-up.

Study methods

At study entry, prevalent symptomatic TB should be ruled out in accordance with the national guidelines for starting PT. The study should not attempt to rule out TB in a more rigorous way than is done in routine practice as this might exclude cases of asymptomatic, incipient TB from the study population that the novel ITT test is intended to identify.

Individuals enrolled in the study should be followed and all, irrespective of their initial test results, should be evaluated for the occurrence of active TB blinded to the initial test result, e.g. by a blinded clinical review panel. Follow-up should preferably be active to limit cohort attrition and the possibility of verification bias. However, passive follow-up for most of the study period with an active visit at the end of the study period may be acceptable in places where migration is limited and systems are in place for tracing study participants. For nested case-control studies, all cases should be captured through robust registries and controls and a random subset of those not registered should be contacted to confirm that they indeed remained TB-free. To prevent further misclassification bias ascertainment of the outcome (development of TB) should be done with a highly specific test (e.g. culture or Xpert assay).

Study analysis

The primary endpoint for the study is the cumulative incidence of TB among individuals with a positive baseline ITT compared to those with a negative ITT. Ideally, bacteriological confirmation (by culture, Xpert or more sensitive future alternatives) should be used to confirm incident TB in those with symptoms suggestive of TB. To rule out incident TB individuals should be free of symptoms suggestive of TB. Secondary analysis may be conducted using less stringent definitions for the diagnosis of an incident TB case.

The predictive ability of the test can be expressed in different ways. In addition to the test accuracy (sensitivity and specificity) the positive and negative predictive values for predicting incident TB cases, the risk ratio, the incident rate after a positive test and after a negative test and the incident rate ratio may also be determined. All these outcomes can be measured using the same study design. These outcomes may be monitored for the total follow-up period of the study (e.g. 2 years) as well as separately for different lengths of follow-up, such as the first 3 months, 6 months, 12 months etc. to assess whether the predictive ability decreases when time increases between sample collection and the moment that active TB developed. An example of such an analysis was conducted by Zak et al. in a prospective cohort study of adolescents in South Africa where blood samples were collected on a 3-monthly interval to assess if a RNA signature predicted progression to active TB in the following 2 years[23]. The predictive ability of the signature increased with decreasing time interval between sample collection and diagnosis of TB.

For tests that allow using different cut-offs for a positive result, trade-offs between sensitivity and specificity may be outlined, e.g. through a ROC-curve.

Important variables to record and stratify results for include the history of previous TB disease, age, gender, BCG vaccination status, risk of re-exposure (high/low incidence country) and comorbidities as listed in Table 1. Additional information on the TST and IGRA results of individuals allows for direct comparison of the new test with currently available LTBI tests and is therefore highly recommended, even though these tests should not be used as the reference standard. To inform policy, subgroups analysis or separate studies that include populations of special interest will be required, including but not limited to, children, people living with HIV, individuals with other forms of immunodeficiency (e.g. TNF-alpha inhibitors), diabetic patients and individuals with extra-pulmonary TB or a history of prior TB or LTBI treatment.

23 Zak, D.E., et al., A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet, 2016. 387(10035): p. 2312-22.

Table 1. List with minimum variables to measure in studies evaluating a TB prediction test

Minimum information needed for all groups	• Age • Gender • BCG-vaccination status • Country of residence • HIV status • Presence of other immune-deficiencies • Presence of other comorbidities • TST result (if possible) • IGRA results (if possible) • Date and time of sample collection (in particular needed, when multiple samples are collected from the same individual) • History of TB
Minimum information needed for incident TB cases	• Location of TB (PTB/EPTB) at time of incident TB • Method of TB detection (self-presented with symptoms or active case finding) at time of incident TB • Symptoms at time of incident TB
Subgroups of specific interest for sub-analysis	• Children • People living with HIV • Individuals with other forms of immunodeficiency • Diabetic patients • Individuals with malnutrition • Patients with incident extra-pulmonary TB • Patients with a history of prior TB treatment • Patients with a history of prior LTBI treatment • Individuals with and without risk of previous TB exposure/ re-exposure during study period (high/low incidence country)

6.2 Health impact studies

Health impact studies are those that aim to evaluate individual patient or health system important outcomes of an ITT. This second set of research questions are intended to provide information related to the potential impact when the test is used in routine practice. Studies that address these questions should be conducted in the settings of intended use, such as non-tertiary care hospitals or primary health care facilities. An important aspect of these studies is to assess the effectiveness and impact of the test when used to guide treatment decisions. Results of these studies may be used in subsequent modeling studies to further assess the potential public health impact of the test.

Research questions

The research questions to evaluate the health impact of an ITT include:

1. What is the effectiveness of the test for reducing incident TB when combined with a strategy to offer preventive treatment (PT) upon a positive test?
2. Is the test combined with PT a cost-effective strategy to reduce incident TB in individuals at high risk of recent TB exposure or high risk of progression to disease?
3. Is the test combined with PT a more effective and cost-effective strategy compared to alternative LTBI test-and-treat strategies using TST and/or IGRA?
4. What is the effect of the ITT combined with PT on the occurrence of adverse effects (e.g. hepatotoxicity), when compared to alternative LTBI test-and-treat strategies (e.g. based on TST and/or IGRA)?
5. What is the effect of the test combined with PT on the uptake and acceptance of PT?

Although in theory health impact studies could run in parallel with clinical evaluation studies, ethical review boards may require data from clinical evaluation studies that indicate that the novel test predicts incident TB equally well as current LTBI tests, such that equipoise can be assumed.

Study design and population

Study designs that allow for these research questions to be answered include comparative studies in which a test-and-treat strategy based on the novel test is compared with the current strategy (for instance TST and/or IGRA followed by PT) or, in settings where there is no alternative strategy in place, no testing. Preferably these should be individually or group-randomized trials. Alternative study designs may include stepped-wedge trials, although these may have limitations with regard to their interpretation[24].

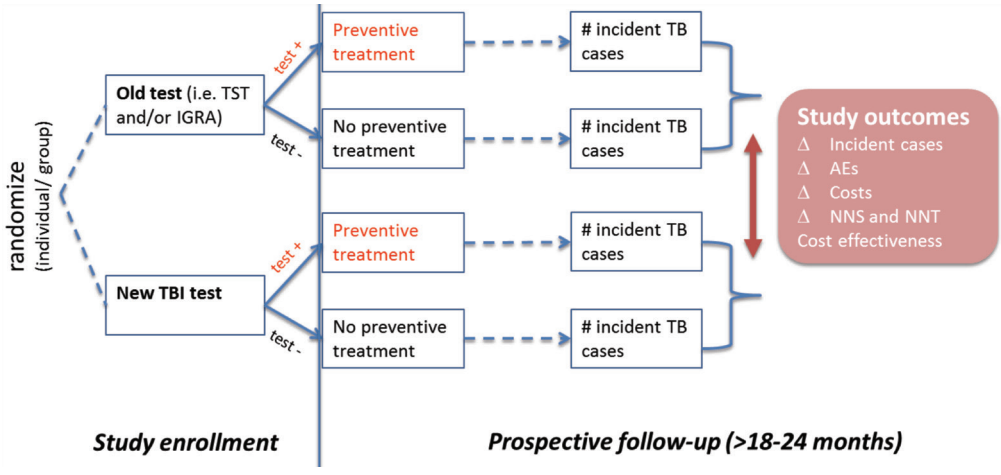
An example of a study design for a pragmatic randomized-controlled trial is given in Figure 4. Individuals or clusters are randomly assigned to receive either the standard of care (in this example a testing strategy based on TST and/or IGRA) or the new testing strategy. Individuals in both arms are offered PT when their test is positive. All individuals, irrespective of their test results, are followed up for the occurrence of incident TB. At the end of the study the difference in the number of incident TB cases, number of patients given PT, number of patients lost-to follow-up, frequency of adverse events and patient and health system costs are compared between both trial arms.

In order to inform WHO guideline development, study populations should include the intended (future) target population for the test, as described in the TPP. Studies should be conducted in low- as well as high-incidence countries and could in principle be of similar design. As studies evaluating the public health impact of a novel test will include preventive treatment of those tested positive there is less concern about the possibility of (indication) bias, as well as less ethical concern, than for studies assessing the predictive ability of the test.

Study methods

All individuals enrolled in the study should be followed for the same pre-specified period, irrespective of their test result and irrespective of whether they receive PT or not. Follow-up should ideally extend until two years after the completion of PT to assess the occurrence of incident TB cases after treatment completion. The whole study population should be assessed

Figure 4. Example study design for the evaluation of public health impact



Abbreviations: AEs=adverse events, D=difference, NNS=number of individuals needed to screen to find one positive test, NNT=number of individuals needed to treat to prevent one incident TB case, TBI=tuberculosis infection.

24 Hemming, K., et al., The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ, 2015. 350: p. h391

for the occurrence of TB disease at least at the end of the study period, but if possible also at interim time points during the study. Ideally, the outcome assessment should be blinded to the initial test result to avoid differential verification or incorporation bias. A form of active follow-up is preferred above passive follow-up, in particular to limit potential loss to follow-up (cohort attrition), which is otherwise more likely to happen in the group that does not receive PT since they do not need to return for follow-up visits. Ascertainment of incident TB may be done according to routine practice, i.e. following nationally recommended steps for diagnosing active TB.

Study analysis

In the analysis, the outcomes (e.g. incidence of TB disease, costs, occurrence of side effects) in the group that received the novel test-and-treat strategy should be compared with those in the alternative arm. The primary analysis should be based on the intention-to-treat cohort, which includes all patients who were enrolled in the arm they were randomly allocated to, irrespective of whether they adhered to all interventions in their assigned arm.

The minimum list of variables to be collected for studies of predictive ability is presented in Table 1. In addition, data should be captured on the acceptance of the novel test, acceptance of PT upon a positive test result, adverse events, cost of the complete test-and-treat intervention as well as of the alternative strategy. Besides a direct comparison on the effectiveness of the test-and-treat strategy, the study may also report on the cost, cost-effectiveness and occurrence of side effects. All these outcomes together would inform the positive and negative implications of scaling up the novel test-and-treat strategy and its potential budget implications.

6.3 Summary of clinical and health impact studies

Studies that address both sets of research questions needed to be done to inform the WHO policy guidance process for the use of a novel ITT. This framework is intended to be used by test developers, manufacturers and others who plan to evaluate ITT candidates to design appropriate studies and make sure that the appropriate outcomes are being recorded. Often diagnostic studies do not report the same outcome measures (i.e. risk ratio's using cumulative incidences vs incidence rates based on person years of follow-up), even though these could easily be distilled, or do not include the appropriate study population and are therefore excluded from the evidence synthesis that informs WHO approval and policy guidance.

Although comparative studies, in particular if randomized, provide the highest quality of evidence, they carry high cost. A way to minimize costs of clinical evaluation studies is to design the study such that multiple research questions can be answered using the same study design. Several examples have been described known[25]. Another option would be to make use of stored specimens (sample banks) that were collected in longitudinal studies and retrospectively analyse the test performance in a nested-case control study design[26,27].

For health impact studies, an alternative to actual studies is to *model* the potential impact of the test-and-treat intervention under different circumstances. While such studies might be cheaper and generate results faster, they bring other challenges. In addition to the research questions outlined earlier, other analyses may be worthwhile to further explore using the data from modelling studies, e.g. to 1) look in more detail at the predictive utility of different cut-off levels of the test for different subgroups, 2) explore if the predictive ability of the test improves when combined with other patient characteristics and 3) model the long term public health impact for varying cut-offs or prediction models in combination with different PT regimens.

25 ClinicalTrials.gov. The Correlate of Risk Targeted Intervention Study (CORTIS). 2016 12-07-2017]; https://clinicaltrials.gov/ct2/show/NCT02735590?term=tuberculosis+cortis&rank=1, accessed 18 July 2017)
26 Rangaka, M.X., et al., Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. Lancet, 2014. 384(9944): p. 682-90.
27 Mahomed, H., et al., The tuberculin skin test versus QuantiFERON TB Gold(R) in predicting tuberculosis disease in an adolescent cohort study in South Africa. PLoS One, 2011. 6(3): p. e17984.

Annex 1: Target Product Profile: Test predicting progression from tuberculosis infection to active disease

Definitions

TB infection: Any person with a positive test for TB infection (TST≥5mm, positive IGRA according to manufacturer's instructions) without microbiological, radiological, or clinical evidence of active TB.

Incipient TB disease: Individuals with tuberculosis infection in whom progression to TB disease has started and who have no symptoms, no radiographic abnormalities suggestive of TB and negative microbiological investigations. Individuals with incipient disease are very likely to develop active TB within a short time of initial evaluation. A subset of patients with incipient disease (primarily immunocompetent patients) will not progress to active disease.

TB disease: Symptomatic patients with compatible clinical and/or radiology and/or histology for TB and a positive microbiological test (confirmed TB), or with compatible clinical and/or radiology and/or histology for TB and started TB treatment (clinical TB).

Characteristic	Optimal	Minimal	Explanations/ Limitations
Intended Use			
Goal of test/ Intended use	Test that can be used to predict risk of progression to active TB from TB infection (TBI) within the next 2 years and provides a quantitative result that correlates with the risk of progression. The test result should decrease or revert to negative with treatment and thus allow an assessment of treatment success or cure and consequentially also reinfection.	Test that can be used to predict risk of progression to active TB from TB infection within the next 2 years. As this test may also be positive in patients with active TB, identification of these individuals needs to be done by a highly sensitive test	TST and IGRAs currently are the mainstay for the diagnosis of TB infection. However, these tests do not predict which individuals are likely to progress to active TB. Progression of TBI may involve varying immuno-pathogenic processes depending on stage (incipient or clinical), type and site of disease (e.g. pulmonary vs. miliary TB). Ideally this test would also have the ability to rule out active TB and a graded test with different cut-offs for incipient and active TB may be useful to guide the choice between different regimens needed depending on the extent of disease. However, it may not be possible to achieve this if we consider that incipient disease is part of the spectrum of active TB. In that case, ruling-out active TB in test-positive patients may need to be done based on other tests/ information (e.g. symptoms, Chest X-Ray, culture). A quantitative test result that correlates with risk of progression could facilitate treatment decisions and a "high" signal could trigger further evaluation for and possible subsequent treatment of active disease before a preventive treatment is given. Algorithm based interpretation will also likely play a part as future changes in immune function may not be available at baseline.
Type of specimen	Capillary whole blood (finger prick sample) / saliva / urine / stool / breath	Whole blood by phlebotomy (or subpopulation of cells if simple processing included) / sputum	

Characteristic	Optimal	Minimal	Explanations/ Limitations
Target population	Asymptomatic individuals who have increased likelihood of exposure to a person with active TB (e.g. close contacts) and individuals with conditions that predispose to progression of TBI to active disease (HIV infection, diabetes, chronic renal failure, chronic medical illness, recent TST converters, children < 5 years of age, and persons receiving anti-TNF).		Testing the general population in a low-risk setting would generate a high number of false-positive test results and would thus likely carry an unfavorable risk-benefit profile for individuals and be costly and inefficient for health systems. This may be different in the general population in settings with high levels of ongoing transmission or with performance characteristics exceeding those set forth in this TPP. Accordingly, programmatic decisions about target populations may vary but need to take these considerations into account.
Target user of the test	Health care workers with minimal laboratory training e.g. nurses	Health care workers with laboratory training e.g. skilled laboratory technicians	
Setting (lowest level of implementation in health care system)	Health post	Referral facilities with some laboratory facilities	
Performance characteristics			
Diagnostic sensitivity for progression to active TB	≥ 90% sensitivity	≥ 75% sensitivity	The performance characteristics are with respect to a two year time horizon over which occurrence of the outcome (progression to active TB) would be observed. A detailed description of rationale for the chosen targets and its measurement is provided in the note below the table. Note that —as per the description below— some deviation from these targets is seen as acceptable. Ideally, the test should perform equally well in all risk groups and all disease presentations. The test should be unaffected by BCG vaccination status and NTM infections. Repeat testing may lead to improved sensitivity and specificity.
Diagnostic specificity for risk of progression to active TB	≥ 90% specificity	≥ 75% specificity	

Characteristic	Optimal	Minimal	Explanations/ Limitations
Reproducibility	Reproducibility: Inter-assay CV ≤ 10.0% at high and low extremes of the assay		For assays that provide a quantitative output (e.g. limit of detection, CT-values)
Operational characteristics			
No. of steps to be performed by operator	< 2, no timed steps	< 10, 1-2 timed steps	
Volume measurements	None	Measuring device provided with kit	
Sample preparation	None or fully integrated	Allows for centrifugation/ incubation	
Data analysis	Integrated	Integrated	
Time to results	< 24 hours	2-5 days	
Biosafety	Universal precautions	Universal precautions, biosafety Level II	
Operating Temperature	Between 5 and 50 °C, 90% humidity	Between 5 and 30 °C, 70% humidity	
Reagents	Self-contained within test kit	Up to 2 external reagent, reconstitution not required	
Stability of test kit / reagent	24 months at 40 °C, 90% humidity, should be able to tolerate stress during transport (3 days at 50 °C)	12 months at 30 °C, 70% humidity, cold chain required for transport	
Instrumentation	Preferably instrument free. If instrument: Small, portable or hand-held instrument (<1kg) that can operate on battery or solar in places with interrupted power supply	Centralized testing platform suitable for use in laboratories.	Note that in some settings centralized, high-throughput instruments may be preferable to small, low-throughput instruments. However, given the cascade of care for identifying and treating TB infection (and the high rate of losses to follow-up), instruments that can be deployed at the lowest level of the health care system have important advantages in many settings.
Waste disposal	Standard infected waste disposal at health center		

Characteristic	Optimal	Minimal	Explanations/ Limitations
Internal Quality control	Included positive control		
External Quality control	Included positive and negative controls	Included positive and negative controls	
Maintenance/ calibration	No calibration/maintenance required	Annual calibration by company staff; maintenance every 1,000 tests or 12 months	
Power requirements	Ideally instrument free test; all equipment with rechargeable battery lasting up to 8 hours	110-220 V AC current; UPS for power failures	
Result capturing, documentation, data display	Ideally instrument free test, but should allow for attaching, or scanning result to the reader to have the ability to save and print the results	Ability to save the results either via instrument or via a separate reader (or alternative). When instrument is used the test menu should be simple with integrated LCD screen; simple key pad or touch screen	
Data export (connectivity and interoperability)	Preferably instrument free but test should allow data export via reader Full data export (on usage of device, error/invalid rates, and personalized, protected results data) over USB port and network. Network connectivity through GSM/UMTS mobile broadband modem. Results should be encoded using a documented standard (such as HL7) and be formatted as JSON text. JSON data should be transmitted through HTTP(S) to a local or remote server as results are generated. Results should be locally stored and queued during network interruptions and sent as a batch when connectivity is restored	Full data export (on usage of device, error/invalid rates, and personalized, privacy-protected results/ data) over USB port and network. Network connectivity through Ethernet, Wi-Fi, and/or GSM/UMTS mobile broadband modem. Results should be encoded using a documented standard (such as HL7) and be formatted as JSON text. JSON data should be transmitted through HTTP(S) to a local or remote server as results are generated. Results should be locally stored and queued during network interruptions and sent as a batch when connectivity is restored	

Characteristic	Optimal	Minimal	Explanations/ Limitations
Electronics and software	None	Integrated	
Training	< 1 day dedicated training for non-laboratory trained health personnel	1-3 days dedicated training for a laboratory trained health personnel	
Pricing			
Cost of equipment	< 500 USD	< 5000 USD	
Cost of consumables (reagents/ test strips)	< 5 USD/ test	10-100 USD/test	As an initial step, it may be acceptable to have an assay costing as much as the currently available IGRAs. Making the test affordable will be an important next step and lower cost will be essential for uptake in lower-income countries.

Global TB Programme
World Health Organization
Avenue Appia 20, CH-1211
Geneva-27, Switzerland

Information Resource Centre
HTM/GTB:
Email: tbdocs@who.int
Website: www.who.int/tb