An evaluation framework for new tests that predict progression from tuberculosis infection to clinical disease

Sandra V. Kik^{1,2}, Samuel Schumacher³, Daniela Maria Cirillo⁴, Gavin Churchyard^{5,6,7}, Catharina Boehme³, Delia Goletti⁸, Molebogeng X. Rangaka^{9,10}, Claudia M. Denkinger³, Christian Lienhardt^{11,12}, Christopher Gilpin¹¹, Alberto Matteelli¹³ and Frank Cobelens^{1,2}

Affiliations: ¹KNCV Tuberculosis Foundation, The Hague, The Netherlands. ²Amsterdam Institute for Global Health and Development and Dept of Global Health, Academic Medical Center, Amsterdam, The Netherlands. ³FIND, Geneva, Switzerland. ⁴IRCCS San Raffaele Scientific Institute, Milan, Italy. ⁵Aurum Institute, Parktown, South Africa. ⁶School of Public Health, University of Witwatersrand, Johannesburg, South Africa. ⁷Advancing Care and Treatment for TB/HIV, South African Medical Research Council, Johannesburg, South Africa. ⁸Translational Research Unit, Dept of Epidemiology and Preclinical Research, National Institute for Infectious Diseases, Rome, Italy. ⁹Institute for Global Health, University College London, London, UK. ¹⁰CIDRI-AFRICA, University of Cape Town, Cape Town, South Africa. ¹¹Global TB Programme, World Health Organization, Geneva, Switzerland. ¹²Unité Mixte Internationale TransVIHMI (UMI 233 IRD – U1175 INSERM - Université de Montpellier), Institut de Recherche pour le Développement (IRD), Montpellier, France. ¹³WHO Collaborating Centre for TB/HIV Co-infection and TB Elimination, Dept of Infectious and Tropical Diseases, University of Brescia, Brescia, Italy.

Correspondence: Sandra V. Kik, KNCV Tuberculosis Foundation, PO Box 146, 2501 CC The Hague, The Netherlands. E-mail: sandra.kik@kncvtbc.org

@ERSpublications

An evaluation framework to evaluate new tests that predict development of clinical tuberculosis
<https://ow.ly/zVZH30kZEsU>

Cite this article as: Kik SV, Schumacher S, Cirillo DM, *et al.* An evaluation framework for new tests that predict progression from tuberculosis infection to clinical disease. *Eur Respir J* 2018; 52: 1800946 [<https://doi.org/10.1183/13993003.00946-2018>].

ABSTRACT Novel accurate tests are needed that identify individuals infected with *Mycobacterium tuberculosis* who have incipient disease and are likely to develop clinical tuberculosis (TB) in the near future to allow for targeted preventive treatment beyond the current risk groups. Recently, a target product profile was developed that outlines the minimal and optimal characteristics for such an incipient TB test. We describe an evaluation framework for generating evidence to inform the development of policy guidance for the use of such a new test by the World Health Organization. Two research objectives are addressed. 1) The predictive ability of an incipient TB test should be assessed in clinical evaluation studies that include the intended target population and follow-up of sufficient duration to observe whether individuals do or do not progress to clinical TB disease. 2) Studies are needed to evaluate the test under routine programmatic conditions and measure its impact on patient- or health-system-important outcomes. For both research objectives, study designs, methods and analysis are described, with the intent to inform the clinical development plans of test manufacturers, researchers and funders.

This article has supplementary material available from erj.ersjournals.com

Received: March 09 2018 | Accepted after revision: July 12 2018

The content of this work is copyright of the authors or their employers. Design and branding are copyright ©ERS 2018.

Introduction

Programmatic management of latent tuberculosis (TB) infection (LTBI) has been poor in many high-burden countries, but if scaled up together with other treatment and prevention strategies it could accelerate the progress towards TB elimination [1–3]. LTBI is currently identified through a response to *in vivo* or *in vitro* immune stimulation by *Mycobacterium tuberculosis* antigens as shown by a positive tuberculin skin test (TST) or interferon- γ release assay (IGRA) in the absence of clinical signs of TB [4]. These tests, however, poorly predict which individuals progress to clinical disease, with a positive predictive value for developing clinical TB over a 2-year period ranging between 1.5% and 6.8%, depending on the risk group and setting [5, 6]. In general, the positive predictive value is lower for populations with a high likelihood of *M. tuberculosis* infection acquired in the past rather than recently, hence the number needed to treat to prevent one TB case is higher. Targeted testing and treatment is therefore only recommended for individuals with a high probability of a recent *M. tuberculosis* infection or with increased risk of progression to TB disease [7, 8].

New tests are needed to identify with high accuracy individuals who are infected with *M. tuberculosis* and are likely to develop disease in the near future. Several promising correlates of risk have been described, although only a few of them have been validated thus far [9–11]. Recently, COBELENS *et al.* [12] proposed to distinguish two conceptually different tests for 1) detecting persistent infection and 2) identifying incipient TB. It was argued that incipient TB tests detect a phase of early disease during which pathology evolves due to active mycobacterial replication and an associated inflammatory response, in the absence of clinical signs of disease (table 1). Such tests are expected to have a higher specificity and positive predictive value for predicting progression to clinical TB disease than current LTBI tests. Moreover, their performance characteristics should be largely independent of the population studied. The tests are considered “rule-in” tests: a negative result provides limited information (because of the limited sensitivity), but a positive result indicates that symptomatic (*i.e.* clinical, overt) clinical TB is likely to develop and therefore preventive treatment is indicated. Here, we outline an evaluation framework for new tests of incipient TB to inform assay manufacturers, researchers and funders about appropriate study designs to guide their clinical development plans and to generate evidence to inform the development of subsequent policy guidance by the World Health Organization (WHO). The process that was followed to reach consensus on the principles described here is depicted in table 2.

Target product profile

The minimum specifications of an incipient TB test are described in a target product profile (TPP) [13]. This TPP was initiated by the Foundation for Innovative New Diagnostics, jointly developed with the Stop TB Partnership’s New Diagnostic Working Group and WHO, and discussed in three expert group meetings (May 2015, July 2016 and February 2017). In short, such a test should be able to identify asymptomatic individuals with incipient TB who are likely to progress to clinical disease within the subsequent 2 years and who would therefore benefit from preventive treatment. A period of 2 years was chosen since progression to disease is highest shortly after infection; about half of the individuals who develop TB will do so within the first 2 years after infection [14]. In addition, the performance of a recently identified “correlate of risk” assay based on a 16-gene mRNA blood signature showed that prediction was greater nearer to the time of TB diagnosis and declined substantially after 2 years [15]. Our expectations of a test that measures incipient TB compared with tests that measure a LTBI, persistent infection or clinical TB are outlined in more detail in the supplementary material.

TABLE 1 Definitions in tuberculosis (TB)

Latent TB infection	Asymptomatic state of infection with <i>Mycobacterium tuberculosis</i> demonstrated by a response to <i>in vivo</i> or <i>in vitro</i> stimulation by <i>M. tuberculosis</i> antigens as identified by a positive TST or IGRA. Note that these assays cannot distinguish individuals who have cleared their <i>M. tuberculosis</i> infection from those that remain infected.
Persistent TB	Asymptomatic, stable quiescent state of infection with <i>M. tuberculosis</i> and with a persistent response upon <i>in vivo</i> or <i>in vitro</i> antigenic stimulation by <i>M. tuberculosis</i> antigens as identified by a positive TST or IGRA.
Incipient TB	Asymptomatic, early pre-clinical disease state during which pathology evolves, such as mycobacterial replication or the inflammatory response. During this state clinical signs of the disease, including TB-compatible symptoms, are still absent. Radiological abnormalities or positive microbiological tests may or may not be present. This state may either evolve and lead to symptomatic clinical TB or regress and remain asymptomatic.
Clinical TB	Active symptomatic disease state as a result of the mycobacterial inflammation, with presence of TB-compatible signs and symptoms, which can be demonstrated by radiological, pathological and/or microbiological evidence.

TST: tuberculin skin test; IGRA: interferon- γ release assay.

TABLE 2 Consensus reaching process

May 2015	WHO convened expert consultation meeting on behalf of NDWG and FIND	First meeting set up to identify the operational and performance characteristics of tests that could predict progression from LTBI to active TB. Following this meeting, two objectives were identified: 1) to develop a TPP for a test of progression and 2) to 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 WHO evaluation.
January 2016	NDWG Task Force on LTBI was established to work on these two objectives	
July 2016	NDWG Task Force meeting with invited experts	Draft versions of the TPP and framework document were presented and discussed. New definitions for different states of LTBI were proposed.
December 2016	Publication of new definitions for different states of LTBI	In this publication new concepts of LTBI are discussed and definitions for incipient and persistent LTBI are provided [12].
December 2016	Online consultation of experts	Pre-final drafts of the TPP and framework document were shared with over 60 international experts from low, middle and high TB incidence countries. These experts were requested to complete an online survey to judge if they agreed with specific items outlined in each of these documents.
February 2017	WHO convened expert consultation meeting on behalf of NDWG	Pre-final TPP and framework document were discussed. Results of the online consultation were presented. Items on which 75% of the experts agreed, or with many comments, were discussed to reach consensus on.
November 2017	Consensus meeting report published	This report outlines the consensus reaching process, and includes the final TPP and guidance recommendations [13].

WHO: World Health Organization; NDWG: New Diagnostic Working Group; FIND: Foundation for Innovative New Diagnostics; LTBI: latent tuberculosis infection; TB: tuberculosis; TPP: target product profile.

WHO evaluation and subsequent development of policy guidance

Since 2008, WHO follows the GRADE (Grading of Recommendations Assessment, Development and Evaluation) process for evidence synthesis and evaluation when developing new guidelines and policy recommendations [16, 17]. The evaluation framework presented here is meant to set a standard to generate the type of evidence that would be acceptable to support the GRADE process (“admissible evidence”) for WHO evaluation of an incipient TB test. Before entering into field evaluation studies, the test will have been scrutinised in earlier stages of development, to assess its reproducibility, robustness and variability under different circumstances. These early analytical studies, combined with data on test accuracy, are usually the focus for regulatory bodies to provide their approval. Requirements to obtain such approval are set out elsewhere [18–20]. However, to inform national and international guideline and policy development, evidence is required that looks beyond test accuracy, and demonstrates the effectiveness, cost-effectiveness and public health impact that can be anticipated when the product is applied in the clinical setting of intended use. Here, we describe the design standards for this last kind of study.

To generate such evidence, clinical evaluation studies are needed to assess the predictive ability of an incipient TB test. These studies are conducted in the intended target population and evaluate test performance in the absence of any additional intervention. To support a WHO policy guidance process, data generated from such studies should further be complemented by cost and transmission modelling in order to anticipate the patient and health system benefits. Health impact studies are needed to assess the intervention and measure its impact on patient- or health-system-important outcomes. Such studies are best done under routine programmatic conditions. We present here the design characteristics of the two types of studies.

Clinical evaluation studies: studies that aim to measure the test’s predictive ability

WHO evaluation will be based on the ability of the test to predict the occurrence of TB disease. This requires longitudinal studies that follow a cohort of individuals over time, conducting the test in the intended target population, in settings where culture, Xpert MTB/RIF (Xpert) or Xpert Ultra are available to confirm or exclude incident clinical TB in the study population.

Study design and population

Studies assessing this question should be longitudinal (prospective), following a cohort of tested individuals at risk of TB progression (e.g. household contacts of smear-positive TB patients or HIV-infected individuals) and evaluating them over 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, where samples from all individuals at risk of TB progression are taken at baseline and incident TB cases are captured either through active or passive follow-up. At study close-out samples from individuals diagnosed with TB and a random subset of individuals who remained TB-free (controls) are then tested with the novel tests. A random subset of individuals who were not diagnosed with TB at study close-out should be contacted to confirm that they remained TB-free. This design is less costly because fewer tests will be done. In these studies, as in others, active follow-up of all study participants is preferred to reduce the chance of loss to follow-up and potential selection bias.

For reasons of efficiency, clinical evaluation studies, regardless of their design, should ideally enrol individuals with likely *M. tuberculosis* exposure, either recent or not, but with a high risk of disease progression. However, a careful assessment and weighing of the potential benefits (e.g. prevention of clinical TB) and harms (e.g. depriving individuals from preventive therapy when potentially beneficial for them) of participating in research of novel incipient TB tests is essential, to make sure that 1) the country's LTBI policies are followed, and 2) for individuals that are not mentioned in the country policy equipoise can be assumed for the intervention and control arm.

Study methods

At study entry, prevalent symptomatic TB should be ruled out in accordance with current guidelines for starting preventive treatment. Studies should not attempt to rule out TB in a more rigorous way than is done routinely, e.g. by doing chest radiography or Xpert in individuals who do not have TB symptoms, as this might exclude cases of asymptomatic, incipient TB from the study population that the novel test is intended to identify.

Enrolled individuals should be tested with the novel test at least at baseline. Individuals are to be followed over time, irrespective of their initial test results, and evaluated for the occurrence of clinical TB at regular intervals (e.g. 3 or 6 monthly). The novel test may be repeated at the time-points when the end-point is being assessed to allow for evaluation of re-infection, disease progression or regression. The diagnosis of incident TB disease should be made blind of the initial test result. Follow-up is preferably active to make sure that individuals are followed with the same rigour regardless of the result of their incipient TB test and to limit the risk of loss to follow-up. Passive follow-up (for most of the study period) with a scheduled visit at the end of the study period may be acceptable in places where migration is limited and systems are in place for tracing back study participants. For nested case-control studies, all cases should be captured through robust registries and controls should be contacted to confirm that they remained TB-free. To prevent misclassification bias, ascertainment of incident TB disease should be with a highly specific test (e.g. culture or Xpert Ultra).

Data analysis

The primary study end-point is the cumulative incidence of confirmed bacteriologically clinical TB among individuals with a positive or negative incipient TB test at baseline. Secondary analyses may use other definitions for diagnosing and ruling out incident TB disease.

The predictive ability of the test can be expressed in different ways. We suggest to report: 1) the test accuracy (sensitivity and specificity for predicting incident clinically confirmed TB); 2) its positive and negative predictive value for clinical TB within a pre-specified period, the corresponding number needed to screen to find one positive test, and the number needed to treat to prevent one incident TB case; 3) the relative risk of a positive compared with a negative test for clinical TB within a pre-specified period; and 4) the incident rate of clinical TB after a positive and negative test, and corresponding incidence rate ratio.

All suggested outcomes can be measured in the same study. Outcomes may be shown for the total duration of follow-up (e.g. 2 years) as well as for shorter durations (i.e. 3, 6 or 12 months) to demonstrate possible decreases in predictive ability with increasing time between sample collection and onset of TB disease. Examples of such analyses were conducted by ZAK *et al.* [15] who assessed the predictive ability of a RNA signature for incident clinical TB disease in a prospective cohort study of adolescents in South Africa, as well as others who assessed the relation between the presence of the correlate of risk and time till onset of TB [10, 11, 21]. For tests with a quantitative readout, different cut-offs for a positive result and trade-offs between sensitivity and specificity may be explored and presented, e.g. through a receiver operating characteristic curve.

Where feasible, studies should record and show stratified results for a range of different variables, including the history of previous TB disease, age, sex, bacille Calmette-Guérin vaccination status, risk of re-exposure (high/low-incidence country) and comorbidities (table 3). Information on individual TST and IGRA results would allow direct comparison of new tests with these existing LTBI tests and is highly

recommended, even though the TST or IGRA should not be used as the reference standard. Dependent on the biological mechanism the novel test is measuring, the TST may be considered to be done only after the incident TB test was performed in order to avoid a boosting phenomenon. To inform policy, subgroups analyses 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 types of immunodeficiency (e.g. under tumour necrosis factor- α inhibitor therapy), diabetic patients and individuals with extrapulmonary TB or a history of prior TB or LTBI treatment.

Challenges

Clinical evaluation studies of an incipient TB test pose a number of design challenges. In areas where TB incidence is low, one might not find sufficient eligible individuals with a history of (recent) *M. tuberculosis* exposure to enrol. Individuals who will receive preventive treatment according to guidelines cannot be included in the study without introducing bias, because preventive treatment reduces the risk of progression to clinical disease and individuals that decline preventive treatment may not be representative for all individuals who were at risk initially. Therefore, these studies should solely include individuals who are currently not routinely recommended for preventive treatment, such as individuals assigned to a nonintervention arm in randomised controlled trials (e.g. post-exposure vaccination trials). Another possibility is to randomise to either preventive treatment or placebo HIV-uninfected individuals with a positive test who are according to national guidelines not recommended for preventive treatment, as is done in the Correlate of Risk Targeted Intervention Study (CORTIS) in South Africa. In this trial, HIV-uninfected individuals with a positive RNA signature are randomised to receive a course of 3 months of weekly high-dose isoniazid and rifapentine (“3HP”) or no preventive treatment (ClinicalTrials.gov identifier NCT02735590) [22]. As all individuals with a positive RNA signature and a sample of those with a negative signature are followed, the predictive ability of the RNA signature for incident TB disease can be determined (figure 1).

Another challenge is the low disease progression rate. Even in populations that carry an increased risk for reactivation, the cumulative TB incidence usually does not exceed 5% over a period of 2 years [23–25]. Studies therefore require large sample sizes to ensure that sufficient numbers of incident TB cases are observed during follow-up. For example, the CORTIS trial will be screening 10 000 HIV-uninfected

TABLE 3 List with minimum variables to measure in studies evaluating an incipient tuberculosis (TB) test

Minimum information needed for all groups	Age Sex Bacille Calmette–Guérin vaccination status Country of residence HIV status Presence of other immune deficiencies Presence of other comorbidities Tuberculin skin test result (if possible) Interferon- γ release assay 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 clinical TB cases	Location of TB (pulmonary TB/extrapulmonary TB) 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 subanalysis	Children People living with HIV Individuals with other forms of immunodeficiency Close contacts Diabetic patients Individuals with malnutrition Patients with incident extrapulmonary TB Patients with a history of prior TB treatment Patients with a history of prior latent TB infection treatment Individuals with and without risk of previous TB exposure/re-exposure during study period (high/low-incidence country)

Reproduced from [13] with permission.

individuals living in a highly endemic community in order to enrol 1500 test-positive and 1700 test-negative individuals [22].

Finally, TB re-infections may occur during the study period after the test was conducted. 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 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 test) is difficult to predict. Potential misclassification bias can be minimised by shortening the follow-up period (especially in studies conducted in high-incidence settings), by repeating the test every 3 or 6 months during the study period and assessing its predictive ability for different lengths of follow-up, and by avoiding enrolment of individuals who are at repeated risk of TB exposure, such as healthcare workers exposed to TB patients (especially in studies conducted in low-incidence settings).

Health impact studies: studies to evaluate patient- or health-system-important outcomes

The second research objective to inform WHO evaluation addresses the potential impact of an incipient TB test on patient- or health-system-important outcomes when used in routine practice. Studies addressing this objective are conducted in settings of intended use, such as nontertiary care hospitals, DOTS (directly observed treatment, short-course) centres or primary healthcare facilities. Importantly, these studies assess the effectiveness and impact of the test when used to guide treatment decisions. Results of these studies may be used in subsequent impact and cost-effective modelling studies, which further assess the potential public health impact of the test.

Research questions related to the public health impact of an incipient TB test include the following. 1) What is the effectiveness of the test for reducing incident TB when combined with a strategy offering preventive treatment upon a positive test? 2) Is the test, combined with preventive treatment, a cost-effective strategy to reduce clinical TB in the intended target population? 3) Is the test, combined with preventive treatment, a more effective and cost-effective strategy compared with the current standard (e.g. alternative LTBI test-and-treat strategies)? 4) What is the effect of the test combined with preventive treatment on the occurrence of adverse effects (e.g. hepatotoxicity) when compared with strategies based on the TST and/or IGRA? 5) What is the effect of the test combined with preventive treatment on the uptake and adherence to therapy?

While health impact studies could run in parallel with clinical evaluation studies, some ethical review boards may require data from clinical evaluation studies indicating that the novel test predicts incident TB disease equally well as, or better than, current LTBI tests, such that equipoise can be assumed.

Study design and population

Studies aiming to answer the aforementioned research questions may compare a test-and-treat strategy using the novel test with the current standard and be individually or group randomised [22, 24, 26]. The

FIGURE 1 Example of study design for the clinical evaluation of a novel incipient tuberculosis (TB) test. Study design is based on the Correlate of Risk Targeted Intervention Study (CORTIS) trial (ClinicalTrials.gov identifier NCT02735590) [22]. RR: risk ratio; 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. Reproduced and modified from [13] with permission.

standard may be TST and/or IGRA testing followed by preventive treatment, or no LTBI testing in settings where this is not routinely done. Alternative study designs may include stepped-wedge trials, although these may have limitations with regard to their interpretation [27].

An example study design for a pragmatic randomised controlled trial is given in figure 2. Individuals or clusters are randomly assigned to receiving either the standard of care (in this example the TST and/or IGRA) or the novel test. Individuals in both arms are offered preventive treatment when their test is positive. All individuals, irrespective of their test results, are followed for the occurrence of incident TB disease. At study close-out the difference in the cumulative incident TB cases, number of patients provided preventive treatment, frequency of adverse events, and patient and health system costs are compared between trial arms.

To inform international guideline development, study populations should include the intended (future) target population for the test, as described in the TPP. Studies may be conducted in low- and high-incidence countries, and could have a similar design. As studies evaluating the public health impact will offer preventive treatment to those tested positive, there is less potential for (indication) bias than for clinical evaluation studies.

Study methods

All individuals enrolled in the study are followed for the same pre-specified period, at least 1 year after the completion of preventive treatment, irrespective of their test result and irrespective of whether they receive preventive treatment or not. Follow-up after completion of preventive treatment is done to measure the incidence of TB post-treatment. The whole study population is assessed for the occurrence of TB disease at the end of the study period and identified cases classified as clinical TB are included in the outcome. Ideally the outcome assessment is done blinded to the initial test result to avoid differential verification or incorporation bias. Active follow-up (e.g. screening participants for TB symptoms during 3-monthly visits) is preferred above passive follow-up in particular to limit cohort attrition, which may otherwise be more likely to happen in the group with no preventive treatment. Ascertainment of incident clinical TB disease may be done according to routine practice, *i.e.* following national guidelines for diagnosing clinical TB.

Data analysis

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

The minimum list of variables to be collected is as suggested for studies of predictive ability (table 3). In addition, data should be collected on the acceptance of the novel test, acceptance of preventive treatment when test-positive, adverse events, cost of the whole test-and-treat intervention and that of the alternative

FIGURE 2 Example of study design for the evaluation of health impact. TST: tuberculin skin test; IGRA: interferon- γ release assay; TB: tuberculosis; AE: adverse event; 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. Reproduced and modified from [13] with permission.

strategy. In addition to a direct comparison of the effectiveness of the strategy, the study may also report on the cost and cost-effectiveness of the tested strategy and the occurrence of adverse effects. All these outcomes inform the positive and negative implications of scaling up the novel test-and-treat strategy and its potential budget implications.

While the research questions outlined earlier serve to generate a minimum set of evidence, other analyses may be worthwhile to further explore using the data from the studies described. For instance, the predictive utility of different cut-off levels of the test may be explored in different subgroups. Furthermore, the predictive ability of the test may be improved when combined with other patient parameters [28]. Lastly, the long-term public health impact may be assessed by varying cut-offs or prediction models in combination with different preventive treatment regimens.

Concluding remarks

We described the design standards for two sets of research questions that need to be addressed to inform development of policy guidance for the use of novel incipient TB tests by WHO, with the intent to help test developers, manufacturers and others to design appropriate studies and report informative outcome measures.

For clinical evaluation, comparative studies, while providing the highest quality of evidence when using randomisation of subjects, are hampered by high cost. A way to minimise these costs is to design the study such that multiple research questions can be answered using the same design, of which several examples are known (ClinicalTrials.gov identifier NCT02735590 [24, 29]). Another option is 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 [10, 11, 15]. For health impact studies, an alternative is to use mathematical modelling approaches to estimate the potential population-level impact of the test-and-treat intervention under different circumstances.

Altogether, the development of novel, highly specific tests for identifying individuals with incipient TB who are likely to develop clinical TB with high accuracy may fill an important gap in the existing repertoire of TB diagnostics. When combined with highly effective preventive treatment in the intended target population, such tests can accelerate the progress towards TB elimination.

Acknowledgements: We thank the Stop TB Partnership New Diagnostics Working Group (NDWG) for their coordination of the Task Force on “Latent TB infection and test of progression” and coordinating the expert and stakeholders’ consultation meetings. We thank Alessandra Varga (NDWG, Geneva, Switzerland) for critical review of an earlier version of this manuscript and for facilitating the expert meetings. We also thank all participants that participated in the technical expert consultation meeting on February 8, 2017: Dick Menzies (McGill University, Montreal, QC, Canada), Helen Ayles (LSHTM/Zambart, London, UK), Pauline Beattie (EDCTP, The Hague, The Netherlands), Grania Brigden (The Union, Paris, France), Hanif Esmail (University of Oxford, Oxford, UK), David Lewinsohn (University of Portland, Portland, OR, USA), Thomas Scriba (University of Cape Town, Cape Town, South Africa), Marieke van der Werf (ECDC, Stockholm, Sweden), Kevin Windthrop (Oregon HS, Portland, OR, USA), Jean-Pierre Zellweger (Swiss Lung Association, Bern, Switzerland), Ruvandhi Nathavitharana (TB Proof, South Africa), Khairunisa Suleiman (Global TB Community Advisory Board Diagnostic Workgroup, Nairobi, Kenya), Norbert Ndjeka (Drug-Resistant TB, TB and HIV, National Dept of Health, Pretoria, South Africa), Rohit Sarin (National Institute of TB and Respiratory Diseases, New Delhi, India), Irina Vasilieva (Ministry of Health, Moscow, Russian Federation), Nguyen Viet Nhugn (National TB Programme and National Lung Hospital, Hanoi, Vietnam), René Becker-Burgos (Global Fund, Geneva, Switzerland), Thomas Forissier (Bill and Melinda Gates Foundation, Seattle, WA, USA), Alexandra Asbach-Nitzsche (Lophius, Regensburg, Germany), Jeff Boyle (Qiagen, Germantown, MD, USA), Anke Coblenz (Abbott, Des Plaines, IL, USA), William Cruikshank (Oxford Immunotec, Oxford, UK), Philippe Jacon (Cepheid, Maurens-Scopont, France), Masae Kawamura (Qiagen, Germantown, MD, USA), Oksana Markova (Generium, Moscow, Russian Federation), Chris Novak (Roche, Basel, Switzerland), Morten Ruhwald (Staten Serum Institute, Copenhagen, Denmark), Andrae Vinson (BD, Research Triangle Park, NC, USA), Haileyesus Getahun (WHO, Geneva, Switzerland), Yohhei Hamada (WHO, Geneva, Switzerland), Alexei Korobitsyn (WHO, Geneva, Switzerland), Karin Weyer (WHO, Geneva, Switzerland).

Conflict of interest: C.M. Denking: FIND (Foundation for Innovative New Diagnostics) is a not-for-profit foundation, whose mission is to find diagnostic solutions to overcome diseases of poverty in low- and middle-income countries (LMICs). It works closely with the private and public sectors and receives funding from some of its industry partners. It has organisational firewalls to protect it against any undue influences in its work or the publication of its findings. All industry partnerships are subject to review by an independent Scientific Advisory Committee or another independent review body, based on due diligence, target product profiles and public sector requirements. FIND catalyses product development, leads evaluations, takes positions and accelerates access to tools identified as serving its mission. It provides indirect support to industry (e.g. access to open specimen banks, a clinical trial platform, technical support, expertise, laboratory capacity strengthening in LMICs) to facilitate the development and use of products in these areas. FIND also supports the evaluation of prioritised assays and the early stages of implementation of WHO-approved (guidance and prequalification) assays using donor grants. In order to carry out test validations and evaluations, FIND has product evaluation agreements with several private sector companies, which strictly define its independence and neutrality *vis-a-vis* the companies whose products get evaluated, and describe roles and responsibilities.

References

- 1 Abu-Raddad LJ, Sabatelli L, Achterberg JT, *et al.* Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Proc Natl Acad Sci USA* 2009; 106: 13980–13985.
- 2 Lonnroth K, Migliori GB, Abubakar I, *et al.* Towards tuberculosis elimination: an action framework for low-incidence countries. *Eur Respir J* 2015; 45: 928–952.
- 3 Houben RM, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS Med* 2016; 13: e1002152.
- 4 Getahun H, Matteelli A, Chaisson RE, *et al.* Latent *Mycobacterium tuberculosis* infection. *N Engl J Med* 2015; 372: 2127–2135.
- 5 Diel R, Loddenkemper R, Nienhaus A. Predictive value of interferon-gamma release assays and tuberculin skin testing for progression from latent TB infection to disease state: a meta-analysis. *Chest* 2012; 142: 63–75.
- 6 Rangaka MX, Wilkinson KA, Glynn JR, *et al.* Predictive value of interferon-gamma release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis* 2012; 12: 45–55.
- 7 Getahun H, Matteelli A, Abubakar I, *et al.* Management of latent *Mycobacterium tuberculosis* infection: WHO guidelines for low tuberculosis burden countries. *Eur Respir J* 2015; 46: 1563–1576.
- 8 World Health Organization. Guideline on the Management of Latent Tuberculosis Infection. Geneva, WHO, 2015.
- 9 Petruccioli E, Scriba TJ, Petrone L, *et al.* Correlates of tuberculosis risk: predictive biomarkers for progression to active tuberculosis. *Eur Respir J* 2016; 48: 1751–1763.
- 10 Sloot R, Schim van der Loeff MF, van Zwet EW, *et al.* Biomarkers can identify pulmonary tuberculosis in HIV-infected drug users months prior to clinical diagnosis. *EBioMedicine* 2015; 2: 172–179.
- 11 Suliman S, Thompson E, Sutherland J, *et al.* Four-gene pan-African blood signature predicts progression to tuberculosis. *Am J Respir Crit Care Med* 2018; 197: 1198–1208.
- 12 Cobelens F, Kik S, Esmail H, *et al.* From latent to patent: rethinking prediction of tuberculosis. *Lancet Respir Med* 2017; 5: 243–244.
- 13 World Health Organization, FIND. 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 (WHO/HTM/TB/2017.18). Geneva, WHO, 2017.
- 14 Comstock GW, Livesay VT, Woolpert SF. The prognosis of a positive tuberculin reaction in childhood and adolescence. *Am J Epidemiol* 1974; 99: 131–138.
- 15 Zak DE, Penn-Nicholson A, Scriba TJ, *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016; 387: 2312–2322.
- 16 Weyer K, Mirzayev F, Migliori GB, *et al.* Rapid molecular TB diagnosis: evidence, policy making and global implementation of Xpert MTB/RIF. *Eur Respir J* 2013; 42: 252–271.
- 17 World Health Organization. Handbook for Guideline Development. Geneva, WHO, 2012.
- 18 European Commission. Regulatory Framework on Medical Devices. 2017. https://ec.europa.eu/growth/sectors/medical-devices/regulatory-framework_en Date last accessed: May 21, 2018.
- 19 Food and Drug Administration. Human Subject Protection; Acceptance of Data From Clinical Investigations for Medical Devices. Final Rule. *Fed Regist* 2018; 83: 7366–7388.
- 20 US food and Drug Administration. Acceptance of Clinical Data to Support Medical Device Applications and Submissions, Frequently Asked Questions, Guidance for Industry and Food and Drug Administration Staff. 2018. www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM597273.pdf Date last accessed: July 17, 2018.
- 21 Duffy FJ, Thompson E, Downing K, *et al.* A serum circulating miRNA signature for short-term risk of progression to active tuberculosis among household contacts. *Front Immunol* 2018; 9: 661.
- 22 Fiore-Gartland A, Carpp LN, Naidoo K, *et al.* Considerations for biomarker-targeted intervention strategies for tuberculosis disease prevention. *Tuberculosis* 2018; 109: 61–68.
- 23 Mathad JS, Bhosale R, Balasubramanian U, *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: 1421–1428.
- 24 Rangaka MX, Wilkinson RJ, Boulle A, *et al.* Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. *Lancet* 2014; 384: 682–690.
- 25 Sharma SK, Vashishtha R, Chauhan LS, *et al.* Comparison of TST and IGRA in diagnosis of latent tuberculosis infection in a high TB-burden setting. *PLoS One* 2017; 12: e0169539.
- 26 Theron G, Zijenah L, Chanda D, *et al.* Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet* 2014; 383: 424–435.
- 27 Hemming K, Haines TP, Chilton PJ, *et al.* The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ* 2015; 350: h391.
- 28 Saunders MJ, Wingfield T, Tovar MA, *et al.* A score to predict and stratify risk of tuberculosis in adult contacts of tuberculosis index cases: a prospective derivation and external validation cohort study. *Lancet Infect Dis* 2017; 17: 1190–1199.
- 29 Mahomed H, Hawkridge T, Verver S, *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: e17984.