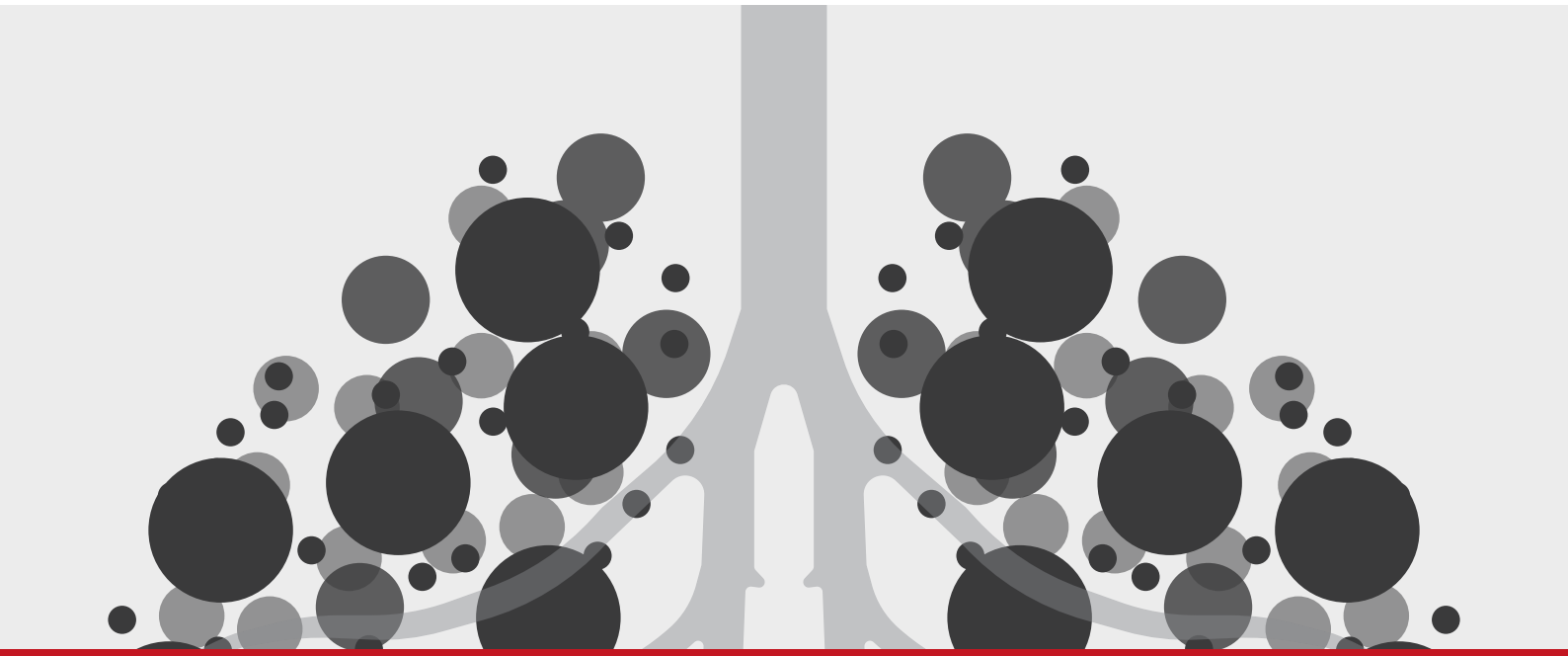
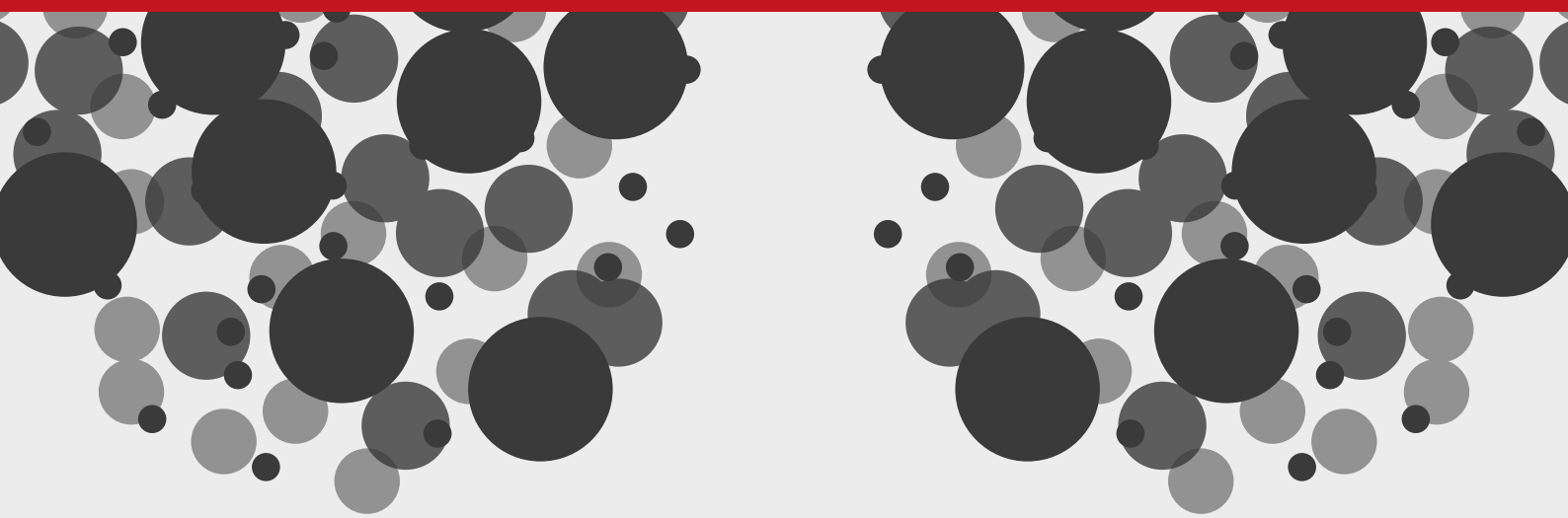
Stop TB Partnership

New Diagnostics Working Group



Meeting Report

SECOND EXPERT WORKSHOP ON THE DEVELOPMENT OF TESTS FOR PROGRESSION OF LATENT TUBERCULOSIS INFECTION (LTBI) TO ACTIVE DISEASE



1st July 2016

San Raffaele Hospital, Milan, Italy

Stop TB Partnership

New Diagnostics Working Group

Meeting Report

SECOND EXPERT WORKSHOP ON THE DEVELOPMENT OF TESTS FOR PROGRESSION OF LATENT TUBERCULOSIS INFECTION (LTBI) TO ACTIVE DISEASE

1st July 2016

San Raffaele Hospital, Milan, Italy

ACKNOWLEDGEMENTS

The New Diagnostics Working Group (NDWG) gratefully acknowledges the contributions of the following individuals:

- Alberto Matteelli (University of Brescia) for his extensive work in coordinating the NDWG Task Force on LTBI and leading its efforts;
- Claudia Denkinger and Samu https://open.spotify.com/track/0fg8CqpjdojMyXLNzM2PaJ el Schumacher (FIND) for their considerable work in developing the Target Product Profile;
- Frank Cobelens (AIGHD) and Sandra Kik (KNCV) for their considerable work in developing the Framework for Evaluation document;
- Other members of the NDWG Task Force on LTBI, who gave valuable feedback on earlier versions of these documents: Gavin Churchyard (Aurum Institute), Christopher Gilpin (WHO), Delia Goletti (National Institute of Infectious Diseases, Italy), Amita Gupta (Johns Hopkins), Richard Menzies (McGill), Lele Rangaka (UCL);
- The session chairs and rapporteurs: Daniela Maria Cirillo (San Raffaele Hospital and NDWG Co-Chair), Hanif Esmail (Oxford University Hospital), Helen Ayles (Zambart/LSHTM), Christian Lienhardt (WHO) in addition to task force members already listed above, as well as presenters: David Lewinsohn (University of Oregon) and Thomas Scriba (University of Cape Town) in addition to task force members already listed;
- All the experts who participated in the meeting and contributed with valuable input in the discussions;
- All the stakeholders who contributed with input through the online consultation which was conducted prior to the meeting;
- Susanna Capone (University of Brescia) for preparing the reports of the online consultation and of the meeting.

EXECUTIVE SUMMARY

Introduction

The fourth component of the first pillar of the WHO End- TB strategy is preventive treatment of persons at high risk of developing active disease; and vaccination against tuberculosis (TB). A direct measurement tool for *M. tuberculosis* (Mtb) infection in humans is currently unavailable. A major research gap for cost-effective management of persons at high risk is the development of diagnostic tests with improved performance and predictive value for progression to active TB.

The Task Force on Latent Tuberculosis Infection of the New Diagnostics Working Group (NDWG) and its partners convened a meeting on 1st July 2016 in Milan, Italy to gather expert advice on:

- new evidence on the nature and significance of LTBI and its relevant implications for the conceptualization of diagnostics;
- development of the Target Product Profile (TPP) for a test of progression of LTBI;
- development of a guidance document for an optimized study design that will produce data for test evaluation and policy development.

Nature and significance of LTBI and limitations of diagnostic tests

Current diagnostic tests – interferon gamma release assays (IGRA) and the tuberculin skin test (TST) – identify historical exposure (and immune sensitization) to Mtb, generally remain positive when infection is cleared (either spontaneously or with preventive treatment), and do not distinguish those most likely to develop disease. Thus, they have a poor positive predictive value (PPV) for predicting active TB, which translates into a very high 'number needed to treat' (NNT) in order to prevent one TB case through preventive therapy.

What tests are needed – two novel options

Tests for persistent infection

The first type should identify those with persistent infection (i.e. results are negative after infection has cleared). Such a test will likely not identify those at greatest risk of progressing to TB due to a low PPV, but could be used to single out those at high risk of progression to severe disease, such as patients with HIV infection, those awaiting anti-TNF-alpha treatment, and infants. These are defined as tests for persistent infection.

Tests for incipient TB

The second test should detect that the disease is active while the patient is still asymptomatic. It would be highly predictive for clinical disease, in particular for those recently exposed. These are defined as tests for incipient disease. The 16-transcript disease risk signature recently described by Zak *et al.* is an example of a biomarker that may fulfil this role. The incipient TB test should have a semi-quantitative read-out and might potentially revert to negative after treatment.

Target product profile for a test of progression

The reasonable time horizon for a test of incipient TB should be prediction of future progression to active TB within two years, taking into account that ~60% of progression occurs in this timeframe (~45% in year one). Acceptable PPV and NNT values – as identified by patients, clinicians and policy makers – were proposed. Minimal performance is represented by an increase of the PPV by factor of ~2 compared to IGRAs. Optimal performance is represented by an increase of the PPV by ~5 compared to IGRAs.

The test should be developed with combinations of sensitivity/specificity that are compatible with such improved values of PPV and NNT. Expectations for accuracy should not be the same for a predictive test as they are for a diagnostic test: even with a very high sensibility and specificity (99%), a low PPV would be reached (67%). The specificity threshold for candidate tests for incipient TB should be 50% under minimal performance.

PPV in a given setting should be considered as an important parameter to guide decisions on implementation. Different country programmes will have differing preferences in the trade-off between sensitivity and specificity (for example, re-testing of individuals with an initial negative result could be an attractive option for programmes that aim to maximize sensitivity).

Guidance to optimized study design

To be consistent with the WHO process for endorsement of a diagnostic test, the test evaluation programme should establish the ability of the test to predict active TB and its health impact from both the patient and community perspectives.

A. Assessing predictive ability

The study population should be represented by individuals at risk of being infected and at risk of disease progression, and who are observed over time. Individuals who do not receive preventive treatment would be desirable in order to avoid biased results (i.e. contacts of MDR-TB patients and other contacts that do not accept the offer of preventive treatment).

B. Assessing the public health impact

Studies around the public health impact should assess efficacy, cost-effectiveness, treatment adherence, and side effects. These should be comparative, randomized intervention studies that compare the impact of a test-and-treat strategy based on the new test with the standard practice (i.e. test-and-treat strategy based on TST and/or IGRA; or no LTBI testing and treating at all).

BACKGROUND

The targets of the End-TB Strategy will not be achieved without accelerating efforts in support of the prevention package, which requires addressing latent TB infection (LTBI) diagnosis and treatment. To this aim, it is essential to develop newer diagnostic tests with significantly increased predictive value for the development of active disease among those who are infected.

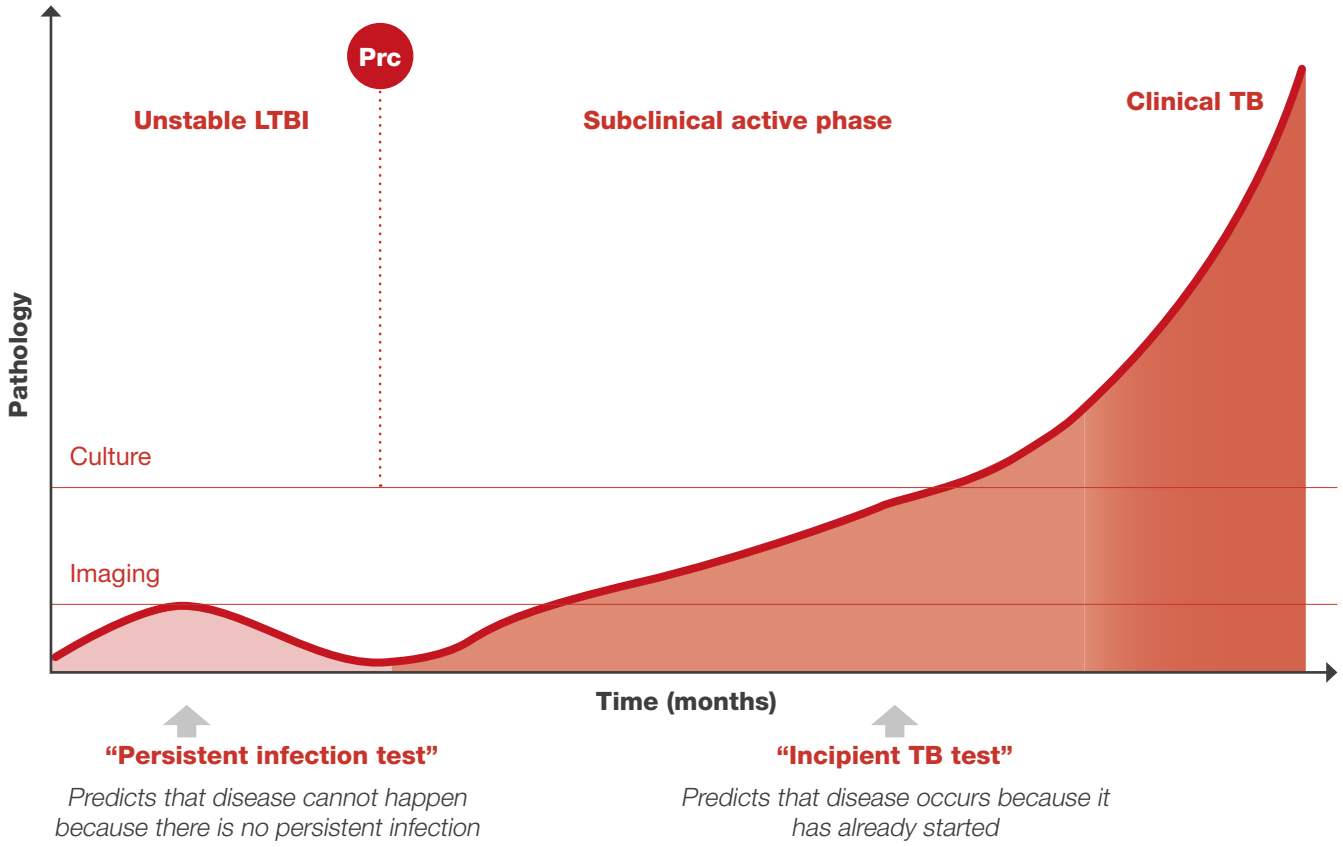
On May 14th 2015, the New Diagnostics Working Group (NDWG) and FIND convened an Expert workshop for the development of best practices for performance and cost-effectiveness studies of tests targeting LTBI. Hosted by the WHO in Geneva, the meeting's aim was to develop a Target Product Profile (TPP) to provide a framework for the test development and its evaluation. A second objective of the meeting was to plan for a document to guide diagnostic developers, by outlining the design of studies that should be conducted to generate the evidence required for test endorsement by WHO.

To support progression and finalization of these two documents, the LTBI Task Force of the NDWG and its partners convened a follow-up meeting on 1st July 2016 in Milan, Italy.

The objectives of the meeting were to gather expert advice on:

- new evidence on the nature and significance of LTBI, and its relevant implications for the conceptualization of diagnostics;
- further development of the Target Product Profile for a test of progression of LTBI;
- further development of a guidance document for an optimized study that will produce data for policy development.

Figure 1 What tests are needed – two novel options



SUMMARY OF PRESENTATIONS AND DISCUSSIONS

LTBI conception: definitions and relevance for diagnostic products

Frank Cobelens presented a concept lecture for this session, in which he proposed a model for understanding the natural history of TB based on recent published reviews with a focus on how novel tests might target different stages of TB natural history.

The inadequacy of the current diagnostics on management of LTBI in the context of the End-TB strategy was highlighted. Current diagnostic tests – interferon gamma release assays (IGRAs) and the tuberculin skin test (TST) – which identify historical exposure (and immune sensitization) to *Mycobacterium tuberculosis* (Mtb), generally remain positive when infection is cleared (either spontaneously or with preventive treatment), and do not distinguish those most likely to develop disease.

Thus, they have a poor positive predictive value (PPV) for active TB, which translates into a very high 'number needed to treat' (NNT) in order to prevent one TB case through preventive therapy. PPV and NNT depend on the average risk of progression from infection to disease and the probability of having earlier, remote infection, which varies between different epidemiological settings. This implies that the use of LTBI tests will also vary in different settings.

Figure 1 outlines the conceptual model showing the transition from latent infection to clinical disease through a period of subclinical active disease during which pathology evolves over a period of time in the absence of symptoms. This translation is triggered by a precipitating event, which may or may not be apparent. The evolving disease may be detected by culture or imaging. (Adapted from Esmail H, Barry CE 3rd, Young DB, Wilkinson RJ. 2014 Phil. Trans. R. Soc. B.)

Although the main strategic purpose of novel LTBI tests should be to identify infected individuals who are at high risk of progression to active disease, it was acknowledged that two types of novel tests might require development.

The first type of tests would identify those with a persistent infection (i.e. becomes negative when infection is cleared). Conceivably, such a test will not distinguish those at greatest risk of progression and will suffer from relatively low PPV; however, it would provide a good rule-out test that could be useful, in particular, for those at high risk of progression to (severe) disease, such as individuals with HIV infection, those who are going to be receiving anti-TNF-alpha treatment, and infants.

The second type of test would detect that the active disease process has already been precipitated while an individual was still asymptomatic, and would potentially be highly predictive for clinical disease and potentially a good rule-in test, in particular for those recently exposed. The recent 16-transcript disease risk signature by Zak *et al.* (Lancet, 2016) is an example of a biomarker that may achieve this.

The subsequent discussion focused on key questions raised at the end of the presentation.

1. Do we agree with the conceptual framework outlined in the presentation? How do we deal with the TPP issue of 'rule-in' vs. 'rule-out'?

There was general consensus on the proposed paradigm and framework for the two categories of novel tests (persistent TB test and incipient TB test), based on current scientific evidence. It was suggested that two separate TPPs might be considered due to their potential different roles (rule-in vs. rule-out) and performance characteristics. However, it was agreed that the test for incipient/subclinical TB is the current priority, and that the term 'incipient TB' is preferable to 'subclinical TB'. While 'incipient TB' more precisely describes the initial appearance of disease, 'subclinical TB' has a number of different circulating definitions, which could potentially lead to confusion.

2. What is the reference time to define the performance for a test of incipient TB?

It was discussed whether the previously proposed two-year period is justified to determine the PPV, and it was noted that if the conceptual framework for the test is accurate, then the duration between the occurrence of a precipitating event and the onset of the clinical signs or symptoms of the disease is the time frame that must be identified. The uncertainties in this area depend mainly on the different diagnostic abilities of the assays. For example, using serial CT scans, evidence of disease might be identified much earlier than by using microbiological or molecular tests for the identification of Mtb. Standard radiology provides similar evidence. Despite these uncertainties, there was consensus that the time period to define PPV and NPV should not be greater than two years, and would more likely be from one year to 18 months. In this and subsequent sessions, it was discussed that sensitivity estimates of an incipient TB test would likely be lower the longer the follow-up period. This was deemed likely for tests that are thought to detect early disease processes, which, by definition, can only be positive if these processes have already started and will not pick up cases whose development is delayed after the time of testing.

3. Should an assay designed to detect incipient TB be able to identify the active TB status? Should it have a semi-quantitative read-out? Should it revert to negative after effective treatment?

There was a general view that a test for incipient TB would detect early active TB – it is therefore unrealistic that such a test could revert to negative in clinical disease. However, there was agreement that it would be desirable for an incipient TB test to have a semi-quantitative read-out, perhaps reflecting the burden of TB, and that it might be possible to identify a threshold within the semi-quantitative scale to distinguish incipient from active disease. This test characteristic would be extremely advantageous, as it would assist in defining the most appropriate treatment strategy (preventive therapy as opposed to TB treatment).

There was consensus that it would be desirable for an incipient TB test to revert to negative following treatment; however, this is not considered a priority at this time.

Target Product Profile for a test of progression

Results of online consultation

Susanna Capone presented the results of the expanded, online survey carried out on the TPP for a test of progression and explained the conception and organization of the survey, as well as the main results.

The survey was created following the structure of the TPP, which included intended use, performance characteristics, operational characteristics, and pricing. Among the 31 items that constitute the TPP, 10 were included in the survey according to their scientific relevance and implementation applicability, and the importance to receive additional input from a broad group of stakeholders.

A total of 473 potential participants were invited to take part in the survey, including:

- members of the NWDG task-force on LTBI;
- LTBI international experts;
- members of the European Respiratory Society (ERS);
- representatives of multilateral and international agencies and organizations;
- members from NGOs, civil society, and patient groups;
- representatives of endemic countries; and
- test developers.

Among them, 76 responded as follows:

- academic and research institutions: 43%
- MOH, national TB Programmes, other national organizations: 19%
- implementer/clinicians, laboratory staff: 12%
- the industry: 8%
- remaining represented by advocacy/NGOs, international bodies, PDP/technical agencies: 4% to 5% each

No funders participated in the survey.

Respondents were from:

- Europe: 47%
- Americas: 21%
- Africa: 20%
- Asia: 12%

The survey questions and results are presented in Annex 3. As shown in the table, most of the answers by respondents showed a concordance with the proposed optimal and minimal targets. In fact, among all respondents there was an agreement of 80% or higher, except for questions 4, 5, and 10, which were related to the diagnostic sensitivity, specificity, and cost of consumables. Regarding 'Diagnostic sensitivity for progression to active TB', 92% of respondents agreed on 90% as an optimal target, whereas 15% of respondents disagreed on 75% as a minimal target, proposing 85% as an alternative. Concerning 'Diagnostic specificity for risk of progression to active TB', 89% of participants agreed on 90% as an optimal target, whereas only 66% of them agreed on 75% as a minimal target, suggesting 90% as a more suitable option.

Finally, regarding 'Cost of consumables (reagents/test strips)', the proposed cost of less than $US5 (under the optimal scenario) was accepted by most of the participants (82%); however, no consensus was reached on the minimal proposed cost of less than $US150, which was considered too high and unaffordable (48% disagreed).

A range of $US30 to $US100 was suggested by respondents, considering the possibility of differentiating costs between high- and low-income countries.

The most relevant comments emerging from the survey are summarized in Annex 4.

Draft Target Product Profile

Samuel G. Schumacher presented an in-depth explanation of the TPP rationale and purpose, which lead to the discussion on the survey results lead by Delia Goletti. The discussion focused particularly on intended use/goal/targets and on performance characteristics.

A) Intended use / goal / target condition

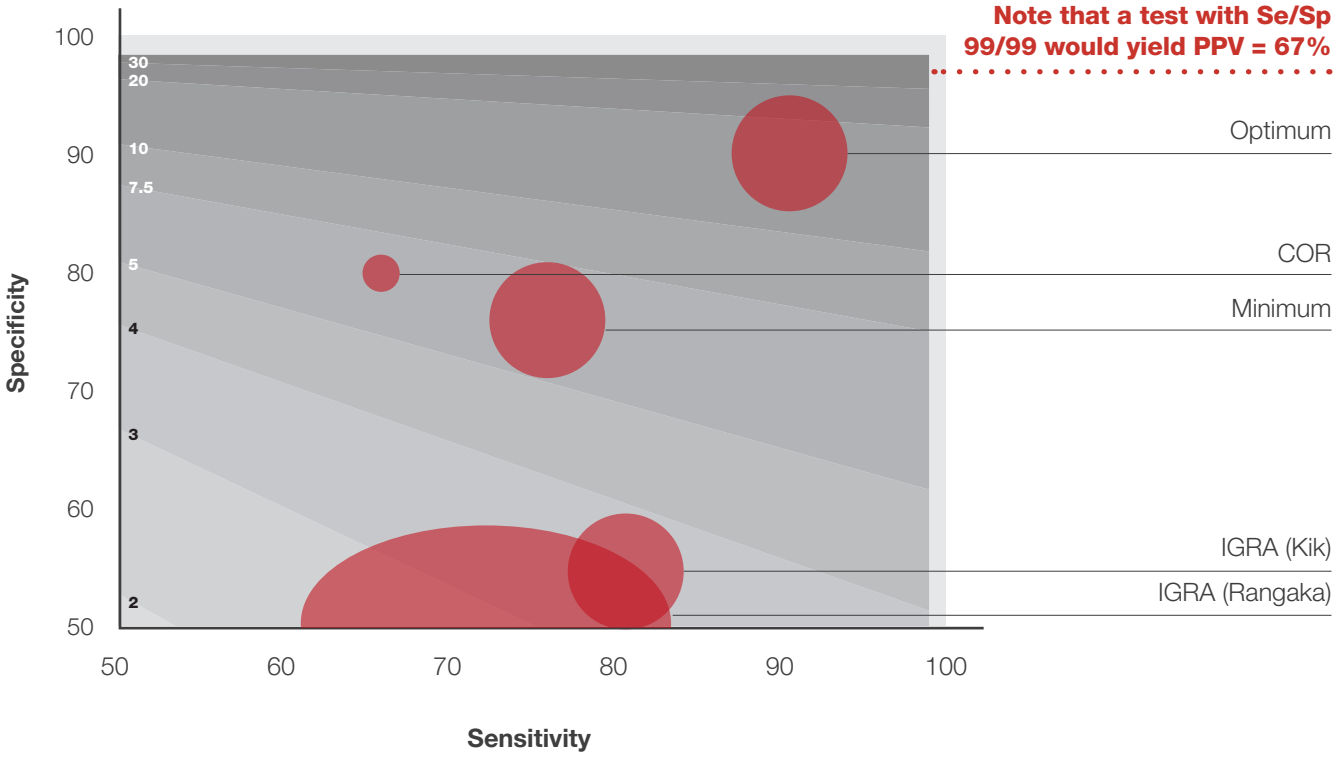
The rationale for a test of incipient TB to predict future progression to active TB in a two-year time horizon was considered a reasonable, pragmatic choice, taking into account that ~60% of progression occurs in the first two years (~45% in year 1). The most promising approach to predict progression is currently thought to be via the detection of incipient TB (which, by definition, would be relatively close to the onset of active disease). Late progression might occur due to precipitating factors, which could not be predicted at the time of testing.

B) Performance targets

One key reason for the limited uptake and adherence of Isoniazid Preventive Therapy (IPT) is that the risk/benefit-profile for preventive treatment is not convincing for patients, clinicians, and public health collaborators. In fact, treatment is imperfect (efficacy, duration, AEs, etc.) and TST/IGRA have a low accuracy for risk of progression (low PPV and high NNT).

The risk/benefit-profile (and its PPV and NNT metrics) is therefore a key element for the overall success of the strategy. The PPV is able to capture the patient perspective: if the test is positive, how likely am I to develop the disease if I don't get treated? whereas NNT captures the clinician/patient perspective: if treating all the individuals who test positive, how many do I need

Figure 2 'Positive Predictive Value' according to Sens/Spec for risk of progression



2% Cumulative incidence

to test-and-treat to prevent one case? While PPV and NNT are thus helpful intuitively to decide on desirable performance, they are a result of both tests' performance and underlying probability of progression.

In a TPP, test accuracy is better defined using metrics relatively invariant to context- and population-specific factors and thus sensitivity and specificity (or likelihood ratios - LR+/-) are more appropriate. According to these premises, a two-step approach to determine the performance targets was followed:

Step 1: Identifying values of PPV and NNT that are currently found acceptable by the patients/clinicians/policy makers. This required estimating current PPV/NNT values in vulnerable groups for whom IPT is currently recommended by WHO using accuracy estimates for TST/IGRA and estimates for risk of progression.

- accuracy estimates for TST/IGRA came from a systematic review (Rangaka *et al.*, 2011) and an updated but unpublished version of this systematic review that was conducted to inform the WHO LTBI guidelines (Kik *et al.*, 2014);
- estimates for risk of progression were based on expert opinion based on a non-systematic review of the published literature for the relevant risk groups.

Step 2: Defining combinations of sensitivity/specificity that are compatible with improved values of PPV and NNT. This required proposing targeted improvements for PPV/NNT (in reference to what was found in Step 1) for the minimal and optimal performance targets and then assessing what sensitivity/specificity is needed to achieve these targets in different key target populations.

- for the minimal performance target, an increase of the PPV by factor of ~2 and (thus cutting NNT by ~1/2) compared to IGRA was proposed;
- for the optimal performance target, an increase of the PPV by factor of ~5 and (thus cutting NNT by ~1/5) compared to IGRA was proposed;
- to assess the various combinations of sensitivity/specificity that are compatible with these proposed values of PPV/NNT and present them in relation to the performance of IGRA, they were shown in contour plots (see Figure 2); the correlate of risk (COR) signature recently found by Zak (Zak *et al*, Lancet 2016) was also plotted for comparison purposes.

It was noted that the PPV achieved is still relatively low (e.g. ~6% for the minimal target) but would represent an important improvement over what is currently available and seems achievable within a five-year time horizon (as evidenced by the COR, meeting this target). Further, it was highlighted that even a test with 99% sensitivity and 99% specificity would only achieve a PPV of 67% (given a 2% cumulative incidence); and that expectations for accuracy cannot be the same for a predictive test as they are for a diagnostic test.

The other main areas of discussion were the following:

- The importance to spell out the rationale behind targets in sufficient detail was pointed out and it was suggested to include the figures in the TPP.
- It was suggested to also show the performance targets in GRADE tables and spelling out the consequences for true positives (TPs), false negatives (FNs), true negatives (TNs), and false positives (FPs).
- It was acknowledged that incipient test performance would vary with duration of follow-up. Sensitivity was generally expected to peak between one and two years, but it was accepted that this may vary with patient population (e.g. children, HIV co-infection). The time for evaluating sensitivity and specificity of tests designed to detect the incipient TB should be minimally 12 months and maximally 24 months, and the follow-up period should be clearly specified when reporting on any accuracy measures. Preferably, accuracy should be reported for 12 months, 18 months, and 24 months if studied.
- For any given sensitivity and specificity of a test, PPV will vary from setting to setting. Therefore, it is the PPV in a given setting that has to be considered as an important parameter to guide decisions on implementation.
- Different country programs will have differing preferences in the trade-off between sensitivity and specificity (and, for example, re-testing of individuals with an initial negative result could be an attractive implementation option for programs wishing to maximize sensitivity – sensitivity will be improved if repeat testing is done and conversion from a negative to a positive test occurs).
- Although there was discussion about minimal sensitivity and specificity, it was agreed that any sensitivity and specificity combinations offering improved performance in terms of PPV and NNT would be of interest. It was recommended that the specificity threshold for candidate tests for incipient TB should be 50%.

Framework for validation of candidate products

Sandra Kik presented the 'Guidance document: how to evaluate TB prediction tests to inform WHO endorsement'.

The overall expectations for a TB prediction test were recalled: being positive in case of progression from LTBI to TB and either positive or negative in case of active TB (which is dependent upon the type of immune response that is being measured by the test); being negative in all other cases (unexposed, with LTBI but not progressing, with treated LTBI and TB). The test would have a low probability of being positive with a consequent high number needed to screen, but a high probability of disease if positive, and thus a low number needed to treat. The novel test should be largely independent from the study population and predict the disease over a short time period (two years). Two test development phases – analytical and field evaluation – were identified.

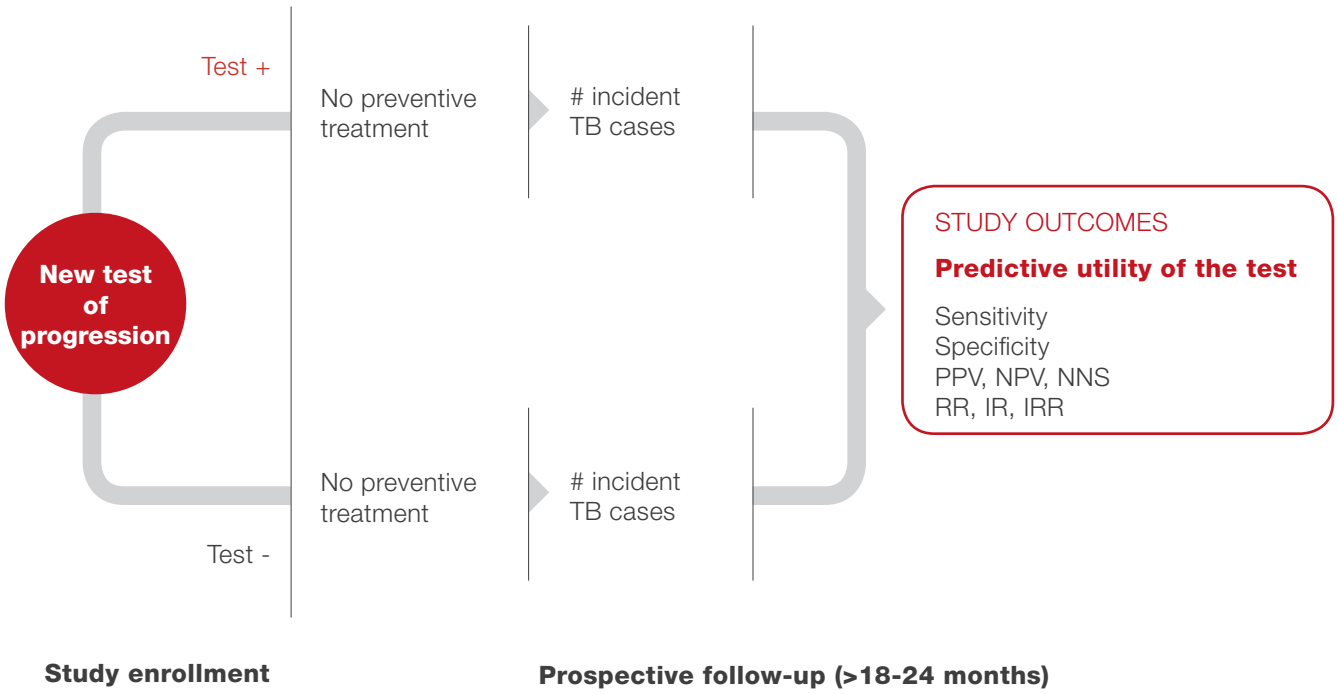
The analytical phase, which was described as the laboratory research phase, is usually implemented by test developers. Easily accessible samples in established repositories would be tested, aiming to assess robustness, variability, and repeatability. This phase is not within the scope of the Guidance Document.

The field evaluation phase, which falls within the scope of the document, concerns clinical studies performed in intended target populations. To be consistent with the WHO grade process endorsement for a diagnostic test, the test evaluation program should mainly establish 1) the predictive ability of the test to predict active TB and 2) the health impact from the patient and community perspectives. Different study designs (longitudinal prospective cohort study, nested control study, etc.) are appropriate to investigate specific questions.

A) Assessing predictive ability

A first example of study design for evaluating predictive ability is shown below. In this phase, the study population identified should be represented by individuals at risk of being infected and at risk of disease progression. Individuals who do not receive preventive treatment would be desirable, in order to avoid biased results of the test's predictive ability, though this may be programmatically difficult to achieve. Contacts of MDR-TB patients and other contacts that do not accept or are not routinely offered preventive treatment were mentioned as a potential population of interest.

Figure 3 Example of study design for evaluating predictive ability of the test



Those enrolled in trials should receive initial screening when active TB is excluded by standard methods and a first test is carried out. During the follow-up, additional testing may be conducted according to a pre-determined time-frame and incident TB should be detected prospectively (by active case finding or passive patients' follow-up). The primary endpoint would be 'bacteriologically confirmed diagnosis of incident TB' vs. 'no or negative bacteriologically tests at end of trial'. While both test negatives and test positives should be followed up to assess the occurrence of incident TB, bacteriological tests during the observational period may be restricted to those presenting with any symptoms during follow-up.

Recommended sub-analyses included assessment of predictive ability for different thresholds or the test, and assessment of predictive ability in sub-groups of individuals according to age, gender, country of origin, bacille calmette guerin (BCG) status, TST/IGRAs status, HIV status, presence of co-morbidities such as diabetes, etc.

Sandra Kik then discussed specificities, potential challenges, and mitigation strategies for study designs in low-income and high-income countries separately.

B) Assessing public health impact

Studies around patient and public health impact should assess efficacy, cost-effectiveness, treatment adherence, or side effects.

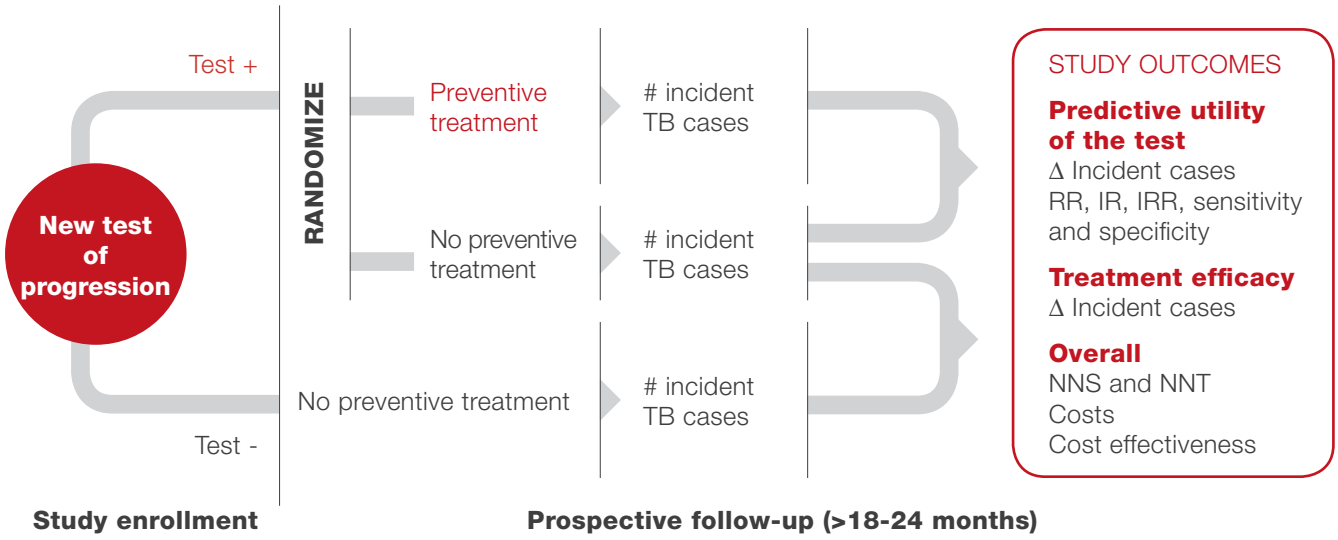
These should be intervention studies with comparative, randomized design (individual or cluster randomized) to minimize biases and increase the level of confidence in the measured effect; alternatively, 'before-after' studies (i.e. stepped-wedge or pre- and post-cohort) could be used even if they are more susceptible to biases.

Study populations should be those intended for the intervention. These studies could be carried out both in high burden countries (HBC) and low burden countries (LBC). As the final aim is to assess the public health impact of the test, these studies should compare new test-and-treat strategies with current practices.

Since current practices of LTBI screening differ between HBC and LBC, different study designs are suggested dependent on whether the study population of interest is currently recommended to undergo LTBI screening and treatment or not.

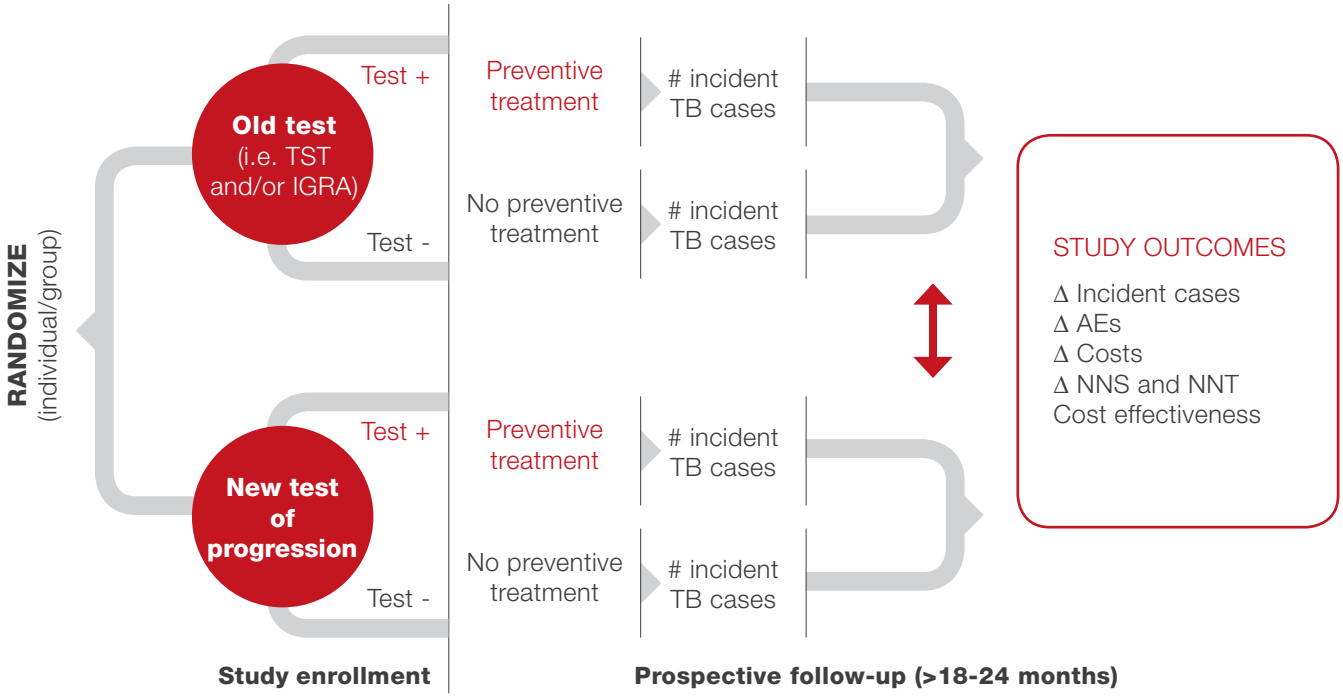
Two different examples are shown below:

Figure 4 Example of study design in populations that are currently not tested for LTBI



Source: Adjusted from M. Hatherill, Union Conference 2015, NDWG symposium, Design of CORTIS trial

Figure 5 Example of study design: evaluating public health impact in populations that arecurrently tested for LTBI tests



Each of these study designs allows addressing slightly different study outcomes, each of them relevant for programmatic evaluation. The first outlines a study where individuals who are not routinely tested for LTBI are evaluated with the new test and are randomized to receive preventive treatment or not if their test is positive. This design can provide estimates of the test performance and its predictive ability for active TB (sensitivity, specificity, Risk Ratio, Incidence Rate, PPV, NPV, etc.) when comparing with the arm that did not receive preventive treatment. In addition, the treatment efficacy can be estimated by comparing the number of incident TB cases among those receiving treatment and those who did not. Study populations of interest may include contacts of MDR-TB patients and other contacts that do not accept or are not routinely offered preventive treatment (i.e. HIV-uninfected contacts in HBCs), and previously treated TB patients.

The second study design outline visualizes a randomized trial where individuals or clusters are randomized to get tested with the new test or current LTBI tests and are recommended preventive treatment based on the test result. Such trials could generate evidence of the impact of new tests at the level of either the patient (if individuals are randomized) or the population (if clusters are randomized) by measuring the difference in the number of incident TB cases, adverse events, costs, NNS and NNT, and cost effectiveness compared to the current standard.

Study populations of interest may include those in whom testing and treatment are currently recommended by the WHO, such as close contacts and HIV-infected individuals with LTBI. In addition, the study of recent immigrants or travellers coming from HBCs who are currently screened for LTBI or adult contacts in HBC may be of interest.

In both cases, the active TB status should be excluded before starting the study, according to current standards. Follow-up practices should preferably be active, similar for those with a positive or negative initial test, and continued after treatment completion.

These studies could offer direct input for modelling studies that would evaluate the longer term public health impact.

A number of open questions remain: Are there sufficient settings and populations to implement these studies? Is the study duration – from 12 to 24 months – appropriate? How many tests (measurements) should be included in the follow-up?

The main points that emerged during the discussion are summarized below:

- Pragmatic studies, to compare a novel vs. standard approach, should have a superiority study design to assess the detection ability and a non-inferiority study design to assess incidence of side effects.
- Prevention of conditions like extra-pulmonary TB that are 'difficult to diagnose' could be an added value for tests of progression.
- Treatment regimens for individuals with a positive test for incipient TB are currently unknown.
- The performance of the test will refer strictly to a given – and pre-specified – time frame. Studies should therefore report the performance of the test at different time intervals – within the first 12 months, 18 months, or 24 months. Modifying the time frame – for example, extending prediction from 18 to 24 months – will change test performance.
- In order to not discourage the industry, requirements for the analytical phase of test development should not be too restrictive.

LTBI diagnostics pipeline and evidence on new candidate tests

Gavin Churchyard introduced the session by presenting and discussing available evidence on commercial tests of LTBI (QuantiFERON [QFT] Plus) and biomarkers for incipient TB (molecular, protein, antibody, and cellular [monocyte/lymphocyte ratio, cell activation markers, antigen specific T cells, response to latency antigens, CD8+ specific] responses). The conclusion was that the current pipeline of tests for incipient TB is extremely limited and that TPPs may help galvanize development of new tests of incipient TB.

Thomas Scriba presented a summary of data on research to develop a transcriptomic correlation of risk (COR) in an adolescent cohort. He recalled criteria for enrolment in the cohort (QFT and/or TST positive, no TB for first six months after enrollment, HIV-negative status) and the study design (retrospective construction of two cohorts, progressors, and controls). An interferon gamma (IFN) response signature (the transcriptomic COR consisting of 16 genes, 62 PCR primers, 257 primer pairs) emerged during the follow-up of two years and was characterized by moderate to high predictive power about TB risk. Based on a cumulative incidence of 2%, the COR showed a PPV ranging from 6% up to 14%, compared to a PPV of 2% to 3% for TST and IGRAs in South Africa. The development research program includes three steps.

First, to simplify the COR by reducing the primers from 62 to six. Second, to adapt the COR to a diagnostic, PCR-based platform. Third, to demonstrate that the COR has potential as a triage test for prevalent TB at time of sampling. This will be tested in a randomized, partially-blinded, clinical trial of isoniazid and rifapentine (3HP) therapy to prevent pulmonary TB in high-risk individuals identified by a transcriptomic COR (the CORTIS trial). Researchers will screen approximately 10,000 HIV-uninfected adults for transcriptomic COR to enroll 3,200 COR+ (randomly distributed to treatment or surveillance) and COR- subjects (all for surveillance) and measure incident TB disease in 15 months.

Finally, David Lewinsohn discussed the role of CD8 T cells as a tool to evaluate the progression from LTBI to active TB. There is evidence that CD8 T cells response reflects Mtb load: response is higher in active TB vs. LTBI, in smear positive patients vs. smear negative, and in pulmonary TB vs. extra-pulmonary TB. Moreover, the response decreases after effective TB treatment. If CD8 T lymphocyte frequency is associated with the bacterial burden, they could positively mirror the progressive steps from exposure, to infection, to incipient disease, to overt disease.

Conclusions and next steps

The meeting achieved substantial progress towards the production of a TPP for a test of progression.

Extensive discussion identified the areas in which the developers need to concentrate for both the TPP for a test of progression and the guidance document on trial design in order to finalize drafts. Further work is ongoing.

There was general agreement that the process for the development of an additional TPP for a test of persistent infection should be started, although scheduled for a later stage.

A decision was taken to propose a WHO-convened meeting to be held at the beginning of 2017 to obtain a larger consensus on the topics discussed at the workshop, and to produce WHO-endorsed TPP and guidelines on study design for a test identifying TB progression.

ANNEX 1. MEETING AGENDA

1st July 2016

Organized by the New Diagnostics Working Group
Hosted by San Raffaele Hospital, Milan, Italy

Meeting room 2 A2
San Raffaele Telethon Institute for Gene Therapy
Building DIBIT 1, 2nd floor
Via Olgettina 58 – 20132 Milan

SESSION 1: Welcome and introduction
Chairs: Alberto Matteelli and Hanif Esmail

09:00	Welcome and opening	Daniela Cirillo
09:15	Objectives of the meeting	Alberto Matteelli
09:30	LTBI conception: Definitions and relevance for diagnostic products	Frank Cobelens
10.15	Discussion	All
11.00	**COFFEE BREAK**	

SESSION 2: Target Product Profile for a test of progression of LTBI to active disease
Chairs: Daniela Cirillo and Delia Goletti

11:15	Preliminary results of the online consultation on the draft TPP	Susanna Capone
11:30	TTP for a test of progression: Key areas to be addressed based on survey results	Samuel Schumacher
11:45	Discussion	All
12:45	Next steps for the completion and dissemination of the TPP	Christopher Gilpin
13:00	**LUNCH**	

SESSION 3: Framework for validation of candidate products
Chairs: Samuel Schumacher and Helen Ayles

13:30	Framework for the evaluation of new LTBI tests	Sandra Kik
14:30	Discussion	All
15:30	Next steps to finalize and disseminate the guidance document	Frank Cobelens
16:00	**COFFEE BREAK**	

SESSION 4: LTBI diagnostics pipelines and evidence on new candidate tests
Chairs: Christian Lienhardt and Gavin Churchyard

16:15	LTBI diagnostics pipeline including biomarkers	Gavin Churchyard
16:30	Genomic signatures	Thomas Scriba (remote)
16:45	Role of CD8 in progression from latent to active TB	David Lewinsohn (remote)
17:00	Wrap-up and conclusive remarks	Daniela Cirillo and Alberto Matteelli
17:15	**Close of the meeting**	

ANNEX 2. LIST OF PARTICIPANTS

NDWG LTBI TASK FORCE

Alberto Matteelli, Coordinator
University of Brescia
Italy

Susanna Capone
University of Brescia
Italy

Gavin Churchyard
Aurum Institute
South Africa

Daniela Cirillo, NDWG Co-Chair
San Raffaele Research Institute
Italy

Frank Cobelens
AIGHD and KNCV
The Netherlands

Christopher Gilpin
WHO Global TB Programme
Switzerland

Delia Goletti
National Institute for Infectious Diseases
Italy

Amita Gupta (attending remotely)
Johns Hopkins School of Public Health
USA

Sandra Kik
KNCV
The Netherlands

Samuel Schumacher
FIND
Switzerland

Alessandra Varga, NDWG Secretariat
FIND
Switzerland

EXPERTS

Helen Ayles
LSHTM / Zambart
UK / Zambia

Hanif Esmail
Oxford University Hospital
UK

David Lewinsohn (attending remotely)
University of Oregon
Chair, Working Group on New Vaccines
USA

Christian Lienhardt
WHO Global TB Programme
Switzerland

Anna Mandalakas
Baylor College of Medicine
USA

Thomas Scriba (attending remotely)
University of Cape Town
South Africa

INDUSTRY REPRESENTATIVES

William Cruikshank
Oxford Immunotec
UK

Ludwig Deml
Lophius Biosciences GmbH
Germany

Paola Legnani
QIAGEN
UK

Oksana Markova
Generium
Russia

Morten Ruhwald
Statens Serum Institut
Denmark

ANNEX 3. QUESTIONS AND RESULTS OF THE WEB-BASED SURVEY ON TPP FOR LTBI TEST OF PROGRESSION – MAY 2016

Questions	Optimal target	Minimal target	Answers *		Note
			% agree on optimal	% agree on minimal	
Q1: Goal of intended use	Biomarker-based test that can be used to predict the risk of progression from TB infection (TBI) to active TB within the next two years, with the ability to rule out active TB. Ideally, the test result should decrease or revert to negative with treatment and thus enable an assessment of treatment success or cure and, consequentially, reinfection.	Biomarker-based test that can be used to predict the risk of progression from TBI to active TB within the next two years. The test would likely be positive in patients with active TB; therefore the presence of active TB needs to be ruled out by another highly sensitive test for active TB.	88	84	
Q2: Type of test	Single or multiple biomarker-based test, providing quantitative results that correlate with the risk of progression as well as qualitative results (positive/negative).	Single or multiple biomarker-based qualitative test (positive/negative).	93	87	
Q3: Target user of the test	Health care workers with no or minimal laboratory training (e.g. nurses).	Health care workers with laboratory training (e.g. skilled laboratory technicians).	84	79	
Q4: Diagnostic sensitivity for progression to active TB	≥ 90%	≥ 75%	92	78	15% of the respondents disagreed on minimal and the remaining 7% neither agreed nor disagreed.
Q5: Diagnostic specificity for risk of progression to active TB	≥90%.	≥ 75%	89	66	20% of the respondents disagreed on minimal and the remaining 14% neither agreed nor disagreed.

Questions	Optimal target	Minimal target	Answers *		Note
			% agree on optimal	% agree on minimal	
Q6: Results capturing, documen-tation, data display	Ideally, this should be an instrument-free test but should allow for attaching or scanning results 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 the instrument is used, the test menu should be simple, with integrated LCD screen, a key pad, or a touch screen.	87	85	
Q7: Training	< 1 day dedicated training for non-laboratory trained health-personnel.	Three to seven days dedicated training for laboratory-trained health-personnel.	87	78	
Q8: Number of steps to be perfor-med by the operator	<2, no timed steps.	<10, one to two timed steps	87	78	
Q9: Cost of equipment	<500 USD	<5000 USD	84	58	27% of the respondents disagreed on minimal and the remaining 15% neither agreed nor disagreed.
Q10: Cost of consu-mables (reagents/ test strips)	< 5 USD/test	<150 USD	82	28	48% of the respondents disagreed on minimal and the remaining 24% neither agreed nor disagreed.

*Answers were given according to the Likert scale (from 1 to 5):

1= disagree
2= somewhat disagree
3= neither agree nor disagree
4= mostly agree
5= fully agree

Disagreement was defined as the sum of 1+2 whereas agreement was defined as the sum of 4+5.

ANNEX 4. MAJOR COMMENTS ORIGINATED BY THE WEB-BASED SURVEY ON TPP

Questions	Optimal target
Q1: Goal of intended use	The biomarkers tests should be positive only for LTBI The test might not be able to differentiate from active TB: too ambitious and probably not feasible Risk assessment on progression for two years time might be too long Risk assessment should ideally define the infection's clearance and therefore be life-time The test might not be easily applied in HIV-infected patients because of HIV interference with TB progression
Q2: Type of test	A qualitative test will probably not be sufficient in view of the spectrum of TB infection A quantitative test will require a lot of data to be validated A qualitative result is not necessary and depends on the immune system of the subject
Q3: Target user of the test	The ideal test should be point of care (POC), community-based, bedside, and laboratory independent Target user should be the patient/candidate themselves The interpretation needs health care workers with good training
Q4: Diagnostic sensitivity for progression to active TB	Both 75% and 90% sound low as minimal/optimal: optimal sensitivity should be at least 95%: missing 1 in 10 is not going to be an acceptable risk; the minimal standard is too low – at least 85% The IGRAs set a very low bar to improve on: anything better than 25% would be a major advance PPV and NPV need to be considered as well, based on expected prevalence, and what the treatment decision implications would be with a positive test
Q5: Diagnostic specificity for risk of progression to active TB	The test should maximize true negatives: 95% specificity optimal, 90% minimal The test should be of high specificity to rule out BCG and NTM infections Sensibility and specificity should be presented to have sufficient high PPV
Q6: Results capturing, documentation, data display	The test should be associated with a mobile-phone application based reader; results should be transmitted wireless Data need to be uploaded and captured; instrument-free/paper results are likely to be lost to analysis Scanning /attaching results is not important as point of care test; data could be recorded in patient notes or logged on a computer
Q7: Training	Three to seven days length of training is too much: <2 days minimal, <4 hours optimal The more complex the test is, the more detailed training is required – i.e. a minimum of one to three days With an automated test, there's little reason to keep staff away from work for such a long time; training can be reduced
Q8: Number of steps to be performed by the operator	Laboratory is at heavy risk to end up as Ziehl-Neelsen method where the intra and inter operator performance and results vary greatly because of multiple steps If the test meets all the other requirements but is of high complexity, the number of steps becomes irrelevant Current tests such as TST and IGRA have many steps: optimal is too ambitious A test with fewer steps is more user-friendly and has a higher probability to be performed correctly

Questions	Optimal target
Q9: Cost of equipment	Must be cheap and affordable The optimal cost of equipment would be $0 if the test is instrument-free Minimal costs depend on the complexity of the test; a wide range of costs is acceptable if a complex test meets all the other requirements Affected communities cannot afford high cost determining low acquisition /under utilization and, consequently, no impact on the global LTBI burden
Q10: Cost of consumables (reagents/test strips)	The cost of the test should be equal to or less than smear microscopy: < $1 optimal /< $10 minimal A test at <$US150 would not be a research tool: $US20 could be negotiated for LMIC but a much higher cost will be applied to HIC <$US5 /test is unrealistic whereas $US150 /test is unaffordable; range suggested: $US30 / test to $US100 /test

Stop TB Partnership

New Diagnostics Working Group

www.stoptb.org