New and improved tuberculosis diagnostics: evidence, policy, practice, and impact

Madhukar Pai[a], Jessica Minion[a], Karen Steingart[b] and Andrew Ramsay[c]

[a]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada, [b]Francis J. Curry National Tuberculosis Center, University of California, San Francisco, California, USA and [c]UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), World Health Organization, Geneva, Switzerland

Correspondence to Madhukar Pai, MD, PhD, Assistant Professor, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, QC H3A 1A2, Canada
Tel: +1 514 398 5422; fax: +1 514 398 4503;
e-mail: madhukar.pai@mcgill.ca

Current Opinion in Pulmonary Medicine 2010, 16:271–284

Purpose of review

The aim is to summarize the evidence base for tuberculosis (TB) diagnostics, review recent policies on TB diagnostics, and discuss issues such as how evidence is translated into policy, limitations of the existing evidence base, and challenges involved in translating policies into impact.

Recent findings

Case detection continues to be a major obstacle to global TB control. Fortunately, due to an unprecedented level of interest, funding, and activity, the new diagnostics pipeline for TB has rapidly expanded. There have been several new policies and guidelines on TB diagnostics. However, there are major gaps in the existing pipeline (e.g. lack of a point-of-care test) and the evidence base is predominantly made up of research studies of test accuracy.

Summary

With the availability of new diagnostics and supporting policies, the next major step is translation of policy into practice. The impact of new tests will depend largely on the extent of their introduction and acceptance into the global public sector. This will itself depend in part on policy decisions by international technical agencies and national TB programs. With the engagement of all key stakeholders, we will need to translate evidence-based policies into epidemiological and public health impact.

Keywords

diagnostics, evidence, impact, policy, tuberculosis

Curr Opin Pulm Med 16:271–284

1070-5287

Introduction

In 2010, poor diagnosis remains a major obstacle to global tuberculosis (TB) control. In most high-burden countries, TB is still diagnosed using tools such as direct sputum microscopy and chest radiographs. Fortunately, the past few years have seen an unprecedented level of interest, funding support, and activity focused on the development of new tools for TB diagnosis, and the new diagnostics pipeline for TB is rapidly expanding. In parallel, there have been several new policy recommendations on TB diagnostics by the WHO. Because recent publications [1•,2,3•,4] have exhaustively reviewed the current pipeline of new diagnostics and the expanding evidence base for their use, we focus our attention on how evidence is translated into policy, limitations of the existing evidence base, deficiencies in the current diagnostics pipeline, and challenges involved in translating policies into practice and impact.

What is the evidence base for tuberculosis diagnostics?

The evidence base for TB diagnostics is ultimately derived from a large body of original research. Because individual studies are seldom sufficient to inform policy and guideline development, the totality of available evidence must be synthesized. Thus, systematic reviews and meta-analyses are often necessary to summarize the evidence on a given diagnostic test. In the past decade, there have been over 35 systematic reviews published on TB diagnostics, on topics ranging from smear microscopy to molecular diagnostics and in-vitro assays for latent TB infection (LTBI). All of these systematic reviews have been made available on a new website 'Evidence-based Tuberculosis Diagnosis' (www.tbevidence.org) compiled by the Stop TB Partnership's New Diagnostics Working Group, in collaboration with several agencies [5•]. While the key findings of published systematic reviews and meta-analyses on TB diagnostics have been reviewed elsewhere [6•], Table 1 provides a brief overview of the evidence base for TB diagnosis, essentially synthesizing the evidence from several systematic reviews [7–37].

What is lacking in current evidence base?

Although a large number of systematic reviews have been published on TB diagnostics, almost all focus on test accuracy (i.e. sensitivity and specificity). This is in part

Table 1 Summary of findings from systematic reviews on tuberculosis diagnostic tests

Diagnostic test	Description	Disease/site	Major findings/results of systematic reviews	Major references
Diagnosis of active TB				
Sputum smear microscopy	Microscopic observation of stained acid fast bacilli	Pulmonary TB	Fluorescence microscopy (FM) is on average 10% more sensitive than conventional microscopy. Specificity of both fluorescence and conventional microscopy is similar. Fluorescent microscopy is associated with improved time efficiency. LED FM performs equivalently to conventional FM, with added benefits of low cost, durability, and ability to use without a darkroom. Centrifugation and overnight sedimentation, preceded with any of several chemical methods (including bleach) is slightly more sensitive (6–9%) than direct microscopy; specificity may be slightly decreased (1–3%) by sputum processing methods. When serial sputum specimens are examined, the mean incremental yield and/or increase in sensitivity from examination of third sputum specimen ranges between 2 and 5%.	[7–9]
Nucleic acid amplification tests (NAATs)	Isolation, replication, and detection of nucleic acid sequences specific for *Mycobacterium tuberculosis*	Pulmonary and extrapulmonary TB	NAATs have high specificity and positive predictive value. NAATs, however, have relatively lower (and highly variable) sensitivity and negative predictive value for all forms of TB, especially in smear-negative and extrapulmonary disease. In-house ('home brew') NAATs produce highly inconsistent results as compared with commercial, standardized NAATs.	[10–15]
Commercial serological antibody detection tests	Detection of host antibody response to *Mycobacterium tuberculosis* antigens	Pulmonary and extrapulmonary TB	Serological tests for both pulmonary and extrapulmonary TB produce highly inconsistent estimates of sensitivity and specificity; none of the current assays perform well enough to replace microscopy.	[12,16,17]
Noncommercial (in-house) serological antibody detection tests	Detection of host antibody response to *Mycobacterium tuberculosis* antigens	Pulmonary TB	Several potential candidate antigens for inclusion in an antibody detection-based diagnostic test for pulmonary TB in HIV-infected and -uninfected individuals were identified. Combinations of select antigens provide higher sensitivities than single antigens.	[18]
Adenosine deaminase (ADA)	Detection of host cellular enzyme released by lymphocytes in response to live intracellular pathogens	TB pleuritis, pericarditis, peritonitis	Measurement of ADA levels in pleural, pericardial, and ascitic fluid is a useful adjunct test for TB pleuritis, pericarditis, and peritonitis. Systematic reviews have reported pooled sensitivity estimates between 88 and 100%, and specificity estimates between 83 and 97%.	[19,20]
Interferon-gamma (IFN-γ)	Measurement of IFN-γ	TB pleuritis	Pleural fluid IFN-γ determination appears to be a useful diagnostic for TB pleuritis, with systematic reviews reporting pooled sensitivity estimates between 89 and 96%, and specificity estimates between 96 and 97%.	[19,21]
Phage amplification assays	Detection of *Mycobacterium tuberculosis*-specific phage viruses, after their infection and amplification of live MTB	Pulmonary TB	Despite high-accuracy estimates, current phage-based assays are limited by high rates of indeterminate results (up to 36%).	[22]
Automated liquid cultures	Automated detection of changes in oxygen, carbon dioxide, or pressure resulting from bacterial growth	Pulmonary TB	Automated liquid cultures are more sensitive than solid cultures; time to detection is more rapid than solid cultures.	[12,23]
Diagnosis of latent TB				
Tuberculin skin test (TST)	Measurement of induration as a result of exposure to intradermal tuberculin	Latent TB infection	Individuals who have received BCG vaccination are more likely to have a positive TST; the effect of BCG on TST results is less after 15 years; positive TST with indurations of >15 mm are more likely to be the result of TB infection than of BCG vaccination. The effect on TST of BCG received in infancy is minimal, especially 10 years after vaccination. BCG received after infancy produces more frequent, more persistent, and larger TST reactions. Nontuberculous mycobacterial (NTM) infection is not a clinically important cause of false-positive TST, except in populations with a high prevalence of NTM sensitization and a very low prevalence of TB infection.	[24,25]
T-cell-based interferon-γ release assays (IGRAs)	Measurement of IFN-γ released from lymphocytes when stimulated by *Mycobacterium tuberculosis*-specific antigens	Latent TB infection	IGRAs have excellent specificity (higher than the tuberculin skin test) and are unaffected by prior BCG vaccination. IGRAs cannot distinguish between LTBI and active TB and have no role for active TB diagnosis in adults. Used as an adjunctive diagnostic, IGRAs may aid in the investigation of pediatric TB. IGRAs correlate well with markers of TB exposure in low-incidence countries. IGRA performance appears to differ in high-endemic vs. low-endemic countries. IGRA sensitivity varies across populations and tends to be lower in high-endemic countries and in HIV-infected individuals.	[12,26–29]

Diagnosis of drug resistance				
Phage amplification assays	Detection of *Mycobacterium tuberculosis*-specific phage viruses, after their infection and amplification of live MTB + inhibition of growth in presence of antituberculous drugs	Rapid detection of rifampicin resistance	When used on culture isolates, phage assays have high sensitivity, but variable and lower specificity. In contrast, evidence is lacking on the accuracy of these assays when they are directly applied to sputum specimens. Recent studies have raised concerns about contamination, false-positive results, and technical assay failures.	[30,31]
Line probe assays: INNO-LiPA Rif. TB [LiPA] and GenoType MTBDR assay	Detection of genetic sequences associated with resistance (after extraction and amplification) using immobilized probes and colorimetric development	Rapid detection of rifampicin resistance	LiPA is a highly sensitive and specific test for the detection of rifampicin resistance in culture isolates. The test has relatively lower sensitivity when used directly on clinical specimens. The GenoType MTBDR assays have excellent sensitivity and specificity for rifampicin resistance even when directly used on clinical specimens.	[32–34]
Colorimetric redox indicators (CRIs)	Determination of MIC using microdilution, followed by addition of reagent which will become reduced in the presence of actively growing MTB resulting in a color change	Rapid detection of rifampicin and isoniazid resistance	Colorimetric methods are sensitive and specific for the detection of rifampicin and isoniazid resistance in culture isolates. CRIs use inexpensive noncommercial supplies and equipment and have a rapid turnaround time (7 days).	[35]
Nitrate reductase assays (NRAs)	Direct or indirect inoculation of drug-free and drug-containing media containing KNO_3. Addition of Greiss reagent detects early growth by reacting with enzymatic byproduct and resulting in a color change.	Rapid detection of rifampicin and isoniazid resistance	NRA has high accuracy when used to detect rifampicin and isoniazid resistance in culture isolates. Limited data are available on its use when directly applied to clinical specimens, but results are promising. The NRA is simple, uses inexpensive noncommercial supplies and equipment, and has a rapid turnaround time (7–14 days) compared to conventional methods.	[36]
Microscopic observation drug susceptibility (MODS)	Direct or indirect inoculation of drug-free and drug-containing liquid media, followed by examination using an inverted microscope to detect early growth	Rapid detection of rifampicin and isoniazid resistance	MODS has high accuracy when testing for rifampicin resistance, but shows slightly lower sensitivity when detecting isoniazid resistance. MODS appears to perform equally well using direct patient specimens and culture isolates. MODS uses noncommercial supplies and equipment, and has a rapid turnaround time (10 days) compared with conventional methods.	[37]
Thin layer agar (TLA)	Direct or indirect inoculation of drug-free and drug-containing solid media, followed by examination using a microscope to detect early growth	Rapid detection of rifampicin and isoniazid resistance	There is a paucity of data evaluating TLA for the detection of drug susceptibility; however, all studies to date have found 100% concordance with their reference standards. TLA uses inexpensive noncommercial supplies and equipment, and has a rapid turnaround time (11 days) compared with conventional methods.	[37]

BCG, bacillus Calmette-Guérin; LED, light emitting diode; LTBI, latent TB infection; MIC, minimal inhibitory concentration; MTB, *Mycobacterium* tuberculosis; TB, tuberculosis. Adapted from [6•]. (Open Access under Creative Commons Attribution License).

because a large proportion of TB diagnostic research studies are focused on measuring test accuracy. Findings from systematic reviews suggest that even relatively straightforward studies of test accuracy are often poorly designed and reported [38,39]. Both researchers of primary TB diagnostic studies and authors of systematic reviews and meta-analyses need to make efforts to follow published guidelines for conducting and reporting their work [40,41], to make the most of their contribution to a useful and unbiased literature base.

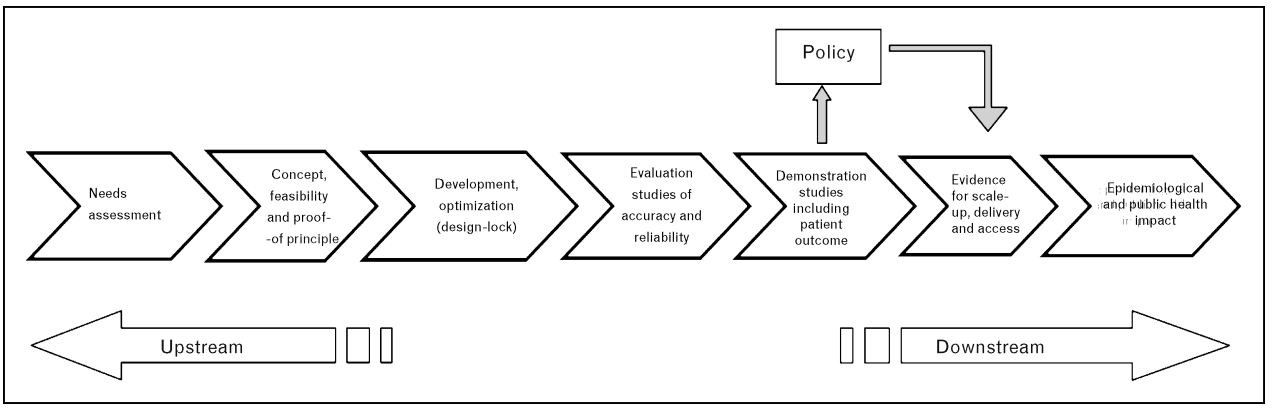
Although the quality of diagnostic studies measuring test accuracy is important, evidence about test accuracy is only one link in a long chain of activities that make up the pathway to developing and implementing a new TB diagnostic. In 2009, the Stop TB Partnership's New Diagnostics Working Group published a scientific blueprint for development of new TB diagnostics [42••]. This publication provides a comprehensive, well referenced plan to guide researchers, clinicians, industry partners, academics, and TB controllers in all sectors in all aspects of TB diagnostics development [42••], starting from needs' assessment, concept, feasibility, proof-of-concept, to test development, validation, and, ultimately, delivery, scale-up, access, and epidemiological and public health impact.

As shown in Fig. 1, evidence on test accuracy is essential, but policy development requires more than estimation of test accuracy. Along with data on test accuracy, we need to consider user-important as well as patient-important outcomes. Patient-important outcomes require more sophisticated and often more resource-intensive research [43,44], wherein a study shows that implementing a diagnostic test in a given situation results in clinically relevant improvements in patient care and/or patient outcomes. For TB diagnostics, this might mean an increased number of patients detected and receiving appropriate treatment, fewer patients defaulting from the diagnostic pathway due to reduced number of patient visits, or more patients cured due to accurate detection of drug resistance. Studies may also investigate the values and preferences patients have when choosing one diagnostic test compared to another. Although the challenges and costs of demonstrating these types of outcomes make them unattractive for many researchers and funding agencies, it is no less important than proving a therapeutic intervention actually changes the course of a disease and not just the level of a biomarker or surrogate endpoint.

User-important outcomes consist of practical concerns for the usability of a test in real-world situations. Although these generally do not require fundamentally different strategies to evaluate, it is important that they are assessed under implementation or real-world settings. These include the ease of use of a technology, the hands-on time of performing the test, the expertise or training required, and the infrastructure needed. It is important to consider biosafety, robustness of any equipment involved, as well as pragmatic issues such as the shelf-life of reagents, the need for special shipping or storage of materials, the availability and reliability of supply chains, and of course cost.

These types of evidence must be taken into account, along with test accuracy and reliability, when policy makers or programs are evaluating a diagnostic for recommendation or widespread use. Systematic reviews of diagnostics should make an effort to summarize data on these outcomes in addition to accuracy, appraise the quality of available evidence, and explore the uncertainty regarding the often assumed values and preferences of patients associated with these tests. However,

Figure 1 Level of evidence required for policy process



Adapted from [42••].

an obstacle here is a lack of methodology for collecting and analyzing such evidence even if the data were reported in primary research. In other words, currently used systematic review methods are mainly aimed at test accuracy.

Where is the current diagnostics pipeline deficient?

Although there are many more TB diagnostics in the pipeline today than in the past, the existing TB diagnostics pipeline itself has limitations and neglects some important aspects of the TB epidemic. Table 2 summarizes the major research priorities for TB diagnostics.

The biggest concern continues to be the lack of a rapid, simple, inexpensive, point-of-care (POC) test for active TB. As yet nothing has emerged from the pipeline or looks likely to emerge from the pipeline in the near future that could supplant smear microscopy. An easy-to-use, inexpensive diagnostic that can perform as well or better than smear microscopy and can deliver results within minutes without sophisticated equipment or highly-trained laboratory personnel would be a major step forward in TB diagnostics and could have a tremendous impact on global TB control [45,46•].

Another area still lacking in adequate diagnostic options is smear-negative TB, especially in HIV-infected persons [47]. Undiagnosed TB is very common in persons infected with HIV; therefore, intensive active case finding is required as strategies that rely on passive detection, or screening with smear microscopy alone, will miss a large number of coinfected patients [47]. Considering the proven benefit of TB preventive therapy using isoniazid in HIV-infected persons, ruling out active TB before initiation of single drug treatment is important not only for the care of the individual patient, but also to prevent the inadvertent selection for drug resistance. The development and validation of an algorithm, taking advantage of newly available tests, to aggressively target this high-risk population remains a priority for TB control.

Childhood TB presents similar challenges [48]. By virtue of the pathophysiology of TB in pediatrics and the inability to obtain adequate sputum samples, microbiologic confirmation of active TB remains an insensitive and inadequate standard. Similar to patients with HIV and smear-negative TB, the development and improvement in diagnostic algorithms that take advantage of available new diagnostics is needed. As good quality sputum specimens are difficult to collect, novel diagnostics that can be used on urine, saliva, breath condensate, and so on could have a greater impact in these populations, especially if a POC format could be developed.

The control of drug-resistant TB requires accurate and rapid diagnostics for the detection of critical patterns of drug resistance. The need to identify cases of multidrug-resistant TB (MDR-TB) through detecting resistance to rifampicin and isoniazid is now well recognized. The next step is to accurately and rapidly identify cases of extensively drug-resistant TB (XDR-TB) through the detection of resistance to key second-line drugs.

Although new tests [such as interferon-gamma release assays (IGRAs)] have emerged for LTBI diagnosis, these tests cannot resolve the various phases of the latent TB spectrum [49,50]. This means existing tests cannot be used to target preventive therapy at the subgroup that is most likely to benefit from treatment. Thus, there is a need for a highly predictive biomarker or combination of biomarkers, which will allow accurate prediction of the subgroup of latently infected individuals who are at highest risk of progression to disease.

How is evidence translated into policy?

The WHO has taken the lead on developing policies and guidelines on TB diagnostics. The WHO policy process is summarized in a recent statement entitled 'Moving research findings into new WHO policies [51•].' The key steps in the WHO policy process are given in Table 3 [51•]. This process takes into account the importance of not only identifying areas in need of policy guidance, but also ensuring that policies are evidence-based and then followed up with dissemination and promotion of new recommendations. For step 2, reviewing the evidence, WHO may commission a systematic review and meta-analysis of available data (published and unpublished) using standard methods appropriate for diagnostic accuracy studies [52•].

Table 4 provides an overview of all the recent WHO policies on TB diagnostics [51•,53–57]. Since 2007, the WHO has endorsed several diagnostic tests and strategies, including liquid cultures, optimized smear microscopy, line probe assays, and noncommercial culture systems for drug-susceptibility testing.

The foundation of the WHO policy process is now the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach [58••]. This is in part a response to the criticism that systematic reviews are rarely used for developing WHO recommendations and that WHO policy processes usually rely heavily on expert opinion [59]. The GRADE approach provides a system for rating the quality of evidence and the strength of recommendations that is explicit, comprehensive, transparent, and pragmatic and is being adopted increasingly by organizations worldwide [58••,60]. The WHO now requires the use of GRADE for

Table 2 Major priorities for research and development and implementation of tuberculosis diagnostics

Research priority	Research methods	Expected outcome	Justification
Development of a rapid, accurate point-of-care (POC) test for pulmonary TB.	Biomarker discovery, followed by incorporation in a highly sensitive POC platform and then clinical validation.	A POC test for pulmonary TB that will meet the user-defined specifications (such as those proposed by MSF).	Currently, there is no POC test for TB that can be used at the health clinic level. Diagnostic delays, therefore, are common.
Development and validation of tools for rapid detection of drug resistance, including for XDR-TB and standardization of DST for second-line drugs.	Identification and characterization of mutations associated with second-line drug resistance; development of newer generation molecular assays for MDR/XDR-TB; improved standardization of existing tests for second-line DST.	Rapid molecular (genotypic) assays for MDR/XDR-TB that will allow rapid identification of drug-resistant TB.	Although line-probe assays are highly accurate for rifampicin resistance, accuracy is lower for isoniazid and other drugs. Second-line DST continues to be a challenge; mutations are not well defined and standardization is a problem with phenotypic methods.
Intensified, active case detection strategies for early detection of active TB in HIV-infected persons (at the clinic level and in the community).	Development and validation of an algorithm (including new tests) for rapid detection of TB in HIV-infected persons.	A validated algorithm that will enable detection of TB in a large proportion of HIV-infected persons with TB disease.	Passive case detection methods do not work well in areas with high HIV prevalence; undiagnosed TB is frequent in HIV-infected persons and can cause enormous morbidity and mortality. Aggressive case detection approaches are needed to enhance case detection, reduce mortality, and reduce transmission.
Improving current diagnostic algorithms to shorten the time required for establishing a diagnosis of smear-negative pulmonary TB and extrapulmonary TB in HIV-infected persons and children.	Development and validation of an algorithm (including new tests) for rapid detection of smear-negative and extrapulmonary TB in HIV-infected persons and children.	A validated algorithm that will rapidly enable detection of smear-negative and extrapulmonary TB in a large proportion of HIV-infected persons and children.	Smear-negative TB, extrapulmonary TB, and childhood TB are diagnostic challenges and available tests perform poorly in these cases of paucibacillary TB. Newer algorithms and tests are needed to get around the limitations of current methods.
Development of a test or algorithm that can accurately rule out active TB disease in HIV-infected persons [to allow initiation of preventive therapy (IPT)]	Development and validation of an algorithm (including new tests) for rapidly ruling out active TB (including smear-negative and extrapulmonary TB) in HIV-infected persons.	A validated algorithm that will enable exclusion of TB in a large proportion of HIV-infected persons prior to IPT.	In HIV-infected persons, undiagnosed active TB is common. Before IPT, it is necessary to rule out active TB. However, there is no easy and accurate method to do this in high-burden countries.
Which biomarkers or combinations of markers will help distinguish the various stages of the spectrum of latent TB infection (from sterilizing immunity to subclinical active disease) and will allow accurate prediction of the subgroup of latently infected individuals who are at highest risk of progression to disease.	Biomarker discovery, followed by validation in clinical and longitudinal (cohort) studies for markers that can predict risk of progression to active TB.	Identification of a biomarker or combination of biomarkers that will allow accurate prediction of the subgroup of latently infected individuals who are at highest risk of progression to disease.	Existing tests for latent TB (TST and IGRAs) cannot resolve the various phases of the latent TB spectrum. This means existing tests cannot be used to target IPT at the subgroup that is most likely to benefit from treatment. This results in overtreatment of a large number of latently infected persons.
Development of a rapid test for childhood TB that will not depend on sputum specimen testing.	Development and validation of a test or an algorithm (including new tests) for rapid detection of TB in children, without requiring sputum specimens.	A test (preferably POC) that can use nonsputum specimens (e.g. urine or breath condensate or saliva) for rapid detection of TB in children.	Childhood TB is a diagnostic challenge and available tests perform poorly in these cases of paucibacillary TB. Also, as young children are unable to produce sputum, it will be helpful to use alternative specimens such as urine, saliva, or breath condensate.
Define different ways of operationalizing and implementing existing policies on HIV testing of TB patients and TB screening of HIV-infected persons.	Operational research on different ways of implementing existing policies on HIV testing of TB patients and TB screening of HIV-infected persons.	Identification of at least one feasible approach that might work best and therefore can be scaled-up.	Existing policies on HIV testing of TB patients and TB screening of HIV-infected persons are poorly implemented. A large proportion of TB patients are not tested for HIV, and HIV-infected persons are not screened for TB. This results in undiagnosed co-infection morbidity/mortality and continued transmission in the community.
Once new diagnostics are approved and available, what factors can enhance their actual delivery and implementation at the programmatic level in high-burden countries?	Operational research on different ways of implementing new diagnostics in national TB programs in high-burden settings.	Identification of at least one feasible implementation approach that might work best and therefore can be scaled-up.	Availability of new tools does not necessarily ensure their adoption and implementation. Translation of policy into practice requires better understanding of barriers to implementation and methods to overcome such barriers.
What is the likely epidemiological impact of widespread LTBI diagnosis and treatment in high-burden countries, and what contribution will LTBI diagnosis and treatment make toward the attainment of the Stop TB Partnership's target for TB elimination?	Mathematical modeling study.	The modeling study will inform the debate on when high-burden countries should begin to focus attention on LTBI diagnosis and treatment.	LTBI diagnosis and treatment is currently not a priority in high-burden countries. However, as TB incidence falls, it can become a priority. Also, some recent modeling studies suggest that TB elimination will require strategies aimed at LTBI management.

DST, drug susceptibility testing; LTBI, latent tuberculosis infection; MDR, multidrug-resistant; MSF, Médecins Sans Frontières; XDR, extensively drug-resistant.

Table 3 World Health Organization policy process for tuberculosis

Major steps	Description of the process
Identifying the need for a policy change	The need to formulate new or revised policies may arise from WHO's ongoing monitoring of technical developments or from interested parties submitting requests and supporting documentation for policy or guideline development. WHO receives information about a new technology or approach via many channels, with the most direct lines coming from national TB programs and researchers themselves. To consider a global policy change, WHO must have solid evidence, including clinical trials or field evaluations in high-TB prevalence settings.
Reviewing the evidence (including systematic reviews)	WHO may carry out or commission a review of the documentation of technology's clinical or programmatic performance, including newly published and 'grey' research or reviews, 'proof-of-principle' reports, large-scale field trials, and demonstration projects in different resource settings. Standardized evaluation criteria have been and are being developed by the New Diagnostics, New Drugs, and New Vaccines Working Groups of the Stop TB Partnership.
Convening an expert panel	If the evidence base is compelling, WHO will convene an external panel of experts, excluding all original principal investigators from the studies. The panel will review the evidence (using the GRADE approach) and make a recommendation or propose draft policies or guidelines to WHO's Strategic and Technical Advisory Group for Tuberculosis (STAG-TB).
Assessing draft policies and guidelines	STAG-TB provides objective, ongoing technical, and strategic advice to WHO related to TB care and control. STAG-TB's objectives are to provide the Director-General, through the Stop TB Department, an independent evaluation of the strategic, scientific, and technical aspects of WHO's TB activities, review progress and challenges in WHO's TB-related core functions, review and make recommendations on committees and working groups, and make recommendations on WHO's TB activity priorities. STAG-TB reviews the policy drafts and supporting documentation during its annual meeting. STAG-TB may endorse the policy recommendation with or without revisions, request additional information and re-review the evidence in subsequent years, or reject the recommendation.
Formulating and disseminating policy	New WHO policies and guidelines will be disseminated through different channels to Member States, including through the World Health Assembly, WHO website, list serves, and journal publications. WHO also disseminates its recommendations to other agencies and donors engaged in TB control activities.

GRADE, Grading of Recommendations Assessment, Development and Evaluation; TB, tuberculosis. World Health Organization: moving research findings into new WHO policies [51•].

all new and revised WHO policies and guidelines, including policies on diagnostics [61]. For example, recent WHO policies on TB infection control [62] and the revised TB treatment guidelines [63] used the GRADE approach.

Grading of Recommendations Assessment, Development, and Evaluation for diagnostic tests: strengths and limitations

The GRADE approach provides a clear separation of quality of evidence and strength of recommendations [58••]. In judgments about quality of evidence, GRADE considers six factors: study design, methodological quality, directness of evidence (patient-important outcomes and generalizability), inconsistency of results, imprecision of results (imprecise or sparse data), and publication bias [58••]. Thus, quality of evidence reflects our confidence that estimates of benefits and downsides from a diagnostic test or strategy generated from research are correct. Quality of evidence is graded as follows:

(1) High quality: further research is very unlikely to change our confidence in the estimate of effect.
(2) Moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
(3) Low quality: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
(4) Very low quality: any estimate of effect is very uncertain.

In the GRADE approach, well designed studies of diagnostic accuracy (cross-sectional or cohort studies on patients with diagnostic uncertainty and use of appropriate reference standard) can provide high-quality evidence on test accuracy. However, these studies may provide only low-quality evidence for guideline development because of uncertainty about the link between test accuracy and outcomes important to patients (discussed below).

The strength of a recommendation refers to the extent to which one can be confident that adherence to the recommendation will do more good than harm [58••]. There are four factors to consider: balance between desirable and undesirable effects; quality of evidence; values and preferences; and costs (resource allocation). GRADE classifies recommendations as strong (most informed patients would choose this option) or weak (patients' choices will vary according to their values and preferences and not all patients would choose this option).

The GRADE process was initially developed for treatment interventions and, therefore, tends to be focused on

Table 4 Highlights of recent WHO policies and statements on tuberculosis diagnostics

Year policy was made	Purpose of testing	Diagnostic test or approach	WHO recommendations
2007	Case detection and drug-susceptibility testing (DST)	Liquid media for culture and DST	WHO recommends, as a step-wise approach: The use of liquid medium for culture and DST in middle-income and low-income countries. The rapid species identification to address the needs for culture and DST. Taking into consideration that liquid systems will be implemented in a phased manner, integrated into a country-specific comprehensive plan for laboratory capacity strengthening.
2007	Case detection	Definition of a new sputum smear-positive TB case	The revised definition of a new sputum smear-positive pulmonary TB case is based on the presence of at least one acid fast bacilli (AFB+) in at least one sputum sample in countries with a well functioning external quality assurance (EQA) system.
2007	Case detection	Reduction of number of smears for the diagnosis of pulmonary TB	WHO recommends the number of specimens to be examined for screening of TB cases can be reduced from three to two, in places where a well functioning EQA system exists, where the workload is very high and human resources are limited.
2008	DST	Molecular line probe assays for rapid screening of patients at risk of MDR-TB	The use of line probe assays is recommended by WHO, with the following guiding principles: Adoption of line probe assays for rapid detection of MDR-TB should be decided by Ministries of Health within the context of country plans for appropriate management of MDR-TB patients, including the development of country-specific screening algorithms and timely access to quality-assured second-line anti-TB drugs. Line probe assay performance characteristics have been adequately validated in direct testing of sputum smear-positive specimens and on isolates of *Mycobacterium tuberculosis* complex grown from smear-negative and smear-positive specimens. Direct use of line probe assays on smear-negative clinical specimens is not recommended. The use of commercial line probe assays, rather than in-house assays, is recommended to ensure reliability and reproducibility of results, as in-house assays have not been adequately validated or used outside limited research settings. Adoption of line probe assays does not eliminate the need for conventional culture and DST capability; culture remains necessary for definitive diagnosis of TB in smear-negative patients, whereas conventional DST is required to diagnose extensively drug-resistant TB (XDR-TB). As current line probe assays only detect resistance to rifampicin and/or isoniazid, countries with documented or suspected cases of XDR-TB should establish or expand conventional culture and DST capacity for quality-assured susceptibility testing of second-line drugs, based on current WHO policy guidance.
2009	Case detection	LED-based microscopy	WHO recommends that conventional fluorescence microscopy be replaced by LED microscopy in all settings and that LED microscopy be phased in as an alternative for conventional ZN microscopy in both high-volume and low-volume laboratories. The switch to LED microscopy should be carried out through a carefully phased implementation plan, using LED technologies that meet WHO specifications.
2009	DST	Noncommercial culture and DST methods	WHO recommends that selected noncommercial culture and DST methods be used as an interim solution in resource-constrained settings, in reference laboratories, or those with sufficient culture capacity, while capacity for genotypic and/or automated liquid culture and DST are being developed. With due consideration of the above issues, WHO endorses the selective use of one or more of the following noncommercial culture and DST methods: Microscopically observed drug susceptibility (MODS), for rapid screening of patients suspected of having MDR-TB, under clearly defined programmatic and operation conditions, and once speciation concerns have been adequately addressed without compromising bio-safety; The nitrate reductase assay (NRA), for screening of patients suspected of having MDR-TB, under clearly defined programmatic and operation conditions, and acknowledging that time to detection of MDR in indirect application would not be faster (but less expensive) than conventional DST methods using commercial liquid culture or line probe assays; Colorimetric redox indicator (CRI) methods, as indirect tests on *M. tuberculosis* isolates from patients suspected of having MDR-TB, under clearly defined programmatic and operation conditions, and acknowledging that time to detection of MDR would not be faster (but less expensive) than conventional DST methods using commercial liquid culture or line probe assays.

LED, light-emitting diode; MDR-TB, multidrug-resistant tuberculosis. From the World Health Organization [51•,53–56].

randomized controlled trials (RCTs). It has been adapted for diagnostic tests and strategies [64••,65], though this area is a work in progress and can be improved based on user's feedback. The first time the GRADE approach was applied to TB diagnostics by the WHO was in September 2009 for use in developing guidelines for improving sputum smear microscopy and using noncommercial culture methods for rapid detection of TB drug resistance. From these experiences, we have found the GRADE approach to have several strengths as well as some limitations.

On the positive side, GRADE offers a systematic, objective, and transparent process and requires the explicit use of systematic reviews and evidence summaries. GRADE forces us to consider several elements, including quality of evidence, cost, values and preferences, and trade-offs between good and bad consequences. One challenge in using GRADE is learning the process itself, as systematic reviewers, policy makers, and TB experts are not necessarily trained in the GRADE approach. We expect this challenge to be overcome as more people receive training and use GRADE. Another challenge recognizes situations in which patient outcomes may not reflect the accuracy or benefit of a diagnostic test/approach because treatment is unavailable (e.g. improved microscopy in facilities where stock-outs of anti-TB drugs occur frequently). Additional limitations and challenges for diagnostic policies are summarized in Table 5. A recent review by Kavanagh [66] provides an interesting perspective on GRADE, especially on the issue of whether GRADE itself is reliable and has been proven to be valid.

By the nature of the GRADE process being based on evidence, it is intrinsically reliant on the availability and quality of the evidence base itself. As we have discussed above, challenges remain to ensure both the quality of primary diagnostic evaluations and the availability of the necessary types of data in systematic reviews. This is brought into clear focus when using the GRADE process, as a lack of objective studies on a topic opens the door to the substitution of expert opinion for evidence. Although expert experiences cannot be discounted, they may often not be generalizable and are subject to being influenced by personal agendas and anecdotal experiences. Experts in TB often rate the same evidence inconsistently, depending on their prior experience with a test, and this can result in poor interrater agreement on GRADE elements. For example, TB researchers who work extensively in resource-poor settings are often skeptical of high-tech tools and tend to undervalue them because of the perceived limited applicability in developing countries.

Conflicts of interest (COI) among guideline panel members and industry involvement in guideline processes are other issues of concern, especially when commercial tests and products are involved. There is some evidence that industry involvement is fairly common with TB diagnostic research, with about 40–50% of TB diagnostic studies reporting some degree of industry involvement or support [26,39]. A recent survey of IGRA guidelines and statements from various countries found that only a small minority had explicit COI disclosures [67]. Some organizations have recognized the need to address the issue of COI. For example, the American Thoracic Society (ATS) published its COI policy for guideline development in 2009 [68]. This policy now recommends procedures such as self-declaration of COI; review of potential participants' COI; disclosure of COI to project participants; refusal or excusal from certain decisions or recommendations when appropriate; and disclosure of COI to users of documents or attendees of conferences. All agencies and bodies involved in guideline development should follow this example.

COI, however, are not restricted to commercial products. Diagnostic tests developers can be academics with no industry involvement. Because of their heavy intellectual investment in new test development and better understanding of the test, they tend to have strong opinions on how policies should be formulated and this can pose conflicts during the guideline development process. Should test developers be included in guideline panels, but excused from voting on recommendations? Sometimes, test developers publish systematic and narrative reviews on their own tests (which invariably tend to be positive) and it is unclear whether such reviews should be included or excluded in the GRADE process. Publication bias is an added concern, especially if industry-supported diagnostic studies are more likely to be published when they report positive findings. Unlike RCTs, inclusion of unpublished diagnostic studies is difficult because of the lack of a diagnostic trials registry.

The involvement of public–private partnerships for product development perhaps increases the complexity. These are often characterized by a partnership between a nonprofit organization and a for-profit diagnostics company with confidential agreements on intellectual property related to a co-developed diagnostic. Test developers from the nonprofit organization may have the same intellectual investment COI as test developers in academia, but may in addition have a COI related to their partnership with a for-profit company. These issues point out a fundamental problem with all guidelines, a problem that GRADE can never address – the fate of a guideline or policy can heavily rest on the group of experts and stakeholders included in the guideline development committee or panel.

The application of the GRADE approach to evidence on diagnostics is relatively new and as a result there are some

Table 5 Challenges and limitations in formulating tuberculosis diagnostic policies

Challenge or limitation	Description and examples
Limitations of the existing evidence base	Majority of TB diagnostic studies are focused on test accuracy (sensitivity and specificity); therefore, systematic reviews are also focused on accuracy Test accuracy studies are often poorly designed, executed, and reported. Impact of tests on patient-important outcomes is rarely available. Accuracy studies are downgraded by GRADE for 'directness' and can never receive a rating of 'high-quality' evidence. Ease of implementation, resources required, cost-effectiveness, biosafety, and programmatic issues are critical for policy, but systematic reviews may not provide such data.
Evidence vs. expert opinion	Existing evidence does not meet the needs of policy makers. Outcomes that experts want and GRADE requires are often not available. In such situations, expert opinion tends to dominate and experts do not always agree; expert opinions are often based on their own unique experiences and anecdotes, which may not necessarily be generalizable or valid.
Difficulties in learning and using the GRADE system	Systematic reviewers, policy makers, and TB experts are not necessarily trained in GRADE. Grading may be done inconsistently across tests by different systematic reviewers; same evidence can be interpreted and rated differently; GRADE ratings may be revised *posthoc*, depending on which tests the experts want to recommend. Some tests are actively 'championed', whereas others are not and this can result in uneven decisions
Conflicts of interest and involvement of test developers	Participation of test developers and industry representatives in the policy process introduces conflicts of interest. There is no consensus on whether test developers and those invested in specific technologies be allowed to do systematic reviews and participate in guideline panel meetings. There can be tension between commercial and noncommercial tests; type and quality of evidence might differ for commercial vs. noncommercial tests, and commercial products might be more actively championed by those with industry involvement.
Patient-important outcomes	Patient outcomes may not reflect the accuracy or benefit of a diagnostic test/approach in settings with weak overall health infrastructure (e.g. rapid or improved microscopy in facilities where stock-outs of anti-TB drugs occur frequently). The possible tension (for TB diagnosis and control) between the importance of individual patient outcomes and public health outcomes (e.g. the notion that false-negative sputum smear results may pose a greater public health risk than false-positive results). For tests used at the central/reference laboratory level, patient-outcome data may not be a good index of a test's impact; the test's impact is confounded by several other factors such as specimen transport, time to get results back to the clinicians, weak healthcare systems, etc. Impact on patient outcomes is affected not just by the test, but the whole package, including treatment, healthcare system efficiency, etc. It can be difficult to separate out the test's impact, and hard/expensive to study the whole package or strategy (which can be time-consuming and expensive). Diagnostic RCTs are rarely available and very hard to do (ethics, cost, etc.) In addition to patient values and preferences, need to acknowledge preferences and values of laboratory technologists and test users. If RCTs and patient-important outcomes are required for noncommercial tests, this will be severely limited by access to funds required to perform these large-scale evaluations.
Systematic review methods	No standardized methodology to search for and objectively synthesize evidence on operational implementation issues, costs to health services, costs to patients, and patient perspectives on new diagnostic tests and approaches. Narrative evidence on the above issues may be excluded from search strategies during systematic reviews of studies on diagnostic accuracy. Results from qualitative and socio-economic studies may not have been captured in the systematic reviews on diagnostic accuracy of the different approaches. Systematic reviews can make an effort to look for, include, and describe outcomes other than sensitivity and specificity, but often do not because they choose to focus instead on easily meta-analyzable outcomes. Policy makers should have a thorough understanding of all the important outcomes (including outcomes that are important to patients) they hope to include in their policy deliberations before commissioning systematic reviews. By explicitly outlining the test characteristics that will influence their decisions in advance, guideline panels can ensure evidence is as complete and objective as possible. This approach will minimize evidence gaps, making the process less susceptible to expert opinion. Weighting the importance of test characteristics in advance can also help to avoid redefining and reinterpreting evidence *posthoc* to suit individual desires to recommend or not recommend.

GRADE, Grading of Recommendations Assessment, Development and Evaluation; RCT, randomized controlled trial; TB, tuberculosis.

difficulties specific to diagnostics, which may be alleviated in time. For example, forcing diagnostic evidence into the RCT framework can be nonintuitive to laboratory researchers who typically conduct diagnostic evaluations. Certainly, the lack of experience using GRADE on the part of systematic reviewers and policy makers currently can lead to inconsistent interpretation of criteria and the revision of ratings *posthoc* in order to create GRADE profiles consistent with predetermined opinions regarding diagnostics that should be recommended. The transition from traditional policy making, which was made primarily based on expert opinion, to the use of more standardized, objective methods is likely to be a struggle for all organizations whether it is clearly acknowledged and dealt with or not.

The absence of diagnostic RCTs and data regarding patient-important outcomes and preferences in the field of TB diagnostics is a major hindrance to their assessments using GRADE, which currently places much weight on these aspects of patient care. As noted above, studies providing estimates of accuracy alone are downgraded for their lack of 'direct' evidence and thus cannot achieve a rating of 'high-quality' evidence. Although it can be agreed that higher levels of evidence need to be encouraged when assessing diagnostics, there are many practical barriers to extrapolating between the use of a diagnostic and the clinical outcomes of patients. Any number of deficiencies in the health system can impact a patient outcome, some of which may prevent the full recognition of benefits clearly provided by a diagnostic. At the same time, many user-important outcomes (as described above), which are of great importance to the feasibility of implementing diagnostics, are not easily captured in the GRADE process.

Diagnostic RCTs are almost nonexistent in TB. Even if they were feasible, there are concerns about their design, interpretation, and ethics [69]. Diagnostic RCTs do not just evaluate a test; they evaluate a strategy or package that includes testing followed by some intervention as a follow-up to the test result [44]. In this context, it is not easy to disentangle the efficacy of the test from the efficacy of the follow-up treatment or intervention. Furthermore, it is not easy to capture patient-important outcomes when ethical considerations prevent clinical decision-making on the basis of a trial product. Evidence from RCTs in highly controlled trial settings may not reflect the real-world conditions in which diagnostics have to be ultimately deployed. Lastly, diagnostic RCTs can take much longer than conventional diagnostic accuracy studies and this can delay the introduction of new policies.

The lack of stringent regulation and licensing of diagnostics certainly contributes to the lack of standardized, high-quality evidence available for the use of decision and policy makers. Additionally, this leads to the need for diagnostic policy processes to not only assess 'added benefit' of one test over another, but often to make the first objective assessment of a test's performance. The imposition of well defined, high standards at the stage of regulatory approval would help guide developers and researchers in their assessments of new diagnostics and provide impetus for the publication of appropriate and needed evidence. Compared to the therapeutics arena wherein strict regulation is imposed before a product is licensed for use, diagnostics require very limited data before they can be used to make patient care decisions. For example, despite a large body of evidence showing poor accuracy of commercial serological, antibody detection tests for TB, several commercial serological tests are on the market and used frequently in developing countries with weak regulatory systems [16,17,70,71]. Poorly performing diagnostics continue to remain on the market despite poor performance in the published literature and there are no mechanisms to 'withdraw' or 'ban' a bad diagnostic.

It needs to be recognized that by the nature of systematic reviews (upon which the GRADE process is reliant), the questions which are asked are of paramount importance [72]. Search criteria, selection processes, and presentation of evidence will all depend on the exact questions posed. If policy makers have a clear understanding of the issues that are important for implementation of a given diagnostic in advance, then evidence can be objectively collected to inform decisions and assessments on both quantitative and qualitative aspects. However, if only issues of test accuracy and technical performance are covered by systematic reviews, then gaps pertaining to other aspects of performance may need to be filled through less objective expert opinion.

All things considered, policy making is a big challenge in TB, as it is in other areas of medicine. Although GRADE has its limitations and can definitely be improved and adapted for TB diagnostics, we believe it is a major advance over the conventional policy making process.

Challenges in translating policies into impact

Availability of new tools does not necessarily ensure their adoption and implementation. Translation of policy into practice requires better understanding of barriers to implementation and methods to overcome such barriers. The impact of new tests will depend largely on the extent of their introduction and acceptance into the global public sector. This will itself depend in part on policy decisions made by international technical agencies such as WHO, by donors, and ultimately by national TB programs. This area has been extensively reviewed by

the Stop TB Partnership's Task Force on Retooling and has led to the creation of a roadmap to guide global, regional, and country-based activities as well as guidelines for engaging stakeholders in retooling and the introduction of specific TB diagnostics [73–75]. The work of the time-limited and now disbanded Task Force on Retooling has been mainstreamed into routine TB control activities led by the DOTS Expansion Working Group and its Subgroup on Introducing New Tools and Approaches (INAT).

The major obstacles to diagnostic retooling for TB control are undoubtedly the poor laboratory infrastructure and weak healthcare delivery systems present in many disease-endemic countries [76]. This has been recognized for many years. Although vastly increased funds are being invested in diagnostics retooling through national investments and funding agencies, there is still little guidance available to countries on what new diagnostic tools, or combinations of these tools, should be implemented in their particular epidemiological/health systems settings, what laboratory capability or capacity should be built to support this implementation, or how this should be done. A roadmap for strengthening TB laboratories that is abreast with recent developments and addresses these issues is urgently needed [77]. Beyond introducing new diagnostics and strengthening laboratories, challenges will remain in the development of accessible, equitable, and high-quality diagnostic services based on them and ensuring that healthcare delivery systems are strengthened so that better diagnostic services translate into better care [78]. In many countries, the private healthcare sector is the dominant source of healthcare. Lack of private sector involvement in TB control is a major weakness in existing programs.

Conclusion

After decades of neglect and poor progress, there is now great excitement about the development and introduction of new diagnostics for TB. The diagnostics pipeline has rapidly expanded and several new tools and strategies have received WHO endorsement for implementation at the country level. There are major gaps in the existing pipeline and the evidence base is predominantly made up of research studies of test accuracy. Future TB diagnostic research needs to focus on clinically meaningful outcomes and also consider obstacles to implementation. The GRADE system has brought greater transparency and evidence-based approaches to policy making, though GRADE for diagnostics is still a work in progress. Future TB policies and guidelines will need to be transparent, evidence-based, and free of COI. Today, despite many years of intensive effort to remedy the situation, weak laboratories remain the major immediate obstacle to translating policy into practice in low-income and middle-income countries. With the engagement of all key stakeholders, these challenges can be addressed to translate all the scientific progress into public health impact.

Acknowledgements

M.P. is a recipient of a New Investigator Award from the Canadian Institutes of Health Research (CIHR). J.M. is a recipient of a Quebec Respiratory Health Training Program (QRHTP) Fellowship. These funding sources had no role in the preparation of this manuscript, nor the decision to submit the manuscript for publication.

The authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. M.P. serves as an external consultant for the Foundation for Innovative New Diagnostics (FIND). FIND is a nonprofit agency that works with several industry partners in developing and evaluating new diagnostics for neglected infectious diseases. M.P. serves as a co-chair of the Stop TB Partnership's New Diagnostics Working Group (NDWG) and A.R. serves as the secretary of the NDWG. K.S. serves as co-chair of the Evidence Synthesis subgroup of the NDWG.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (pp. 286–287).

1 • Pai M, Minion J, Sohn H, *et al.* Novel and improved technologies for tuberculosis diagnosis: progress and challenges. Clin Chest Med 2009; 30:701–716.
This paper reviews the existing evidence base on TB diagnostics, describes the progress of new technologies, and ends with a review of cost-effectiveness and modeling studies on the potential effect of new diagnostics in TB control.

2 Perkins MD, Cunningham J. Facing the crisis: improving the diagnosis of tuberculosis in the HIV era. J Infect Dis 2007; 196 (Suppl 1):S15–S27.

3 • World Health Organization & Stop TB Partnership. New laboratory diagnostic tools for tuberculosis control. Geneva: World Health Organization; 2008.
Detailed review of the existing TB diagnostics pipeline.

4 Pai M, O'Brien R. New diagnostics for latent and active tuberculosis: state of the art and future prospects. Semin Respir Crit Care Med 2008; 29:560–568.

5 • Pai M, Ramsay A, O'Brien R. Comprehensive new resource for evidence-based TB diagnosis. Expert Rev Mol Diagn 2009; 9:637–639.
This study describes the development of a new website resource on 'evidence-based TB diagnosis'.

6 • Pai M, Ramsay A, O'Brien R. Evidence-based tuberculosis diagnosis. PLoS Med 2008; 5:e156.
A survey of all the systematic reviews on TB diagnostic tests and the contribution made by systematic reviews for informing clinical practice and policy.

7 Steingart KR, Henry M, Ng V, *et al.* Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis 2006; 6:570–581.

8 Steingart KR, Ng V, Henry M, *et al.* Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis 2006; 6:664–674.

9 Mase SR, Ramsay A, Ng V, *et al.* Yield of serial sputum specimen examinations in the diagnosis of pulmonary tuberculosis: a systematic review. Int J Tuberc Lung Dis 2007; 11:485–495.

10 Greco S, Girardi E, Navarra S, Saltini C. The current evidence on diagnostic accuracy of commercial based nucleic acid amplification tests for the diagnosis of pulmonary tuberculosis. Thorax 2006; 61:783–790.

11 Ling DI, Flores LL, Riley LW, Pai M. Commercial nucleic-acid amplification tests for diagnosis of pulmonary tuberculosis in respiratory specimens: meta-analysis and meta-regression. PLoS One 2008; 3:e1536.

12 Dinnes J, Deeks J, Kunst H, *et al.* A systematic review of rapid diagnostic tests for the detection of tuberculosis infection. Health Technol Assess 2007; 11:1–196.

13 Pai M, Flores LL, Hubbard A, *et al.* Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis. BMC Infect Dis 2004; 4:6.

14 Pai M, Flores LL, Pai N, *et al.* Diagnostic accuracy of nucleic acid amplification tests for tuberculous meningitis: a systematic review and meta-analysis. Lancet Infect Dis 2003; 3:633–643.

15 Daley P, Thomas S, Pai M. Nucleic acid amplification tests for the diagnosis of tuberculous lymphadenitis: a systematic review. Int J Tuberc Lung Dis 2007; 11:1166–1176.

16 Steingart KR, Henry M, Laal S, *et al.* A systematic review of commercial serological antibody detection tests for the diagnosis of extra-pulmonary tuberculosis. Thorax 2007; 62:911–918.

17 Steingart KR, Henry M, Laal S, *et al.* Commercial serological antibody detection tests for the diagnosis of pulmonary tuberculosis: a systematic review. PLoS Med 2007; 4:e202.

18 Steingart KR, Dendukuri N, Henry M, *et al.* Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: a meta-analysis. Clin Vaccine Immunol 2009; 16:260–276.

19 Greco S, Girardi E, Masciangelo R, *et al.* Adenosine deaminase and interferon gamma measurements for the diagnosis of tuberculous pleurisy: a meta-analysis. Int J Tuberc Lung Dis 2003; 7:777–786.

20 Liang QL, Shi HZ, Wang K, *et al.* Diagnostic accuracy of adenosine deaminase in tuberculous pleurisy: a meta-analysis. Respir Med 2008; 102:744–754.

21 Jiang J, Shi HZ, Liang QL. Diagnostic value of interferon-g in tuberculous pleurisy: a meta-analysis. Chest 2007; 131:1133–1141.

22 Kalantri S, Pai M, Pascopella L, *et al.* Bacteriophage-based tests for the detection of Mycobacterium tuberculosis in clinical specimens: a systematic review and meta-analysis. BMC Infect Dis 2005; 5:59.

23 Cruciani M, Scarparo C, Malena M, *et al.* Meta-analysis of BACTEC MGIT 960 and BACTEC 460 TB, with or without solid media, for detection of mycobacteria. J Clin Microbiol 2004; 42:2321–2325.

24 Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and nontuberculous mycobacteria? Int J Tuberc Lung Dis 2006; 10:1192–1204.

25 Wang L, Turner MO, Elwood RK, *et al.* A meta-analysis of the effect of Bacille Calmette Guerin vaccination on tuberculin skin test measurements. Thorax 2002; 57:804–809.

26 Pai M, Zwerling A, Menzies D. T-cell based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med 2008; 149:177–184.

27 Menzies D, Pai M, Comstock G. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. Ann Intern Med 2007; 146:340–354.

28 Pai M, Riley LW, Colford JM Jr. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect Dis 2004; 4:761–776.

29 van Zyl-Smit RN, Zwerling A, Dheda K, Pai M. Within-subject variability of interferon-g assay results for tuberculosis and boosting effect of tuberculin skin testing: a systematic review. PLoS One 2009; 4:e8517.

30 Pai M, Kalantri S, Pascopella L, *et al.* Bacteriophage-based assays for the rapid detection of rifampicin resistance in Mycobacterium tuberculosis: a meta-analysis. J Infect 2005; 51:175–187.

31 Minion J, Pai M. Bacteriophage assays for rifampin resistance detection in Mycobacterium tuberculosis: updated meta-analysis. Presented at the World Health Organization Expert Group Meeting on Non-commercial Culture Methods and Mycobacteriophage-based Assays for Rapid Screening of Patients at Risk of Drug-Resistant Tuberculosis. September 8–9, 2009, Geneva, Switzerland.

32 Ling DI, Zwerling A, Pai M. GenoType MTBDR assays for the diagnosis of multidrug-resistant tuberculosis: a meta-analysis. Eur Respir J 2008; 32:1165–1174.

33 Morgan M, Kalantri S, Flores L, Pai M. A commercial line probe assay for the rapid detection of rifampicin resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. BMC Infect Dis 2005; 5:62.

34 Bwanga F, Hoffner S, Haile M, Joloba ML. Direct susceptibility testing for multi drug resistant tuberculosis: a meta-analysis. BMC Infect Dis 2009; 9:67.

35 Martin A, Portaels F, Palomino JC. Colorimetric redox-indicator methods for the rapid detection of multidrug resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. J Antimicrob Chemother 2007; 59:175–183.

36 Martin A, Panaiotov S, Portaels F, *et al.* The nitrate reductase assay for the rapid detection of isoniazid and rifampicin resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. J Antimicrob Chemother 2008; 62:56–64.

37 Minion J, Leung E, Menzies D, Pai M. Microscopic-Observation Drug Susceptibility (MODS) and Thin Layer Agar (TLA) assays for the detection of drug resistant tuberculosis: a systematic review. Presented at the World Health Organization Expert Group Meeting on Non-commercial Culture Methods and Mycobacteriophage-based Assays for Rapid Screening of Patients at Risk of Drug-Resistant Tuberculosis. September 8–9, 2009, Geneva, Switzerland.

38 Pai M, O'Brien R. Tuberculosis diagnostics trials: do they lack methodological rigor? Expert Rev Mol Diagn 2006; 6:509–514.

39 Fontela PS, Pai NP, Schiller I, *et al.* Quality and reporting of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD standards. PLoS One 2009; 4:e7753.

40 Banoo S, Bell D, Bossuyt P, *et al.* Evaluation of diagnostic tests for infectious diseases: general principles. Nat Rev Microbiol 2006; 4 (9 Suppl):S21–S31.

41 Bossuyt PM, Reitsma JB, Bruns DE, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Ann Intern Med 2003; 138:40–44.

42 •• Stop TB Partnership's New Diagnostics Working Group and World Health Organization. Pathways to better diagnostics for tuberculosis: a blueprint for the development of TB diagnostics. Geneva: World Health Organization; 2009.

Comprehensive, well referenced blueprint to guide researchers, clinicians, industry partners, academics, and TB controllers in all sectors in all aspects of TB diagnostics development, from concept to evaluation, implementation, scale-up, delivery, and impact.

43 Bossuyt PM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making 2009; 29:E30–E38.

44 Lord SJ, Irwig L, Bossuyt PM. Using the principles of randomized controlled trial design to guide test evaluation. Med Decis Making 2009; 29:E1–E12.

45 Keeler E, Perkins MD, Small P, *et al.* Reducing the global burden of tuberculosis: the contribution of improved diagnostics. Nature 2006; 444 (Suppl 1): 49–57.

46 • Abu-Raddad LJ, Sabatelli L, Achterberg JT, *et al.* Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. Proc Natl Acad Sci U S A 2009; 106:13980–13985.

Recent modeling study on potential impact and epidemiological benefits of new TB diagnostic tools.

47 Getahun H, Harrington M, O'Brien R, Nunn P. Diagnosis of smear-negative pulmonary tuberculosis in people with HIV infection or AIDS in resource-constrained settings: informing urgent policy changes. Lancet 2007; 369: 2042–2049.

48 Marais BJ, Pai M. New approaches and emerging technologies in the diagnosis of childhood tuberculosis. Paediatr Respir Rev 2007; 8:124–133.

49 Barry CE 3rd, Boshoff HI, Dartois V, *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol 2009; 7:845–855.

50 Pai M. Spectrum of latent tuberculosis: existing tests cannot resolve the underlying phenotypes. Nat Rev Microbiol 2010 Jan 19 [Epub ahead of print].

51 • World Health Organization. Moving research findings into new WHO policies. http://www.who.int/tb/dots/laboratory/policy/en/index4.html. Geneva: World Health Organization; 2008.

This WHO document describes the new process for policies on tuberculosis.

52 • Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008; 149:889–897.

This study describes the Cochrane Collaboration methodologies and approaches for diagnostic test accuracy meta-analyses.

53 World Health Organization. The use of liquid medium for culture and DST. http://www.who.int/tb/dots/laboratory/policy/en/index3.html. Geneva: World Health Organization; 2007.

54 World Health Organization. Reduction of number of smears for the diagnosis of pulmonary TB [cited 27 January 2010]. http://www.who.int/tb/dots/laboratory/policy/en/index2.html

55 World Health Organization. Definition of a new sputum smear-positive TB case [cited 27 January 2010]. http://www.who.int/tb/dots/laboratory/policy/en/index1.html

56 World Health Organization. Policy statement. Molecular line probe assays for rapid screening of patients at risk of multidrug-resistant tuberculosis (MDR-TB) [cited 27 January 2010]. http://www.who.int/tb/dots/laboratory/policy/en/index4.html

57 World Health Organization. Report of the 9th Meeting of the Strategic and Technical Advisory Group on Tuberculosis (STAG-TB). Geneva: World Health Organization; 2009 http://www.who.int/tb/advisory_bodies/stag/en/index.html

58 Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on •• rating quality of evidence and strength of recommendations. BMJ 2008; 336:924–926.
This study provides a comprehensive introduction to the GRADE approach for developing policy recommendations.

59 Oxman AD, Lavis JN, Fretheim A. Use of evidence in WHO recommendations. Lancet 2007; 369:1883–1889.

60 Schunemann HJ, Oxman AD, Brozek J, *et al.* GRADE: assessing the quality of evidence for diagnostic recommendations. Evid Based Med 2008; 13:162–163.

61 World Health Organization. Guidelines for WHO Guidelines. Geneva: World Health Organization; 2003.

62 World Health Organization. WHO policy on TB infection control in health-care facilities, congregate settings and households. Geneva: World Health Organization; 2010.

63 World Health Organization. Treatment of tuberculosis. 4th ed. Geneva: World Health Organization; 2009.

64 Schunemann HJ, Oxman AD, Brozek J, *et al.* Grading quality of evidence and •• strength of recommendations for diagnostic tests and strategies. BMJ 2008; 336:1106–1110.
This report outlines the GRADE approach to diagnostic tests and strategies.

65 Brozek JL, Akl EA, Jaeschke R, *et al.* Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy 2009; 64:1109–1116.

66 Kavanagh BP. The GRADE system for rating clinical guidelines. PLoS Med 2009; 6:e1000094.

67 Pai M. Guidelines on IGRAs: concordant or discordant? 2nd Global Symposium on IGRAs. Dubrovnik, Croatia; 2009.

68 Schunemann HJ, Osborne M, Moss J, *et al.* An official American Thoracic Society Policy statement: managing conflict of interest in professional societies. Am J Respir Crit Care Med 2009; 180:564–580.

69 Biesheuvel CJ, Grobbee DE, Moons KG. Distraction from randomization in diagnostic research. Ann Epidemiol 2006; 16:540–544.

70 Steingart KR, Ramsay A, Pai M. Commercial serological tests for the diagnosis of tuberculosis: do they work? Future Microbiol 2007; 2:355–359.

71 World Health Organization & UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. Laboratory-based evaluation of 19 commercially-available rapid diagnostic tests for tuberculosis. Geneva: World Health Organization; 2008.

72 Pai M, McCulloch M, Enanoria W, Colford JM Jr. Systematic reviews of diagnostic test evaluations: what's behind the scenes? ACP J Club 2004; 141:A11–A13.

73 World Health Organization. New technologies for tuberculosis control: a framework for their adoption, introduction and implementation [WHO/HTM/STB/2007.40]. Geneva: World Health Organization; 2007.

74 Stop TB Partnership Retooling Task Force. Engaging stakeholders for retooling TB control. Geneva: World Health Organization; 2008.

75 Stop TB Partnership Retooling Task Force. Checklist of key actions for the use of liquid culture and drug susceptibility testing. Geneva: World Health Organization; 2008.

76 Ridderhof JC, van Deun A, Kam KM, *et al.* Roles of laboratories and laboratory systems in effective tuberculosis programmes. Bull World Health Organ 2007; 85:354–359.

77 World Health Organization. Strategic approach to the strengthening of laboratory services for tuberculosis control: 2006–2009. Geneva: World Health Organization; 2006.

78 Yagui M, Perales MT, Asencios L, *et al.* Timely diagnosis of MDR-TB under program conditions: is rapid drug susceptibility testing sufficient? Int J Tuberc Lung Dis 2006; 10:838–843.