The Journal of Infectious Diseases

SUPPLEMENT ARTICLE

Guidance for Studies Evaluating the Accuracy of Biomarker-Based Nonsputum Tests to Diagnose Tuberculosis

Paul K. Drain,[1,2,3,a] Jennifer Gardiner,[4,a] Haylea Hannah,[3] Tobias Broger,[5] Keertan Dheda,[6] Katherine Fielding,[7] Gerhard Walzl,[8] Myrsini Kaforou,[9] Katharina Kranzer,[7,10] Simone A. Joosten,[11] Christopher Gilpin,[12] Karin Weyer,[12] Claudia M. Denkinger,[5,13] and Samuel G. Schumacher[5]

Departments of [1]Global Health, [2]Medicine, and [3]Epidemiology, University of Washington, and [4]Bill & Melinda Gates Foundation, Seattle, Washington; [5]Foundation for Innovative New Diagnostics, Geneva; [6]Centre for Lung Infection and Immunity, Department of Medicine and University of Cape Town Lung Institute, University of Cape Town, South Africa; [7]London School of Hygiene and Tropical Medicine, United Kingdom; [8]Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa; [9]Imperial College London, United Kingdom; [10]Research Centre Borstel, Germany; [11]Department of Infectious Diseases, Leiden University Medical Centre, The Netherlands, [12]Global Tuberculosis Programme, World Health Organization, Geneva, Switzerland; and [13]University Hospital Heidelberg, Division of Tropical Medicine, Centre of Infectious Diseases, Germany

The World Health Organization's (WHO) "End TB" strategy calls for development and implementation of novel tuberculosis (TB) diagnostics. Sputum-based diagnostics are challenging to implement and often less sensitive in high-priority populations. Nonsputum, biomarker-based tests may facilitate TB testing at lower levels of the healthcare system, accelerate treatment initiation, and improve outcomes. We provide guidance on the design of diagnostic accuracy studies evaluating nonsputum, biomarker-based tests within the context of WHO's target product profile for such tests. Study designs should account for the intended use when choosing the study population, setting, and reference standards. Although adults with respiratory symptoms may be an initial target population, other high-priority populations regardless of symptoms—including people living with human immunodeficiency virus, those unable to produce sputum samples or with extrapulmonary TB, household contacts, and children—should be considered. Studies beyond diagnostic accuracy that evaluate feasibility and population-level impacts are also needed. A biomarker-based diagnostic may be critical to ending the TB epidemic, but requires appropriate validation before implementation.

Keywords. tuberculosis; diagnosis; biomarker; diagnostic accuracy; WHO End TB strategy; target product profiles; study design guidance.

Tuberculosis (TB) is the leading infectious cause of death worldwide [1]. In 2017, >10 million people developed TB disease and an estimated 1.6 million people died from TB [1]. Each year, approximately one-third of all TB cases are either not diagnosed or not reported [1]. In India and South Africa, 2 high-TB-burden countries, approximately 40% and 20%, respectively, of people with TB disease remain undiagnosed, much of which may be due to complicated diagnostic processes [2, 3].

The recommended diagnostic algorithms that rely on sputum-based tests to identify TB disease have many limitations, particularly in resource-limited settings [4]. Sputum can be difficult for some people to produce, especially certain high-risk groups, including people living with human immunodeficiency virus (PLWH) and children [5]. The generation of a sputum sample may expose healthcare workers to infectious aerosols, resulting in *Mycobacterium tuberculosis* infection [6]. Although extrapulmonary TB accounts for approximately 15% of all incident TB cases, people without any pulmonary involvement are unlikely to have a positive sputum result [1, 7]. People may undergo asymptomatic states of TB disease, such as incipient TB (asymptomatic infection likely to progress to TB disease) or subclinical TB (disease not causing clinical TB-related symptoms but resulting in other detectable abnormalities) [8, 9]. These disease states are less likely to be identified with sputum-based tests, given that the person has not developed a symptomatic cough [8]. Furthermore, many sputum-based TB tests, including the Xpert MTB/RIF assay (hereafter "Xpert") and smear microscopy, are processed in a centralized laboratory, which results in treatment delays and opportunities for loss to follow-up of contagious TB patients [10]. Consequently, in the absence of biomarker-based nonsputum diagnostic test for use at the clinical point of care, many people are either empirically treated for TB, receive inappropriate preventive therapy, or do not receive TB therapy at all.

The World Health Organization's (WHO) End TB Strategy calls for development and implementation of novel TB diagnostic tests [11]. Testing nonsputum samples, such as urine, breath, or blood, may allow for an expedited and more comprehensive diagnosis in primary healthcare settings, allowing for rapid initiation of anti-TB treatment.

[a]P. K. D. and J. G. contributed equally to this work.

Correspondence: P. K. Drain, MD, MPH, University of Washington, 325 Ninth Ave, UW Box 359927, Seattle, WA 98104-2420 (pkdrain@uw.edu).

The Journal of Infectious Diseases® 2019;220(S3):S108–15

DOI: 10.1093/infdis/jiz356

Despite new interest in developing biomarker-based diagnostic tests using nonsputum samples for TB, there has been little guidance on conducting appropriate evaluations of these novel tests. In a recent systematic review of TB biomarker studies published between 2010 and 2015, only 23 of 44 TB biomarkers were validated in a study with a low risk of bias (appropriate study design, participant selection, and gold standard diagnostics), and only 1 TB biomarker test, urine lipoarabinomannan (uLAM), had been reviewed by WHO [12]. In this manuscript, we discuss considerations for conducting diagnostic validation studies for evaluating novel biomarker-based, nonsputum tests to diagnose TB in high-burden settings.

Scope for Biomarker-Based, Nonsputum Tests for TB

Target product profiles (TPPs) are developed to align the needs of test users with the specifications that test developers should meet for test performance and operational characteristics [13]. The high-priority TPPs for TB, as defined by the WHO, specified the optimal and minimal requirements for performance characteristics of a biomarker-based, nonsputum diagnostic test (Appendix). The goal of such a test is to diagnose pulmonary and extrapulmonary TB using nonsputum samples with the aim of initiating TB treatment during the first clinical encounter to minimize the number of patient visits to a healthcare facility. The ideal target populations are adults and children, including PLWH, who are being investigated for TB disease and reside in a country with a medium to high TB prevalence. The test should meet the minimal requirements regarding technical equipment and environment, and be available at the lowest level of the healthcare system. The target users should be trained healthcare workers or microscopy technicians. A biomarker-based nonsputum test for TB should serve as a stand-alone diagnostic test that does not require further confirmatory testing. As some people may experience waxing and waning of TB symptoms [8], repeated serial testing may be necessary. While case definitions for incipient and subclinical TB have been proposed [8, 9], evaluating diagnostic tests and using the existing gold-standard test that clearly meets those criteria will be challenging. In general, biomarkers evaluating incipient and subclinical TB will need to have longitudinal study designs with repeated serial testing of individuals [14].

Ideal Criteria of Biomarker-Based, Nonsputum Tests for TB

The biomarker-based, nonsputum diagnostic TB test is intended to lead to the initiation of anti-TB treatment and must therefore have high specificity compared against a microbiological reference standard (target 98%). Sensitivity is outlined for optimal and minimal requirements with the highest targets set for adults with smear-positive, culture-positive pulmonary TB (≥98% optimally; overall pooled among all groups ≥65% minimally). Targets for diagnostic sensitivity vary between children (≥66% optimally; no minimal target set) and adults with extrapulmonary TB (≥80% optimally; no minimal target set). Extrapulmonary targets are lower compared to the targets set for pulmonary TB to reflect an opportunity to improve diagnosis for these patient populations. The established targets were selected to match the currently available best-performing tests (eg, smear microscopy) for each patient population in high-burden-TB settings. Of note, the accuracy targets for a biomarker-based, nonsputum diagnostic test are in contrast to a triage test, which should have higher diagnostic sensitivity (>90% minimally), lower diagnostic specificity (>70% minimally), and require confirmatory TB testing prior to treatment initiation (see paper by Nathavitharana et al in this supplement).

The desired samples for the biomarker-based, nonsputum diagnostic TB test should be performed with minimal sample preparation and may include urine, blood, saliva, and exhaled air (breath) (Appendix). The test should provide results in <1 hour with no instrument (optimal) or a small, portable handheld instrument operated by a battery (minimal). During the development of a biomarker-based diagnostic, the intended user, such as a trained nurse in a district-level hospital or healthcare worker based in a community clinic, should be defined. Feasibility of conducting and interpreting a certain test, and hence rollout of the test, is directly influenced by the intended user profile.

General Study Design Considerations

Despite the opportunity to transform TB diagnosis, a biomarker-based, nonsputum TB test has several key challenges to overcome before it can be established as a useful diagnostic tool. Diagnostic test accuracy should be evaluated using a cross-sectional or cohort study of people who have signs or symptoms consistent with pulmonary or extrapulmonary TB disease (Table 1). It is important to avoid using known and severe cases, as this may introduce bias. Test performance may vary by clinical setting due the different mixture of patients. For example, testing of patients in centralised settings may lead to overestimation of accuracy as patients often present at later stages of disease. Other study designs, including ease of use and analytical studies, may be considered in parallel to investigate other test aspects.

Studies should be designed with necessary statistical power in adult pulmonary TB patient populations. For example, for a test with 65% sensitivity (minimum TPP target), 160–320 reference standard–positive samples would be required to achieve 95% confidence intervals (CIs) with a total CI width of 15% and 10%, respectively (Figure 1). To demonstrate the required specificity of 98% with a 5% or 4% CI width, 140–250 reference standard–negative samples would be required, respectively. Assuming a 30% TB prevalence, this would translate into enrollment of 530–1070 participants, whereas a 15% TB prevalence would require 1070–2130 participants. Later sections discuss considerations for populations and settings when designing such studies, which may influence sample size and statistical power considerations.

Table 1. Overview of Recommendations for Study Design

Topic	Recommendation
General study design	• Use a prospective study design enrolling participants who require evaluation for TB (avoid using known and severe cases)
	• Consider how many reference standard–positive and –negative samples are required to obtain a precise estimate of the sensitivity and specificity, respectively
	• Conduct follow-up visits evaluating vital status and presence of TB symptoms at 2 mo from the initial diagnostic test at minimum
	• Follow STARD and QUADAS-2 guidelines for reporting and designing diagnostic accuracy studies
Population and setting	• Avoid selecting participants in whom TB has already been diagnosed by another test or who have already started on TB treatment
	• Focus on adults, including PLHIV, who have respiratory symptoms suggestive of TB in initial studies, but do not exclude patients based on inability to provide a sputum sample
	• Studies evaluating the diagnostic accuracy of tests targeting a biomarker in high-priority groups (eg, children, extrapulmonary TB) may prioritize a different initial study population.
	• Perform testing in intended use setting if quality of testing can be assured
	• Provide stratified accuracy estimates for key subpopulations (by HIV status and smear status)
Index test	• Consider specifics of the test being evaluated and how they might influence the study design, ideal population, and ideal setting for evaluation
	• For example, if tests have a non-automated read-out, test users should be blinded to the clinical information, reference standard, and comparator tests results
Reference standard and comparators	• Report results from >1 type of reference standard to better understand any bias that may be associated with the choice of reference standard (eg a microbiological reference standard may underestimate specificity; see further explanation under reference standard in the text and Figure 2)
	• Include a minimum of 2 cultures to either diagnose or exclude TB disease
	• Consider smear microscopy and GeneXpert as comparators when assessing performance for novel diagnostics for pulmonary TB, and modify the reference standard based on the target population, as needed (ie, sputum-based tests may not be an appropriate with sufficient reference standard for PLHIV)
Flow and specimen issues	• Carefully design and report the study sample flow and preparation, considering the limitations of each possible approach
Key issues beyond accuracy	• Consider the willingness and training that clinical teams would require to adopt a test into the decision-making process as well as training needs for laboratory personnel
	• The potential clinical, cost, and population-level impacts of new tests need to be assessed through modeling and empirical studies

Abbreviations: HIV, human immunodeficiency virus; PLHIV, people living with human immunodeficiency virus; QUADAS-2, The Quality Assessment of Diagnostic Accuracy Studies 2; STARD, Standards for Reporting of Diagnostic Accuracy Studies; TB, tuberculosis.

Baseline participant information (eg, TB symptoms, TB history, human immunodeficiency virus [HIV] status) should be collected and researchers should consider follow-up visits for a minimum of 2 months from the initial diagnostic test to ensure resolution of symptoms in those not diagnosed with TB. A follow-up visit at 6 months may also be warranted to further confirm disease status but should be weighed with concerns about loss to follow-up and resource limitations.

Population and Setting

The ideal population for initial studies of a new biomarker-based nonsputum diagnostic may be symptomatic adults seeking care at a healthcare facility for pulmonary TB, due to the availability of a reliable diagnostic reference standard. However, patients who cannot provide sputum specimens should not be excluded as they represent a population that may benefit most from biomarker-based tests. The interpretation of test results for people with clinically diagnosed TB (ie, clinical TB treated for TB without bacteriological confirmation) presents challenges in initial evaluation studies. Because children and adults with extrapulmonary TB have a high rate of clinical TB, these groups should be included in diagnostic studies, once the diagnostic accuracy has been reliably established in patients with pulmonary TB, unless the diagnostic TB test specifically targets these populations. Additional validation studies conducted in pediatric and/or extrapulmonary TB cohorts could complement the data on adults with pulmonary TB disease and inform recommendations for use in these important populations.

Tests should be evaluated in multicountry studies to capture and assess the genetic heterogeneity of study populations and the different epidemiology of TB. Last, researchers should explicitly state the target population and setting, as diagnostic accuracy depends on the burden of disease, study populations, and research setting.

Ideally, studies would be conducted in the target setting of intended use. The aim for the biomarker-based, nonsputum TB test is deployment across the health system, including in primary care clinics and challenging settings where end users have minimal technical knowledge. Alternatively, initial data could be generated in controlled laboratory settings with special attention to testing requirements that will be challenging to implement in the target setting of intended use. With this approach,

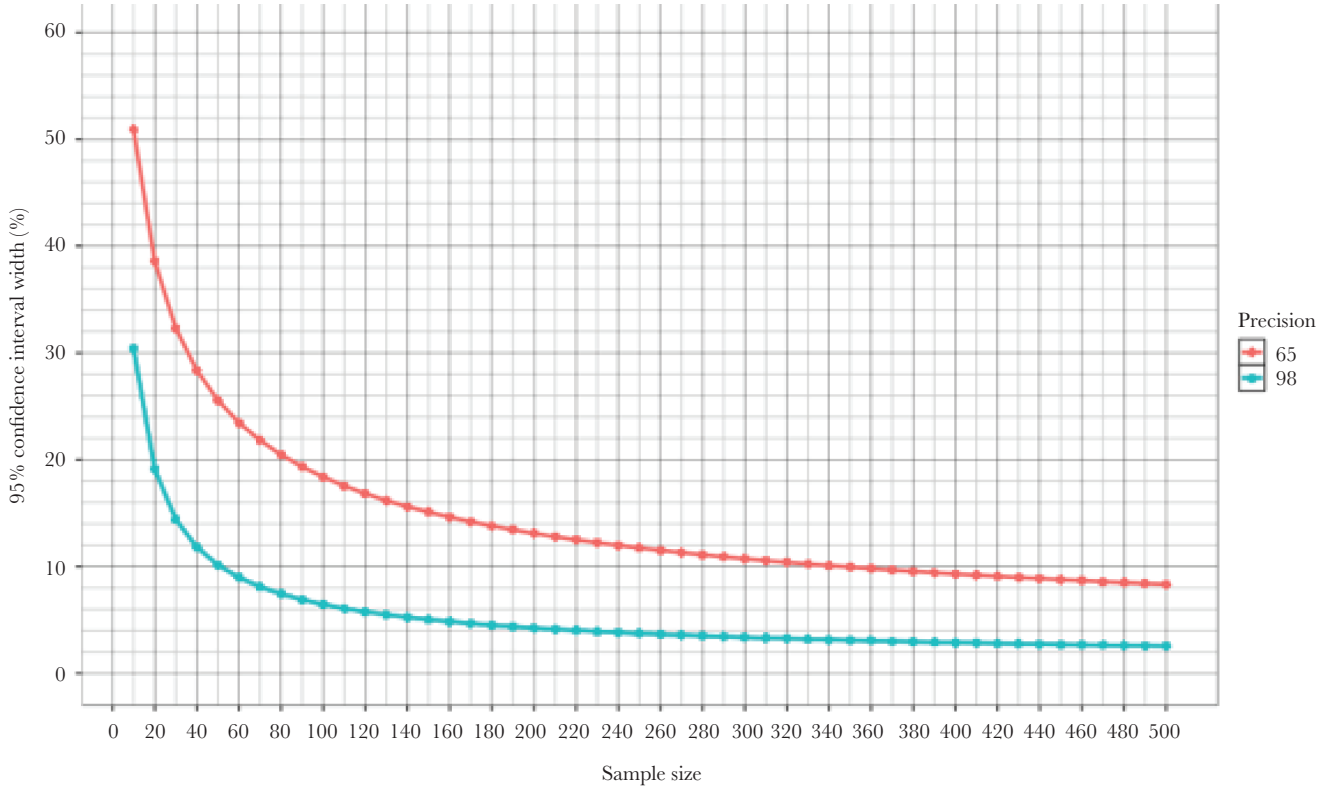

Figure 1. Precision of accuracy estimates as function of sample size. The lines show the precision of accuracy estimates as a function of sample size; accuracy point estimates are chosen according to the minimal target based on the target product profile: sensitivity (65%, red line) and specificity (98%, green line). The y-axis shows total width of the 95% confidence interval for sensitivity and specificity for a given sample size. The x-axis shows the necessary number of participants with tuberculosis (TB) to achieve a given precision for sensitivity (number of TB patients) and the number of participants without TB to achieve a given precision for specificity (number of non-TB patients).

an assessment of whether test performance varies depending on level of training of the operators will be important. Depending on the sample tested, participant enrollment may need to occur at the same or nearby location as the collection site before being transported to a central laboratory.

Index Test

The index test refers to the diagnostic test under evaluation in the research study. When designing diagnostic accuracy studies, researchers should consider test administration, interpretation, and setting and whether the assay read-out is automated or requires a degree of subjective interpretation (eg, visual reader). Test readers should be blinded to clinical information and results of the reference standard and comparator tests when reading and interpreting results for the index test. Interreader reliability from independent and blinded visual interpretation needs to be assessed and considered when interpreting results from diagnostic accuracy studies.

Reference Standards and Comparators

The inclusion of multiple target populations will require relevant reference standards to provide a complete assessment of performance of a biomarker-based, nonsputum TB diagnostic. With current diagnostic tools, the reference standard for TB disease may not be clear, but will generally rely on sputum culture for a microbiological test result. Some participants may have subclinical TB or paucibacillary TB, only to develop TB-related symptoms shortly after diagnostic testing. For participants who are negative by the reference standard at enrollment, and not started on empiric treatment, researchers may conduct repeated reference standard testing within the following 2 months for a more accurate classification.

In general, there are several options for the reference standard for comparing novel TB diagnostics (Figure 2). Each reference standard addresses 3 categories of participants tested for TB differently: microbiologically confirmed TB (true positive), clinical TB, and no TB (true negative). People with clinical TB may present with symptoms or signs suggestive of TB and are diagnosed with TB disease by a healthcare provider, but TB is not detectable [15]. These reference standards have associated advantages and disadvantages depending on the target diagnosis and population being studied. It is critical to clearly describe how the reference standard is defined; an example of each reference standard is provided in Table 2.

1. Excluding clinical TB: In this approach, participants with clinical TB are excluded from the analysis. This analysis easily distinguishes between true-positive and true-negative cases based

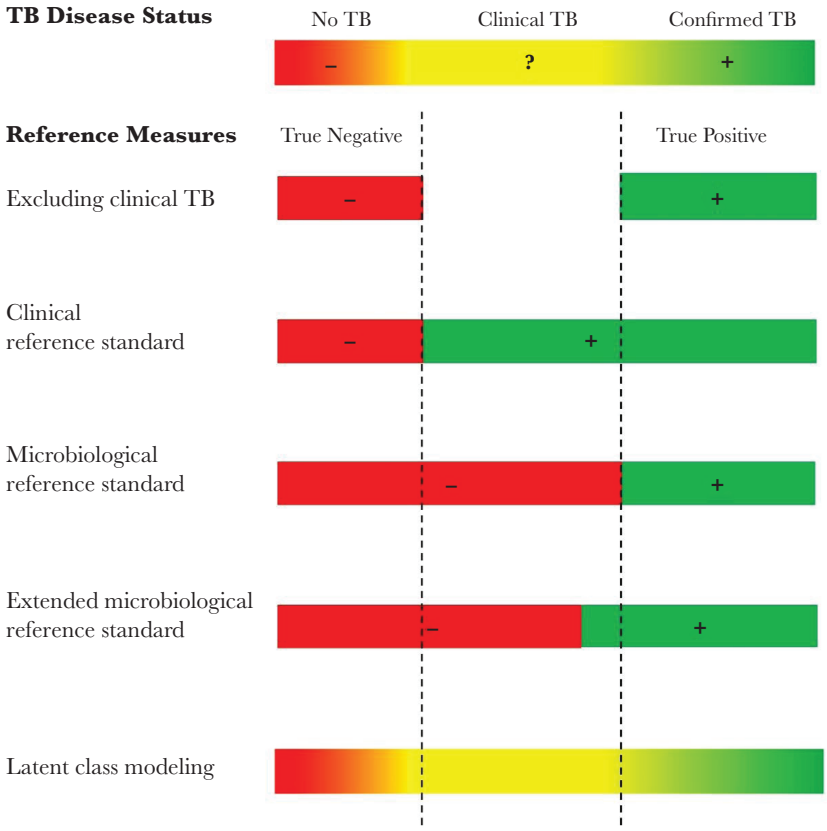

Figure 2. Defining outcomes for the reference standard test when designing tuberculosis diagnostic accuracy studies. For excluding clinical tuberculosis (TB), those with clinical TB (in yellow) are excluded from the analysis. For the clinical reference standard (CRS), those with clinical TB are often included as a confirmed TB case. For the microbiological reference standard (MRS), those with clinical TB are included as a noncase (no TB). The extended MRS classifies a few additional clinical TB cases into the confirmed TB category. Latent class modeling uses statistical modeling to incorporate all available test results to estimate the probability of TB in each individual, and test accuracy is estimated while accounting for uncertainty in disease classification [15, 16]. Ideally, results from all of the above should be presented but at least the "case/noncase," CRS, and MRS are easily done without much additional work.

only on the microbiological reference standard. However, this approach usually yields overestimates of sensitivity and specificity by removing the individuals who are the most difficult to diagnose.

2. Clinical reference standard (CRS): In this approach, researchers may apply an algorithm based on the results from clinical tests (eg, chest radiograph [CXR]) to determine a final diagnosis. As such, those participants with clinical TB may be grouped with participants who have confirmed TB disease as positive TB cases. This approach may underestimate sensitivity, since not all clinical TB cases truly have disease. The specificity may be overestimated because the no-TB group is restricted to those in whom disease was ruled out with a high degree of certainty.
3. Microbiological reference standard (MRS): In this approach, only participants who have positive microbiological test results are classified as TB cases, while participants with clinical TB are classified as not having TB disease. In contrast to the CRS, this approach may yield overestimates for sensitivity, but underestimates for specificity.
4. Extended microbiological reference standard: In this approach, additional combined clinical, microbiological, or pathological tests are used to further classify participants with clinical TB as a TB case. In comparison to MRS, this approach may yield less-biased overestimates for sensitivity and underestimates for specificity.

In addition to using different reference standards for classification of cases, researchers may also use latent class modeling to account for uncertainty in disease classification. In this approach, statistical modeling is used to estimate diagnostic accuracy by estimating the probability of TB in each individual [16, 17]. This method acknowledges some uncertainty that becomes incorporated into each parameter estimate. These models can be complex to construct but have the potential to generate more accurate estimates of diagnostic sensitivity and specificity.

While a reference standard should be chosen before conducting the primary statistical analyses, reporting results according to several possible reference standards may be valuable.

Table 2. Sample Applications of Tuberculosis Reference Standards

Parameter	Excluding Clinical TB	MRS	eMRS	CRS
Tests and follow-up considered				
1–2 MGIT[a]	X	X	X	X
1–2 LJ[a]	X	X	X	X
Blood culture[a]	X	X	X	X
Urine Ultra	X	X	X	X
Sputum Xpert/Ultra	X	X	X	X
Additional testing[b]	…	…	X	X
Clinical follow-up	…	…	…	X
Symptoms and treatment at 2–3 mo				
Persistent symptoms	…	…	…	X
Initiation of ATT	…	…	…	X
Reference standard positive	Any of the tests considered is positive	Any of the tests considered is positive	Any of the tests considered is positive	Any of the tests considered is positive and/or TB treatment was started
Reference standard negative	None of the tests considered is positive and at least 1 test is negative	None of the tests considered is positive and at least 1 test is negative	None of the tests considered is positive and at least 1 test is negative	None of the tests considered is positive and the patient has no symptoms and TB treatment was not started
Unclassifiable (excluded from analysis)	Neither reference standard positive nor reference standard negative; patients with clinical TB	Neither reference standard positive nor reference standard negative	Neither reference standard positive nor reference standard negative	Neither reference standard positive nor reference standard negative

Abbreviations: ATT, anti-tuberculosis therapy; CRS, clinical reference standard; eMRS, extended microbiological reference standard; LJ, Löwenstein–Jensen; MGIT, mycobacterial growth indicator tube; MRS, microbiological reference standard; TB, tuberculosis.

[a] *Mycobacterium tuberculosis* complex needs to be confirmed.

[b] Any additional mycobacterial culture or GeneXpert/Ultra from other respiratory and/or nonrespiratory samples (eg, pleural fluid, tissue biopsy, cerebrospinal fluid) that were performed based on clinical indication.

The direction and extent of the bias for the CRS may vary, depending on the specific use case. A biomarker-based diagnostic has the potential to detect culture-negative TB at earlier stages, which may not be captured by using an MRS [13].

Regardless of the reference standard used in analyses, high-quality diagnostic accuracy studies should ideally include a minimum of 2 sputum cultures to either diagnose or exclude pulmonary TB disease. However, obtaining 2 sputum cultures may be impractical for pragmatic studies in settings with fewer resources. For example, a single negative Xpert test may not be sufficient in a research setting to accurately classify an individual as a non-TB case, particularly when performed on acid-fast bacilli smear-negative sputum from PLWH. For certain populations, such as PLWH, for whom sputum-based tests are less sensitive, the reference standard chosen may be based on the availability of TB test results, clinical information, CD4 cell count, and HIV treatment status. In children and adults with extrapulmonary disease, culture should ideally be performed on multiple available samples (eg, cerebrospinal fluid, blood, urine, gastric aspirate, stool, biopsies). However, a CRS may be most appropriate for children, since microbiological confirmation can be rare. Researchers should also consider including CXR and other radiological imaging within a reference standard based on the target study population.

Smear microscopy and Xpert assays are important comparator tests when assessing performance for novel diagnostics for pulmonary TB disease to put results into context with these widely used tools and provide further insights into the spectrum of the study population. In PLWH, alternative tests leveraging nonsputum specimens should be included (eg, uLAM or CXR) due to data showing increased diagnostic yield in participants with advanced HIV [18]. Depending on the objective and planned analyses, additional testing of the reference measures may include bacillary load, time to culture positivity, or drug susceptibility testing. In general, diagnostic accuracy of the index test should be reported according to each available reference standard.

Flow and Specimen Issues

Initial studies may leverage banked specimens for a cost-efficient investigation of biomarkers in diverse samples and patient populations. Prior to initiating such studies, stability in frozen samples should be demonstrated. In prospective studies where fresh samples are tested, attention must be paid to the timing of specimen collection and testing if the biomarker can degrade or if the biological specimen changes over time.

Key Considerations Beyond Accuracy

Current practice relies on sputum culture to identify people along the spectrum of latent to active TB who require treatment. Therefore, an inherent risk in moving to a nonsputum-based diagnosis for pulmonary TB is the challenge in matching the

sputum-based definition with an alternate biomarker-based, nonsputum TB test. However, there may be potential benefits from earlier treatment (possibly a shortened or modified regimen) for less ill participants who are not producing sputum. Furthermore, the use of such a test in primary healthcare setting will likely lead to testing of participants who may be less ill and may be earlier in the progression from latent to active TB, potentially reducing transmission. Together, studies of nonsputum biomarkers may identify a portion of participants who are negative on the reference standard assay and yet positive by the biomarker assay being evaluated. These scenarios will require follow-up studies to further characterize these participants, identify those who will benefit from treatment, and then demonstrate clinical benefit.

Studies evaluating the training required for test performers and readers, and if index test performance varied by time of day or operator, are of interest for test implementation. Studies should also consider the training that clinical teams would require to adopt a test into the decision-making process. Diagnostic tests that are well designed and with high performance have not resulted in improved clinical outcomes in absence of appropriate buy-in [19]. Studies should gather early feedback from users and clients to inform sample handling and device characteristics. Additional considerations for evaluation are both the feasibility and acceptability of testing among participants and healthcare workers. Although manufacturers may evaluate test stability in simulated, well-controlled environments, studies may assess the stability of tests to varying storage methodologies to complement these studies (eg, sunlight, temperature, freezing, exposure to poor air quality or dust). Test developers should specify the ideal time to testing, as well as test and specimen storage temperature. These parameters will be essential for evaluating the test in low-resource communities with hot and humid climates.

Studies of new technologies often suffer from bias and inconsistent reporting of results. A self-evaluation of risk of study bias, using a tool for quality assessment of diagnostic accuracy studies (The Quality Assessment of Diagnostic Accuracy Studies 2) is useful both during study design and when reporting results [20]. Diagnostic studies should be reported using Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines [21]. Last, empiric treatment may lead to diagnostic misclassifications and underestimates of the incremental benefit of the diagnostic, so any use of clinical reference definitions need to be well described in a study protocol [22].

The effectiveness of a new TB diagnostic depends largely on its accuracy; however, its cost-effectiveness depends strongly on test complexity and price [23]. National TB programs may weigh diagnostic accuracy, beneficiaries, price, ease of use, and potential impact when recommending a new diagnostic test. Where possible, researchers should consider collecting cost data alongside TB outcomes to inform cost-effectiveness analyses.

Because efforts to develop a biomarker-based nonsputum diagnostic will be novel, there is an expectation that iterative studies and later design improvements will be required. Mathematical modeling studies can leverage these data to provide useful information on the population-level impact of nonsputum, biomarker-based tests, and inform diagnostic test implementation [24]. The field will benefit from openly available data to enable iterative analyses and refine future work based on prior efforts [25].

CONCLUSIONS

The limitations of currently available diagnostic tools allow the transmission of TB to continue in resource-limited settings with the highest TB burden [26]. The development of novel biomarker-based, nonsputum tests may be essential to eliminating TB [11, 27, 28]. Simple biomarker-based tests may reach high-priority populations currently being missed by sputum-based tests, such as PLWH, children, and those with extrapulmonary TB. Such tests also have the potential to be used at lower levels of the healthcare system and to diagnose people at earlier stages of TB disease progression, which may accelerate treatment initiation, improve clinical outcomes, and reduce TB transmission. Researchers need to carefully consider population, setting, and reference standards when designing diagnostic accuracy studies of biomarker-based tests, and ideally also assess feasibility and cost, aligning each with the scope and target of the diagnostic test. A novel biomarker-based TB test may be critical to ending the TB epidemic but will require appropriate validation before widespread implementation.

Notes

Supplement sponsorship. This supplement is sponsored by FIND (Foundation for Innovative New Diagnostics) and was made possible through the generous support of the Governments of the United Kingdom, the Netherlands, Germany and Australia.

Potential conflicts of interest. All authors: No reported conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

REFERENCES

1. World Health Organization. Global tuberculosis report 2018. Geneva, Switzerland: WHO, **2018**.
2. Naidoo P, Theron G, Rangaka MX, et al. The South African tuberculosis care cascade: estimated losses and methodological challenges. J Infect Dis **2017**; 216:S702–13.
3. Subbaraman R, Nathavitharana RR, Satyanarayana S, et al. The tuberculosis cascade of care in India's public sector: a systematic review and meta-analysis. PLoS Med **2016**; 13:1–38.

4. European Tuberculosis Laboratory Initiative, WHO Europ ean Region. Expert opinion of the European Tuberculosis Laboratory Initiative core group members for the WHO European Region. Geneva: World Health Organization, **2017**. http://www.euro.who.int/en/health-topics/communicable-diseases/tuberculosis/publications/2017/european-tuberculosis-laboratory-initiative-regional-tb-and-mdr-tb-diagnosis-workshop-report-2017. Accessed 15 July 2019.
5. Kendall EA. Tuberculosis in children: under-counted and under-treated. Lancet Glob Health **2017**; 5:e845–6.
6. Menzies D, Joshi R, Pai M. Risk of tuberculosis infection and disease associated with work in health care settings. Int J Tuberc Lung Dis **2007**; 11:593–605.
7. Purohit M, Mustafa T. Laboratory diagnosis of extra-pulmonary tuberculosis (EPTB) in resource-constrained setting: state of the art, challenges and the need. J Clin Diagnostic Res **2015**; 9:EE01–6.
8. Drain P, Bajema K, Dowdy D, et al. Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. Clin Microbiol Rev **2018**; 31. doi:10.1128/CMR.00021-18.
9. Achkar JM, Jenny-Avital ER. Incipient and subclinical tuberculosis: defining early disease states in the context of host immune response. J Infect Dis **2011**; 204(Suppl 4):1179–86.
10. Cohen GM, Drain PK, Noubary F, Cloete C, Bassett IV. Diagnostic delays and clinical decision making with centralized Xpert MTB/RIF testing in Durban, South Africa. J Acquir Immune Defic Syndr **2014**; 67:e88–93.
11. World Health Organization. The End TB strategy. Geneva, Switzerland: WHO, **2015**.
12. MacLean E, Broger T, Yerliyaka S, Fernandez-Carballo L, Pai M, Denkinger CM. Biomarkers to detect active tuberculosis: a systematic review of the evidence, study quality and progress. Nat Microbiol **2019**; 4:748–58.
13. World Health Organization. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva, Switzerland: WHO, **2014**. http://www.who.int/tb/publications/tpp_report/en/. Accessed 15 July 2019.
14. Kik SV, Schumacher S, Maria Cirillo D, et al. An evaluation framework for new tests that predict progression from tuberculosis infection to clinical disease. Eur Respir J **2018**; 52. pii:1800946.
15. World Health Organization. Definitions and reporting framework for tuberculosis—2013 revision. Geneva, Switzerland: WHO, **2014**. http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345_eng.pdf. Accessed 15 July 2019.
16. Schumacher SG, van Smeden M, Dendukuri N, et al. Diagnostic test accuracy in childhood pulmonary tuberculosis: a Bayesian latent class analysis. Am J Epidemiol **2016**; 184:690–700.
17. Stout JE, Wu Y, Ho CS, et al. Evaluating latent tuberculosis infection diagnostics using latent class analysis. Thorax **2018**; 73:1062–70.
18. Shah M, Hanrahan C, Zy W, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. Cochrane Database Syst Rev **2016**; 5:CD011420.
19. Huang DT, Yealy DM, Filbin MR, et al. Procalcitonin-guided use of antibiotics for lower respiratory tract infection. N Engl J Med **2018**; 379:236–49.
20. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med **2011**; 155:529–36.
21. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open **2016**; 6:e012799.
22. Theron G, Peter J, Dowdy D, Langley I, Squire SB, Dheda K. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? Lancet Infect Dis **2014**; 14:527–32.
23. Dowdy DW, Brien MAO, Bishai D. Cost-effectiveness of novel diagnostic tools for the diagnosis of tuberculosis. Int J Tuberc Lung Dis **2008**; 12:1021–9.
24. Egger M, Johnson L, Althaus C, et al. Developing WHO guidelines: time to formally include evidence from mathematical modelling studies. F1000Research **2017**; 6:1584.
25. Hey S, Kesselheim A. Countering imprecision in precision medicine. Science **2016**; 353:448–9.
26. Sigal GB, Segal MR, Mathew A, et al. Biomarkers of tuberculosis severity and treatment effect: a directed screen of 70 host markers in a randomized clinical trial. EBioMedicine **2017**; 25:112–21.
27. Stop TB Partnership. Global plan to end TB 2016–2020: the paradigm shift. Geneva, Switzerland: United Nations Office for Project Services, **2015**. http://www.stoptb.org/assets/documents/global/plan/GlobalPlanToEndTB_TheParadigmShift_2016-2020_StopTBPartnership.pdf. Accessed 15 July 2019.
28. Wallis R, Kim P, Cole S, et al. Tuberculosis biomarkers discovery: developments, needs, and challenges. Lancet Infect Dis **2013**; 13:362–72.