

Guidance for Studies Evaluating the Accuracy of Tuberculosis Triage Tests

Ruvandhi R. Nathavitharana,^{1,9} Christina Yoon,² Peter MacPherson,^{3,4} David W. Dowdy,⁵ Adithya Cattamanchi,^{2,6} Akos Somoskovi,⁷ Tobias Broger,⁸ Tom H. M. Ottenhoff,⁹ Nimalan Arinaminpathy,¹⁰ Knut Lonnroth,¹¹ Klaus Reither,^{12,13} Frank Cobelens,¹⁴ Christopher Gilpin,¹⁵ Claudia M. Denkinger,^{8,16} and Samuel G. Schumacher¹⁶

¹Department of Infectious Diseases, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston; ²Division of Pulmonary and Critical Care Medicine, Department of Medicine, Zuckerberg San Francisco General Hospital, University of California San Francisco; ³Department of Clinical Sciences, Liverpool School of Tropical Medicine, United Kingdom; ⁴Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Malawi; ⁵Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; ⁶Center for Tuberculosis and Division of Pulmonary and Critical Care Medicine, University of California San Francisco; ⁷Intellectual Ventures Laboratory, Global Good Fund, Bellevue, Washington; ⁸Foundation for Innovative Diagnostics (FIND), Geneva, Switzerland; ⁹Department of Infectious Diseases, Leiden University Medical Center, The Netherlands; ¹⁰MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London; ¹¹Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden; ¹²University of Basel, Switzerland; ¹³Department of Medicine, Swiss Tropical and Public Health Institute, Basel, Switzerland; ¹⁴Department of Global Health, Amsterdam UMC, University of Amsterdam, The Netherlands; ¹⁵World Health Organization, Geneva, Switzerland; ¹⁶University Hospital Heidelberg, Division of Tropical Medicine, Centre of Infectious Diseases, Germany

Approximately 3.6 million cases of active tuberculosis (TB) go potentially undiagnosed annually, partly due to limited access to confirmatory diagnostic tests, such as molecular assays or mycobacterial culture, in community and primary healthcare settings. This article provides guidance for TB triage test evaluations. A TB triage test is designed for use in people with TB symptoms and/or significant risk factors for TB. Triage tests are simple and low-cost tests aiming to improve ease of access and implementation (compared with confirmatory tests) and decrease the proportion of patients requiring more expensive confirmatory testing. Evaluation of triage tests should occur in settings of intended use, such as community and primary healthcare centers. Important considerations for triage test evaluation include study design, population, sample type, test throughput, use of thresholds, reference standard (ideally culture), and specimen flow. The impact of a triage test will depend heavily on issues beyond accuracy, primarily centered on implementation.

Keywords. diagnostics; study design guidance; target product profiles; triage; tuberculosis.

Of the estimated 10 million new active tuberculosis (TB) cases each year, approximately 3.6 million are not notified and are potentially undiagnosed [1], resulting in poor individual outcomes and ongoing TB transmission within families and communities [2–4]. Systematic reviews have reported that reasons for delayed TB diagnosis include persons seeking care in the informal or private sectors or in community or primary healthcare settings [5, 6], where access to rapid, sensitive TB diagnostics is particularly limited [7], hindering the potential for early diagnosis and prompt treatment initiation.

Despite continued reliance on smear microscopy in high TB incidence settings, there have been considerable advances in the development of molecular diagnostic tests for TB. GeneXpert MTB/RIF (Xpert) (Cepheid Inc., Sunnyvale, CA) is a cartridge-based nucleic acid amplification test for rapid diagnosis of TB and rifampicin resistance that was first endorsed by the World Health Organization (WHO) in 2010 [8, 9]. The WHO currently

recommends the use of the next-generation Xpert MTB/RIF Ultra (with improved sensitivity for TB detection) as the initial test to be used for all persons being evaluated for TB [10, 11]. However, Xpert has not been deployed at most lower level health facilities due to considerable implementation barriers, including cost and infrastructure requirements, low throughput, and a relatively long testing turnaround time in field settings [12, 13]. A 2017 report evaluating policies in 29 high TB incidence countries highlighted that only 15 (52%) have adopted a policy of “Xpert for all”, only 7 (47%) of which have widely implemented Xpert [14]. There are some TB molecular diagnostic alternatives to Xpert that are available, such as line probe assays [15] or TB-loop-mediated isothermal amplification (TB-LAMP) [16], or those under development and evaluation, such as the Truenat MTB or AccuPower TB&MDR Real-Time PCR (for comprehensive list see FIND diagnostics pipeline) [17]. However, these tests in development and evaluation are not yet shown to be field applicable and deployable at lower level health facilities, hence the need for a test that can be used in these settings.

Triage tests are typically simple and low-cost tests aiming to improve ease of access and implementation (compared with confirmatory tests) and decrease the proportion of patients requiring more expensive confirmatory testing (ie, a rule-out test with a high sensitivity and negative predictive value) [18].

Correspondence: R. Nathavitharana, MBBS, MPH, Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Lowry Building, Suite GB, 110 Francis Street, Boston MA 02215 (rnathavi@bidmc.harvard.edu)

The Journal of Infectious Diseases® 2019;220(S3):S116–25

© The Author(s) 2019. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com. DOI: 10.1093/infdis/jiz243

A TB triage test (Box 1) is designed to be used in adults and children identified as having symptoms compatible with TB or having risk factors for any form of active TB (or at a minimum for pulmonary TB) (Figure 1). Triage testing should stratify individuals for either confirmatory TB diagnostic testing (for triage test-positive patients) or further investigation of likely non-TB aetiologies (for triage test-negative patients). The key characteristics defined for a TB triage test at a WHO consensus meeting to develop target product profiles (TPPs) for new TB diagnostic tests in 2014 [19] were that it should be as follows: non-sputum based; easy to use; rapid; accurate (optimally 95% sensitive and 80% specific for any form of active TB when compared with the confirmatory test, or minimally 90%

sensitive and 70% specific for pulmonary TB when compared with the confirmatory test); affordable; and usable with only minimal infrastructure and training needs. An optimized triage test for TB would likely have a large global market and high potential to reduce TB burden [20].

The current recommended approach to identifying patients presenting to healthcare facilities who should be evaluated for TB primarily relies on (1) patients reporting symptoms compatible with TB (cough, hemoptysis, fever, night sweats, weight loss) through passive case finding and/or (2) systematic symptom screening of individuals with certain risk factors for TB such as human immunodeficiency virus (HIV) [21]. Patients who self-report symptoms or have a positive symptom screen should next undergo confirmatory testing with Xpert or other WHO-endorsed molecular diagnostic tests where feasible and/or sputum smear microscopy when molecular tests are unavailable onsite. Of note, chest x-ray (CXR) typically remains a secondary diagnostic test after an initial negative sputum-based test and is often unavailable in community and many primary healthcare settings. The accuracy of symptom screening for pulmonary TB is highly variable, with one review demonstrating a sensitivity ranging from 25% to 50% for prolonged cough (longer than 2 weeks) to 77%–84% for any TB symptom. Specificity dropped from around 92%–95% for prolonged cough to around 67%–74% for any TB symptom [22]. In addition, studies that have evaluated patients using exit interviews after presentation to healthcare clinics in high TB incidence settings have demonstrated that, although a high proportion of clinic attendees reported 1 or more TB compatible symptoms (approximately 5%–15% of whom will have TB), only a subset of those who were identified as having TB symptoms (between 10%–25%) actually underwent sputum-based TB diagnostic testing, resulting in a substantial early diagnostic gap in the cascade of care [23–26].

Prior evaluations of TB diagnostics, including the existing but suboptimal tests used for triage (eg, CXR), have often identified a lack of rigor with respect to sources of bias related to patient sampling, study design, and issues related to reference standard [27, 28] (see also Denking et al, Paper 1). In this article, we will focus on providing guidance for the design of diagnostic accuracy studies of novel TB triage tests, in view of the urgent need for these data to inform WHO review and guide potential policy recommendations. We will review general study design considerations and other key issues including each of the Quality Assessment of Diagnostic Accuracy Study (QUADAS-2) domains [29]. Impact data beyond accuracy, although not the focus of this article, will also be discussed.

INTENDED USE SCENARIOS FOR TRIAGE TESTS

The cost efficiencies offered by introduction of a triage test (with respect to decreased numbers of patients requiring confirmatory testing) may improve case finding by reducing health system overload. This may make it more feasible and affordable

Box 1. Definitions of tests and risk factors.

Triage test: a test that can be used in people presenting to a health facility and reporting one or a combination of symptoms compatible with TB (cough, fever, night sweats, weight loss, chest pain, haemoptysis) or in those with risk factors for TB (such as HIV or those who have had contact with someone who had infectious TB) to determine those who need confirmatory TB testing.

Confirmatory test: a TB diagnostic test that provides a definitive diagnosis of TB. This will typically be Xpert or other WHO endorsed confirmatory tests such as mycobacterial culture. Based on the results of confirmatory testing, TB treatment can be initiated.

Comparator test: this is a test or procedure that is comparative to the index test (which in this case is a triage test). This could consist of an alternative triage test or the standard of care without triage testing.

Risk factors for TB:

- People previously treated for TB
- Household or other close/prolonged contacts
- People with an untreated fibrotic lesion identified by chest radiography
- People living with HIV and people attending HIV testing
- People with diabetes mellitus
- People with chronic respiratory disease and smokers
- Undernourished people
- People with gastrectomy or jejunio-ileal bypass
- People with an alcohol-use disorder and people who use intravenous drugs
- People with chronic renal failure
- People who are immunocompromised or are having treatments that compromise their immune system
- Elderly people
- People in mental health clinics or institutions
- People in high risk congregate settings such as prisons or those who are homeless

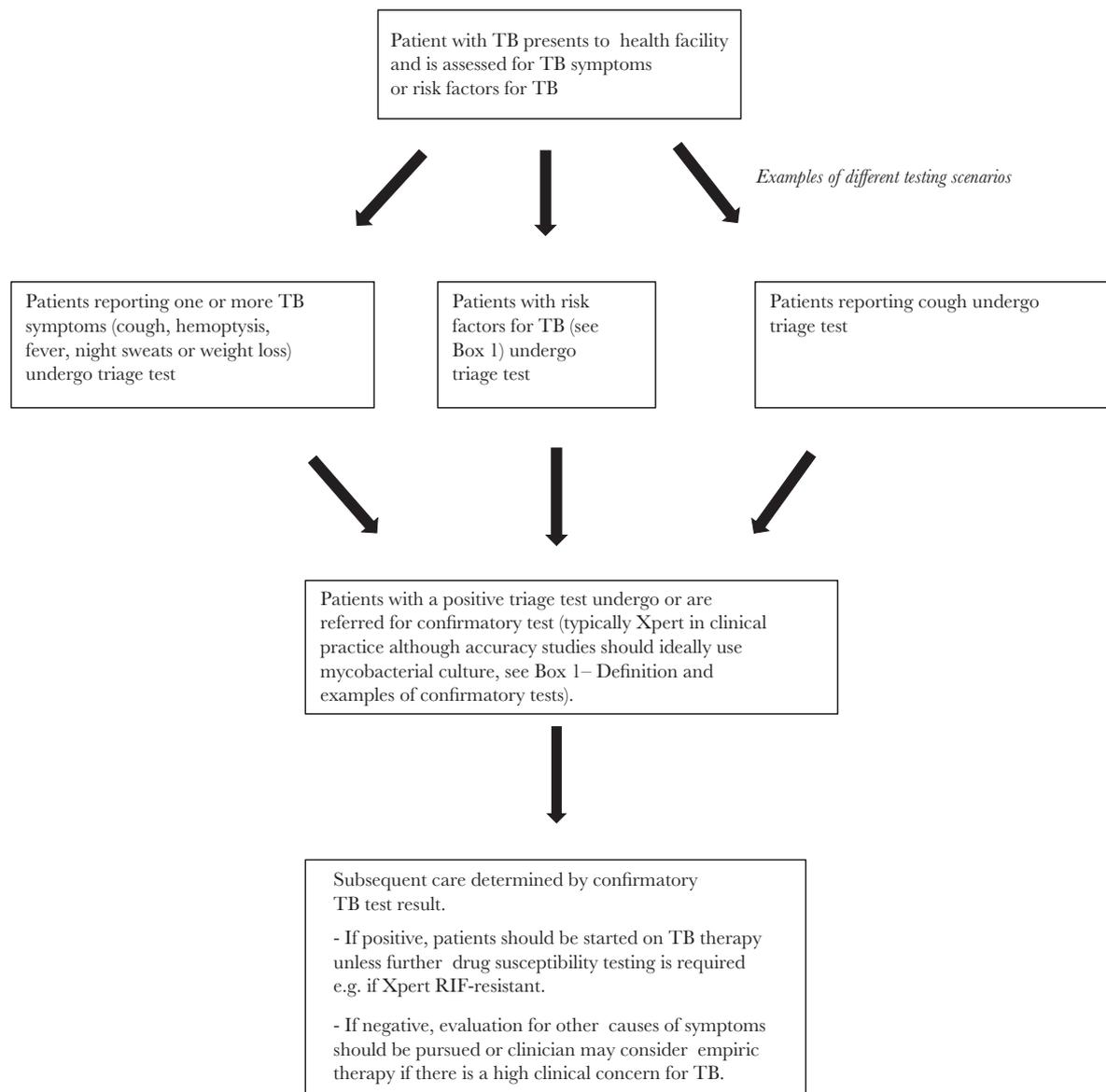


Figure 1. Clinical pathway for tuberculosis (TB) triage test.

to offer confirmatory testing to patients with a higher pretest probability of TB, based on presenting with TB symptoms or TB risk factors and having a positive triage test. If a triage test can be implemented at lower levels of the healthcare system, triage testing may also facilitate earlier diagnosis by expanding TB case finding in these settings [19], because access to triage test should be greater than for the confirmatory test. Triage testing might also be used for initial testing at higher levels of the healthcare system, such as tertiary or reference hospitals, as part of a transmission control screening approach such as FAST (Find cases Actively, Separate safely and promptly Treat effectively) [30] to identify infectious patients. Although access to confirmatory TB testing is likely to be better in these settings, high rates of initially undiagnosed TB have been documented in hospitals, including referral centers [31].

CLINICAL PATHWAY

Triage testing would be used to determine which patients with symptoms or risk factors for TB require confirmatory testing, which should be performed in all triage test-positive patients (Figure 1). Identification of patients with TB symptoms or risk may be done passively by patient self-reporting or by active symptom screening, either at the point of entry to the healthcare facility (often by a triage nurse) or by the clinician seeing the patient (including outreach workers evaluating household TB contacts). As mentioned previously, the accuracy of TB symptom screening across settings is highly variable and will thus affect the impact of a triage test being performed for patients identified to have symptoms (trriage test sensitivity will likely be lower for active rather than passive case

finding). Given the limitations of symptom screening, some patients whom clinicians consider sufficiently high-risk may be referred directly for triage (versus confirmatory) testing, even in the absence of symptoms [32]. We acknowledge that testing patients with TB risk factors but without TB symptoms may blur boundaries between triage testing (typically done in symptomatic patients) and active screening (typically done in asymptomatic people), but, for the purposes of this manuscript, we consider an initial, nonconfirmatory test in such patients to be included as a triage test.

CURRENT LANDSCAPE OF TUBERCULOSIS TRIAGE TESTS

No optimal triage test for TB currently exists. After symptom screening, CXR is currently the most commonly used triage test. However, the rollout of CXR has been largely limited by the infrastructure and training needs required. Although CXR has typically been evaluated by human readers, it is increasingly being assessed by computer-assisted detection (CAD) software [33, 34]. The CAD software performance has varied widely compared with a microbiological reference standard: sensitivity (47%–100%) and specificity (23%–94%) [28]. In a simulated algorithm, based on data from South Africa, CXR with CAD software as a triage test before Xpert in patients with suspected TB (based on self-presentation with symptoms) resulted in decreased costs per screened case (\$6.72 versus \$13.09 for patients tested with Xpert alone) and increased throughput from 45 to 113 patients per day [35]. The interest in an easy-to-use, nonsputum biomarker-based test has been substantial, and interest has focused primarily on host biomarkers. A recent systematic review and meta-analysis suggested that point-of-care C-reactive protein (CRP) may be able to meet TPP performance targets [36], and a large study in Uganda demonstrated 89% sensitivity (95% confidence interval [CI], 83–93) and 72% specificity (95% CI, 69–75) for culture-confirmed TB in people living with HIV (PLHIV) in a clinic-based setting [37]. One consortium reported a 7-marker serum protein-based biosignature that had a sensitivity of 94% and specificity of 73% in a training cohort of 491 participants with symptoms of pulmonary TB [38]. A more recent proteomic analysis on 1470 specimens from patients with symptoms and signs suggestive of pulmonary TB revealed a 6-marker-based protein signature that had a sensitivity of 90% and specificity of 80% [39]. Alternatives, including exhaled breath tests that measure volatile organic compounds (which are altered in disease states such as TB), have thus far not met triage test criteria [40].

GENERAL STUDY DESIGN CONSIDERATIONS

Diagnostic test accuracy of TB triage tests should be assessed in cross-sectional or cohort studies by evaluating a consecutive series or a random sample of patients with symptoms of

TB or risk factors for TB who are attending healthcare facilities. Using healthy controls and/or patients with severe disease can introduce spectrum bias, which may overestimate test accuracy. When designing triage test accuracy studies, investigators should ensure data can be reported according to the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines [41].

Well characterized specimens from patients being evaluated for TB symptoms are critical for the development of triage tests. Banked specimens can also play an important role to supplement data collected from prospective evaluations. However, few sample banks have specimens from patients presenting to peripheral healthcare settings, which may introduce spectrum bias, because specimens may be more likely to originate from patients with higher disease severity. Thus, studies demonstrating reproducibility of test results on fresh and banked specimens are needed to confirm data based on studies with banked specimens.

The choice of sample size is a critical consideration for any study that aims to be informative in its own right (ie, outside of a systematic review and meta-analysis). Figure 2 shows a plot of the precision of accuracy estimates as a function of sample size for sensitivity and specificity in line with the TPP minimal requirements. Although sample size planning needs to account for a multitude of factors, reasonable precision (CI width of 10%–15%) of the sensitivity estimate can be achieved with ~100 TB patients (this assumes that a given test being evaluated meets TPP minimal requirements and that standard methods are used to estimate CI based on sample size and point estimate), with additional precision gains requiring significant additional enrollment. At a TB prevalence of 5% versus 10% versus 20% (reflective of different healthcare settings) [42, 43], this would require enrolling 1500 versus 1000 versus 500 presumptive TB patients depending on the setting, to ensure that sensitivity estimates are reasonably precise.

POPULATION AND SETTING

The study populations selected for triage test evaluation studies should reflect the target populations in settings of intended use. The key initial study population for the evaluation of triage tests may often be adults (≥ 15 years, including PLHIV) with symptoms suggestive of pulmonary TB that include cough who are able to produce a sputum specimen. This group is a useful starting point because there is a reasonably strong reference standard (liquid culture) and it is the group that is also already prioritized in international guidelines for infection control purposes. Important subgroups and additional patient groups include patients being evaluated for paucibacillary/smear-negative TB (more common in PLHIV), extrapulmonary TB, and/or pediatric TB. We acknowledge that certain triage tests may be more applicable to children than adults, in which case evaluation in children may occur before or concurrently with

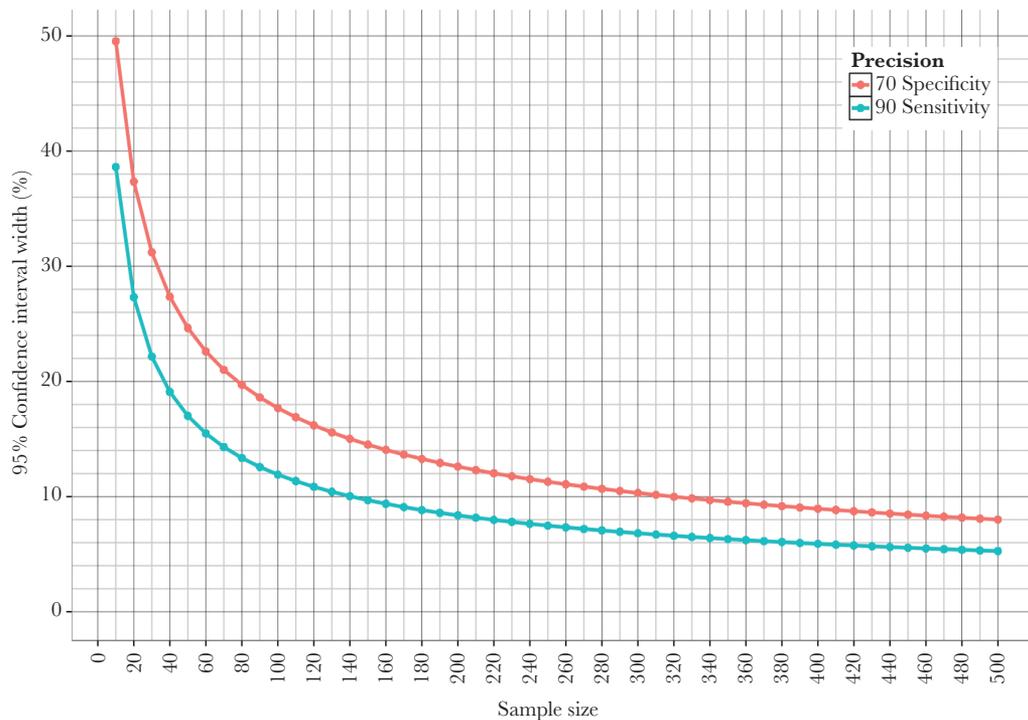


Figure 2. Precision of accuracy estimates as function of sample size. The lines show the precision of accuracy estimates as function of sample size; accuracy point estimates are chosen according to the minimal target based on the target product profile, ie, sensitivity for TB: 90% (blue line) and specificity: 70% (red line). The y-axis shows total width of the 95% confidence interval (CI), ie, upper limit of the 95% CI minus the lower limit of the 95% CI for sensitivity and specificity for a given sample size. The x-axis shows the necessary number of patients with TB to achieve a given precision for the sensitivity estimate and the number of patients without TB to achieve a given precision for the specificity estimate. Of note, this figure should serve to highlight that sample size calculations always represent a reasonable compromise between precision of estimates versus the need to recruit more participants, because an optimal sample size is often hard to achieve.

evaluation in adults. To expand diagnosis to other forms of TB beyond pulmonary disease, the test would ideally need to be nonsputum based. Specimens with the greatest potential are blood (venous or preferably fingerstick), urine, and breath.

Patient enrollment and testing should ideally be performed in the primary settings of intended use, ie, L0: community health outposts, L1: primary healthcare centers, and L2: district hospitals, where other TB testing modalities and potentially L2 district hospitals are often not present or only to a minimal extent [7]. Because the reference standard tests are often not available for routine clinical use in these settings, it may be that an initial evaluation takes place in a more centralized location such as a hospital-based outpatient department. However, assessments that include the more challenging settings of intended use (ie, L0/1) are essential to investigate robustness and ease of use of the triage test by nonlaboratory personnel.

The number and choices of study sites for triage test evaluation should ideally be based on the distribution of factors that are known, or hypothesized, to lead to variability in triage test performance. This will vary according to the specific triage test being evaluated, but such factors may include host variability (eg, immune system status), comorbidities (such as HIV coinfection, type-2 diabetes), or environmental conditions (such as temperature, humidity, dust, nontuberculous

mycobacterial exposure) that may vary by geographic location. Initial assessments regarding implementation feasibility and triage test performance variability would ideally be incorporated into early diagnostic accuracy studies (and considered at the design stage) and subsequently be assessed in greater detail in dedicated implementation studies.

INDEX TEST

Triage test evaluation studies should clearly report how the index test (the test under investigation) is performed (administration, interpretation, and setting). Reporting should include indeterminate or invalid results and instrument failures. If the assay readout is not automated and requires a degree of subjective interpretation (eg, visual reader), cutoffs for positivity must be prespecified and readers of index-test results must be blinded to results of the reference standard and other tests. Interreader reliability also needs to be assessed.

Triage test developers and evaluators will also need to consider special issues pertaining to tests that use machine learning techniques to classify patients. Machine learning may be used to analyze the results generated from x-ray patterns detected by CAD or breath biomarkers to provide a test score. Test scores, above which a result is labeled as “test-positive”, are usually determined based on fixed probability thresholds. However,

some test developers have proposed the use of variable (eg, population-specific) thresholds, which will pose regulatory and feasibility issues. In either case, prespecification of thresholds (whether a single threshold or multiple) is essential for late-stage studies aiming to provide unbiased estimates of sensitivity and specificity, particularly pertaining to tests as they might be used in actual practice [44]. It is possible that there will not be a single threshold appropriate for all use cases because these may be chosen depending on the epidemiological characteristics and resources available in a given healthcare setting where triage testing is used. Biomarker-based tests that use probability thresholds should prespecify and report the thresholds used.

REFERENCE STANDARD AND COMPARATORS

We recommend that mycobacterial culture (with speciation) should be used as the primary reference standard for diagnostic test accuracy evaluations of TB triage tests. This should be performed on commercial liquid media (MGIT), alone or in addition to solid culture (Lowenstein Jensen or Middlebrook 7H10 or 7H11 agar), in line with the WHO recommendation that low- and middle-income countries should implement liquid culture systems [45]. However, we suggest that triage test diagnostic accuracy results should also be compared with the confirmatory test used in practice at a given setting (typically Xpert), as has been done in some prior studies [37].

It is conceivable that biomarker-based triage tests detect early or incipient TB [46], which may be culture negative. For this form of TB and other forms such as pediatric, extrapulmonary TB, and TB in PLHIV, sputum is difficult to obtain and, even if available, sputum culture alone is an imperfect reference standard. Attempts should be made for the reference standard to include extrapulmonary sampling. The use of a composite reference standard should be considered, particularly in populations with a higher likelihood of early, incipient, extrapulmonary, and/or paucibacillary disease. Follow-up of culture-negative patients that are not started on TB therapy empirically should ideally be performed, to detect those who may become culture-positive subsequently. In addition, studies should work with clinicians to standardize the approach to deciding when to start empiric therapy using predefined criteria. Researchers may consider the use of several reference standards for analysis, for example, using clinical or composite reference standards, or more sophisticated estimation approaches such as latent class analysis [47] or sensitivity analysis including different reference standards for analysis (see also Drain et al, Paper 3).

When conducting diagnostic test accuracy evaluations for new TB triage tests, it is also important to consider relevant comparator tests (Box 1) when and where available, to determine the additional contribution of the new test. As mentioned earlier, CRP could be considered as a comparator test for triage tests designed for PLHIV [37]. Of note, studies comparing the

diagnostic accuracy of 2 or more tests will require a much larger sample size to detect small differences in sensitivity and/or specificity.

FLOW AND SPECIMEN ISSUES

Depending on the type of triage test being evaluated, rigorous attention to sample type and test throughput should be part of the diagnostic test evaluation. Triage testing, as well as reference standard testing, should ideally be performed on the same day and always before treatment initiation because this could influence reference standard and index test results. Because mycobacterial culture is not typically available at the primary settings of intended use, courier systems should be set up such that specimens can be transported to the site where reference testing can be performed on the same day to minimize bias as far as possible.

For nonsputum-based tests, other issues may arise depending on the type of specimen, such as blood sample volume restrictions, issues with sample storage and biomarker stability, and sample transportation. If the triage test is sputum-based, considerations may include evaluation of the test performance using expectorated versus induced sputum specimens (see also Schumacher et al, Paper 2). In addition, if a test is sputum-based, performing the index test and reference standard(s) on the same specimen enables the most direct comparison of accuracy, although the potentially large sputum volume required may be prohibitive (see also Schumacher et al, Paper 2).

KEY ISSUES BEYOND ACCURACY

It is important to emphasize that triage tests are expected to offer advantages, such as cost, feasibility, acceptability, and scalability, that are not captured by evaluations that solely evaluate accuracy. The primary advantage of a triage test is that it would expand TB diagnostic test capability to lower levels of the healthcare system and allow improved targeting of patients that require confirmatory testing. The decision to implement a triage test at the lowest levels of the healthcare system or in patients who do not report symptoms but have risk factors for TB will depend on its characteristics including feasibility of implementation, cost, and potential variation in test performance. Factors such as access to confirmatory testing, and whether this requires transporting patients or specimens, will also affect the likelihood of successful implementation and impact of a triage test at a given level of the healthcare system. For subsequent impact or cost-effectiveness evaluations, complete algorithms should be compared (eg, with triage test versus without triage test) [48]. Studies looking beyond accuracy and directly evaluating the effects of implementing a triage test are important to assess whether expectations about the impact of triage tests [22] hold in practice. Implementation studies should evaluate process indicators that may be affected by the use of triage testing such as the number of patients presenting with symptoms who undergo triage testing,

the number who receive confirmatory testing, the number testing positive for TB, and the time to TB diagnosis. It must be remembered that improving patient-centered outcomes, such as time to effective treatment initiation, treatment completion, cure, and mortality (and subsequently population metrics such as annual risk of infection), relies on improved linkage to and retention in care. More importantly, a nonsputum-based triage test such as CRP could also expand the diagnostic net to patients who are currently unable to produce sputum, including some PLHIV as well as other groups such as children and those with extrapulmonary TB, but overall diagnostic yield is limited by the performance of existing nonsputum-based confirmatory tests. In an ideal setting, a triage test would not serve as a confirmatory test in patients unable to produce sputum, but the lack of reliable and validated biomarker-based nonsputum diagnostic

tests remains yet another gap in the TB diagnostic landscape (see Drain et al, Paper 3). Avoidance of sputum-based sampling in the L0 and L1 healthcare settings may also help to reduce the risk of transmission to healthcare workers. Some triage tests may point to diagnoses other than TB, for example, CXR-based tests may reveal other pulmonary diseases, or a multiplexed assay could reveal other diagnoses such as HIV or malaria; thus, triage testing could potentially improve diagnosis and patient care more broadly. Triage testing also has the potential to reduce costs, both due to the reduced number of confirmatory tests needed as well as the number of triage test-negative patients who would not require potential travel to sites where confirmatory testing is available. Implementation studies should evaluate some of these other potential benefits of the use of triage tests in different contexts.

Table 1. Overview of Recommendations for TB Triage Test Diagnostic Accuracy Evaluations Grouped by QUADAS Domains

Topic	Recommendation
General Study Design	<ul style="list-style-type: none"> Use a cross-sectional or cohort study enrolling a consecutive series or random sample of patients who require evaluation for TB (avoid using patients with known severe disease or healthy controls, because this introduces spectrum bias and can overestimate test accuracy) Banked specimens may play an important role to supplement data collected from prospective evaluations (while recognizing the possibility of spectrum bias) Consider how many reference standard positive and negative samples are required to obtain a precise estimate of the sensitivity and specificity respectively (sample size calculations should take into account factors including TB prevalence) Refer to the STARD (Standards for Reporting of Diagnostic Accuracy) guidelines in addition to the more detailed advice pertaining to TB triage test evaluation in this article
Population and Setting	<ul style="list-style-type: none"> Avoid selecting patients in whom TB has already been diagnosed by another test or who have already started on TB treatment For initial studies focus on adults, including PLHIV, who have respiratory symptoms suggestive of TB; subsequent evaluation should include other key groups such as children and people being evaluated for extrapulmonary TB Studies should potentially enroll patients in the primary settings of intended use, ie, L0: community health outposts, L1: primary health-care centers, and L2: district hospitals Perform testing (particularly for reference standard because triage test may be a point-of-care assay) in quality assured laboratories; followed by testing in settings of intended use Provide stratified accuracy estimates for key subpopulations (by HIV-status, smear-status, presence of comorbidities such as chronic lung disease that may present with TB symptoms such as cough)
Index Test	<ul style="list-style-type: none"> Studies should report the specifics of the triage test under investigation (administration, interpretation, and setting) Indeterminate or invalid results and instrument failures should be reported If a test has a nonautomated readout, blinding is essential to make sure the index test is interpreted independently of the reference test or comparators
Reference Standard and Comparators	<ul style="list-style-type: none"> Use automated liquid mycobacterial culture as the primary reference standard Studies may also compare triage tests to the confirmatory test used in practice at the setting where the test is being evaluated (eg, Xpert or other WHO-endorsed molecular diagnostic tests), but this should optimally be done in addition to culture Avoid partial or differential verification bias; ie, all those who received the index test should also receive the same reference standard Include clinical case definition, additional measures, as well as follow-up, to understand discordant (triage-test-positive, culture-negative) results Studies may compare triage tests to other comparator triage tests such as CXR or CRP, but these should be done in addition to the reference standard
Flow and Specimen Issues	<ul style="list-style-type: none"> Studies should carefully design and report the sample flow and specimen processing The triage test and reference standard should ideally be performed on the same day (and same specimen if the triage test is sputum-based) For tests that use machine learning techniques, test results may be based on probability thresholds. Prespecification of thresholds (whether a single or multiple) is essential for late-stage studies aiming to provide unbiased estimates of sensitivity and specificity
Key Issues Beyond Accuracy	<ul style="list-style-type: none"> Test characteristics other than diagnostic accuracy, such as cost, feasibility, acceptability, and scalability, are often not captured by evaluations that solely evaluate accuracy but are critical and need to be evaluated systematically Implementation studies should evaluate factors such as the testing infrastructure, which includes access to confirmatory testing, and whether this requires transporting patients or specimens, as well as test performance in different environments (temperature, humidity, dust) Implementation studies should include process indicators that may be affected by the use of triage testing The potential clinical and population level impact of new triage tests needs to be assessed through empirical studies, cost-effectiveness evaluations, and modeling, which should compare complete algorithms (eg, with triage test versus without triage test)

Abbreviations: CRP, C-reactive protein; CXR, chest x-ray; HIV, human immunodeficiency virus; PLHIV, people living with HIV; QUADAS, Quality Assessment of Diagnostic Accuracy Studies; TB, tuberculosis; WHO, World Health Organization.

CONCLUSIONS

A TB triage test has the potential to expand and improve TB diagnostic testing and identify at least a subset of the estimated 3.6 million so-called “missing patients” with TB who are currently not detected or notified. In this paper, we provide guidance on the design for TB triage test evaluation studies (see summary of recommendations in Table 1). Test evaluators should ensure that their studies are designed to best answer key questions regarding performance in populations and settings of intended use, to provide high-quality evidence for the development of WHO policy recommendations. Although diagnostic test accuracy is a critical step in triage test evaluation, evaluating other aspects of test implementation beyond accuracy is also essential. A TB triage test that cannot be easily implemented in L0 and L1 settings may have limited impact on earlier stages of the patient diagnostic pathway. We acknowledge that designing a triage test algorithm involves an explicit assessment of the prior probability of disease and rationalization of resources, with consequences for the patient and health system. However, an accurate triage test used to determine which patients require confirmatory testing, primarily at the initial point of contact by patients into the healthcare system, could be a critical tool to decrease the TB diagnostic gap.

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.

Disclaimer. The authors do not have a commercial or other association that might pose a conflict of interest to the publication of this work.

Financial support. R. R. N. is funded by a National Institutes of Health Career Development Award (National Institute of Allergy and Infectious Diseases [NIAID] K23 AI13264801A1) and American Society of Tropical Medicine and Hygiene Burroughs Wellcome Fellowship. She acknowledges prior funding from a Harvard Center for AIDS Research Scholar Award (NIAID 2P30AI060354-11). C. Y. is funded by a National Institutes of Health Career Development Award (NIAID K23 AI114363). P. M. is funded by the Wellcome Trust (206575).

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.

References

1. World Health Organization. Global Tuberculosis Control. WHO Report 2018. Geneva: World Health Organization; 2018.
2. Cheng S, Chen W, Yang Y, et al. Effect of diagnostic and treatment delay on the risk of tuberculosis transmission in Shenzhen, China: an observational cohort study, 1993-2010. *PLoS One* 2013; 8:e67516.
3. Cohen T, Murray M, Wallengren K, Alvarez GG, Samuel EY, Wilson D. The prevalence and drug sensitivity of tuberculosis among patients dying in hospital in KwaZulu-Natal, South Africa: a postmortem study. *PLoS Med* 2010; 7:e1000296.
4. Bates M, Mudenda V, Shibemba A, et al. Burden of tuberculosis at post mortem in inpatients at a tertiary referral centre in sub-Saharan Africa: a prospective descriptive autopsy study. *Lancet Infect Dis* 2015; 15:544–51.
5. Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC Infect Dis* 2009; 9:91.
6. Getnet F, Demissie M, Assefa N, Mengistie B, Worku A. Delay in diagnosis of pulmonary tuberculosis in low-and middle-income settings: systematic review and meta-analysis. *BMC Pulm Med* 2017; 17:202.
7. Huddart S, MacLean E, Pai M. Location, location, location: tuberculosis services in highest burden countries. *Lancet Glob Health* 2016; 4:e907–8.
8. World Health Organization. Policy Statement: Automated Real-Time Nucleic Acid Amplification Technology for Rapid and Simultaneous Detection of Tuberculosis and Rifampicin Resistance: Xpert MTB/RIF. Geneva: World Health Organization; 2011.
9. Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010; 363:1005–15.
10. World Health Organization. Automated Real-Time Nucleic Acid Amplification Technology for Rapid and Simultaneous Detection of Tuberculosis and Rifampicin Resistance: Xpert MTB/RIF Assay for the Diagnosis Of Pulmonary and Extrapulmonary TB in Adults and Children: Policy Update. Geneva: World Health Organization; 2016.
11. World Health Organization. WHO Meeting Report of a Technical Expert Consultation: Non-inferiority analysis of Xpert MTB/RIF ultra compared to Xpert MTB/RIF. Geneva: World Health Organization; 2018.
12. Albert H, Nathavitharana RR, Isaacs C, Pai M, Denkinger CM, Boehme CC. Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J* 2016; 48:516–25.
13. Theron G, Zijenah L, Chanda D, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet* 2014; 383:424–35.

14. Médecins Sans Frontières. Out of Step: TB policies in 29 countries. 3rd ed. 2017. Geneva, Switzerland: MSF Access Campaign.
15. Nathavitharana RR, Cudahy PG, Schumacher SG, Steingart K, Pai M, Denkinger CM. Accuracy of line probe assays for the diagnosis of pulmonary TB and detection of resistance to rifampicin and isoniazid: a systematic review and meta-analysis. *Eur Respir J* 2017; 49:pii: 1601075.
16. Shete PB, Farr K, Strnad L, Gray CM, Cattamanchi A. Diagnostic accuracy of TB-LAMP for pulmonary tuberculosis: a systematic review and meta-analysis. *BMC Infect Dis* 2019; 19:268.
17. FIND diagnostic pipeline. Available at: <https://www.finddx.org/dx-pipeline-status/>. Accessed 1 May 2019.
18. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006; 332:1089–92.
19. World Health Organization. High-priority target product profiles for new tuberculosis diagnostics: Report of a consensus meeting. Geneva: World Health Organization; 2014.
20. Kik SV, Denkinger CM, Casenghi M, Vadnais C, Pai M. Tuberculosis diagnostics: which target product profiles should be prioritised? *Eur Respir J* 2014; 44:537–40.
21. World Health Organization. Systematic screening for active tuberculosis: Principles and recommendations. Geneva: World Health Organization; 2013.
22. Van't Hoog AH, Onozaki I, Lonroth K. Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden. *BMC Infect Dis* 2014; 14:532.
23. Chihota VN, Ginindza S, McCarthy K, Grant AD, Churchyard G, Fielding K. Missed opportunities for TB investigation in primary care clinics in South Africa: experience from the XTEND trial. *PLoS One* 2015; 10:e0138149.
24. Claassens MM, Jacobs E, Cyster E, et al. Tuberculosis cases missed in primary health care facilities: should we redefine case finding? *Int J Tuberc Lung Dis* 2013; 17:608–14.
25. Kweza PF, Van Schalkwyk C, Abraham N, Uys M, Claassens MM, Medina-Marino A. Estimating the magnitude of pulmonary tuberculosis patients missed by primary health care clinics in South Africa. *Int J Tuberc Lung Dis* 2018; 22:264–72.
26. Roy M, Muyindike W, Vijayan T, et al. Implementation and operational research: use of symptom screening and sputum microscopy testing for active tuberculosis case detection among HIV-infected patients in real-world clinical practice in Uganda. *J Acquir Immune Defic Syndr* 2016; 72:e86–91.
27. MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. Author Correction: a systematic review of biomarkers to detect active tuberculosis. *Nat Microbiol* 2019; 4:899.
28. Pande T, Cohen C, Pai M, Ahmad Khan F. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. *Int J Tuberc Lung Dis* 2016; 20:1226–30.
29. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155:529–36.
30. Barrera E, Livchits V, Nardell E. F-A-S-T: a refocused, intensified, administrative tuberculosis transmission control strategy. *Int J Tuberc Lung Dis* 2015; 19:381–4.
31. Nathavitharana RR, Daru P, Barrera AE, et al. FAST implementation in Bangladesh: high frequency of unsuspected tuberculosis justifies challenges of scale-up. *Int J Tuberc Lung Dis* 2017; 21:1020–5.
32. Yoon C, Dowdy DW, Esmail H, MacPherson P, Schumacher SG. Screening for tuberculosis: time to move beyond symptoms. *Lancet Respir Med* 2019; 7:202–4.
33. Ahmad Khan F, Pande T, Tessema B, et al. Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy. *Eur Respir J* 2017; 50:1700953. doi: 10.1183/13993003.00953-2017.
34. Hwang EJ, Park S, Jin KN, et al. Development and validation of a deep learning-based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. *Clin Infect Dis* 2018; Nov 12. doi: 10.1093/cid/ciy967.
35. Philipsen RH, Sánchez CI, Maduskar P, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. *Sci Rep* 2015; 5:12215.
36. Yoon C, Chaisson LH, Patel SM, et al. Diagnostic accuracy of C-reactive protein for active pulmonary tuberculosis: a meta-analysis. *Int J Tuberc Lung Dis* 2017; 21:1013–9.
37. Yoon C, Semitala FC, Atuhumuza E, et al. Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *Lancet Infect Dis* 2017; 17:1285–92.
38. Chegou NN, Sutherland JS, Malherbe S, et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax* 2016; 71:785–94.
39. De Groote MA, Sterling DG, Hraha T, et al. Discovery and validation of a six-marker serum protein signature for the diagnosis of active pulmonary tuberculosis. *J Clin Microbiol* 2017; 55:3057–71.
40. Phillips M, Basa-Dalay V, Blais J, et al. Point-of-care breath test for biomarkers of active pulmonary tuberculosis. *Tuberculosis (Edinb)* 2012; 92:314–20.

41. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* **2015**; 351:h5527.
42. Nliwasa M, MacPherson P, Gupta-Wright A, et al. High HIV and active tuberculosis prevalence and increased mortality risk in adults with symptoms of TB: a systematic review and meta-analyses. *J Int AIDS Soc* **2018**; 21:e25162.
43. Datiko DG, Guracha EA, Michael E, et al. Sub-national prevalence survey of tuberculosis in rural communities of Ethiopia. *BMC Public Health* **2019**; 19:295.
44. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clin Chem* **2008**; 54:729–37.
45. World Health Organization. Policy statement: liquid media for culture and DST. Geneva: World Health Organization; **2007**.
46. Esmail H, Lai RP, Lesosky M, et al. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18F]fluoro-D-glucose positron emission and computed tomography. *Nat Med* **2016**; 22:1090–3.
47. 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.
48. Yoon C, Semitala FC, Asege L, et al. Yield and efficiency of novel intensified tuberculosis case-finding algorithms for people living with HIV. *Am J Respir Crit Care Med* **2019**; 199:643–50.