Diagnostic yield as an important metric for the evaluation of novel tuberculosis tests: rationale and guidance for future research

Tobias Broger, Florian M Marx, Grant Theron, Ben J Marais, Mark P Nicol, Andrew D Kerkhoff, Ruvandhi Nathavitharana, Helena Huerga, Ankur Gupta-Wright, Mikashmi Kohli, Brooke E Nichols, Monde Muyoyeta, Graeme Meintjes, Morten Ruhwald, Rosanna W Peeling, Nitika Pant Pai, Nira R Pollock, Madhukar Pai, Adithya Cattamanchi, David W Dowdy, Puneet Dewan, Claudia M Denkinger

Better access to tuberculosis testing is a key priority for fighting tuberculosis, the leading cause of infectious disease deaths in people. Despite the roll-out of molecular WHO-recommended rapid diagnostics to replace sputum smear microscopy over the past decade, a large diagnostic gap remains. Of the estimated 10·6 million people who developed tuberculosis globally in 2022, more than 3·1 million were not diagnosed. An exclusive focus on improving tuberculosis test accuracy alone will not be sufficient to close the diagnostic gap for tuberculosis. Diagnostic yield, which we define as the proportion of people in whom a diagnostic test identifies tuberculosis among all people we attempt to test for tuberculosis, is an important metric not adequately explored. Diagnostic yield is particularly relevant for subpopulations unable to produce sputum such as young children, people living with HIV, and people with subclinical tuberculosis. As more accessible non-sputum specimens (eg, urine, oral swabs, saliva, capillary blood, and breath) are being explored for point-of-care tuberculosis testing, the concept of yield will be of growing importance. Using the example of urine lipoarabinomannan testing, we illustrate how even tests with limited sensitivity can diagnose more people with tuberculosis if they enable increased diagnostic yield. Using tongue swab-based molecular tuberculosis testing as another example, we provide definitions and guidance for the design and conduct of pragmatic studies that assess diagnostic yield. Lastly, we show how diagnostic yield and other important test characteristics, such as cost and implementation feasibility, are essential for increased effective population coverage, which is required for optimal clinical care and transmission impact. We are calling for diagnostic yield to be incorporated into tuberculosis test evaluation processes, including the WHO Grading of Recommendations, Assessment, Development, and Evaluations process, providing a crucial real-life implementation metric that complements traditional accuracy measures.

Introduction

Tuberculosis diagnosis has relied on low sensitivity sputum smear microscopy for more than 100 years. In 2022, of the estimated 10·6 million people who developed tuberculosis, 3·1 million were not diagnosed and reported.[1] The persistent tuberculosis diagnostic gap is closely associated with the inability of countries to reach the WHO standard of universal access to rapid molecular tuberculosis diagnostics.[2] In 2022, for example, only 47% of patients diagnosed with tuberculosis were initially tested with a WHO-recommended rapid diagnostic test.[1,3] Constraints in diagnostic access are central to these institutional failures, spanning from being identified as needing a test, obtaining specimens, and receiving testing to starting and completing treatment.[4]

Figure A illustrates key moments in the past century that have been instrumental in defining how tuberculosis diagnostics are evaluated. Since 2007, WHO has been applying the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE)[23] process to guideline development. GRADE prioritises randomised controlled trials that directly evaluate the effect of a diagnostic test on patient-important outcomes in real-life conditions.[24] However, in the absence of direct evidence from randomised controlled trials, WHO's guideline development groups usually link accuracy studies to patient-important outcomes, such as cure, mortality, time to diagnosis, and time to treatment, and integrate these into GRADE's evidence to decision framework to infer the probable effects of tests and develop recommendations.[25] If a test is not likely to improve patient-important outcomes or population health, then the health-care system has no reason to use it, regardless of its accuracy.

A new pipeline of tuberculosis diagnostics is emerging, in part as a dividend from the massive diagnostic investments made during the COVID-19 pandemic.[26] While traditional accuracy metrics and individual patient-important outcomes remain critical, population health-focused measures are equally important in defining how diagnostics can be deployed to achieve public health objectives. Such measures are less emphasised within GRADE, and accordingly recommendations for new diagnostics could fail to focus on the potential to identify more people with tuberculosis and close the diagnostic gap.[27] In this Health Policy, we aim to emphasise the importance of diagnostic yield for emerging tuberculosis diagnostics and how yield relates to effective population coverage.

Diagnostic yield

Since the mid-20th century, diagnostic yield has become a metric in evaluating the use of various diagnostic tests and procedures across different medical specialties. Many of the initial publications focused on cancer screening, but it is increasingly used in infectious diseases as well.[10,12–15,17–19,28–31]

Lancet Glob Health 2024; 12: e1184–91

Department of Infectious Disease and Tropical Medicine (T Broger MSc, F M Marx MD, A Gupta-Wright MD, Prof C M Denkinger MD) **and German Center for Infection Research** (Prof C M Denkinger), **Heidelberg University Hospital, Heidelberg, Germany; DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Faculty of Science, Stellenbosch University, Stellenbosch, South Africa** (F M Marx); **DSI-NRF 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** (Prof G Theron PhD); **The University of Sydney Infectious Diseases Institute, Sydney, NSW, Australia** (Prof B J Marais MD); **Children's Hospital at Westmead, Sydney, NSW, Australia** (Prof B J Marais); **Division of Infection and Immunity, School of Biomedical Sciences, University of Western Australia, Perth, WA, Australia** (Prof M P Nicol MD); **Division of HIV, Infectious Diseases, and Global Medicine, Zuckerberg San Francisco General Hospital and Trauma Center, San Francisco, CA, USA** (A D Kerkhoff MD); **Center for Tuberculosis, University of California San Francisco, San Francisco, CA, USA** (A D Kerkhoff, A Cattamanchi MD); **Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Boston, MA, USA** (R Nathavitharana MBBS); **Department of Epidemiology, Epicentre, Paris, France**

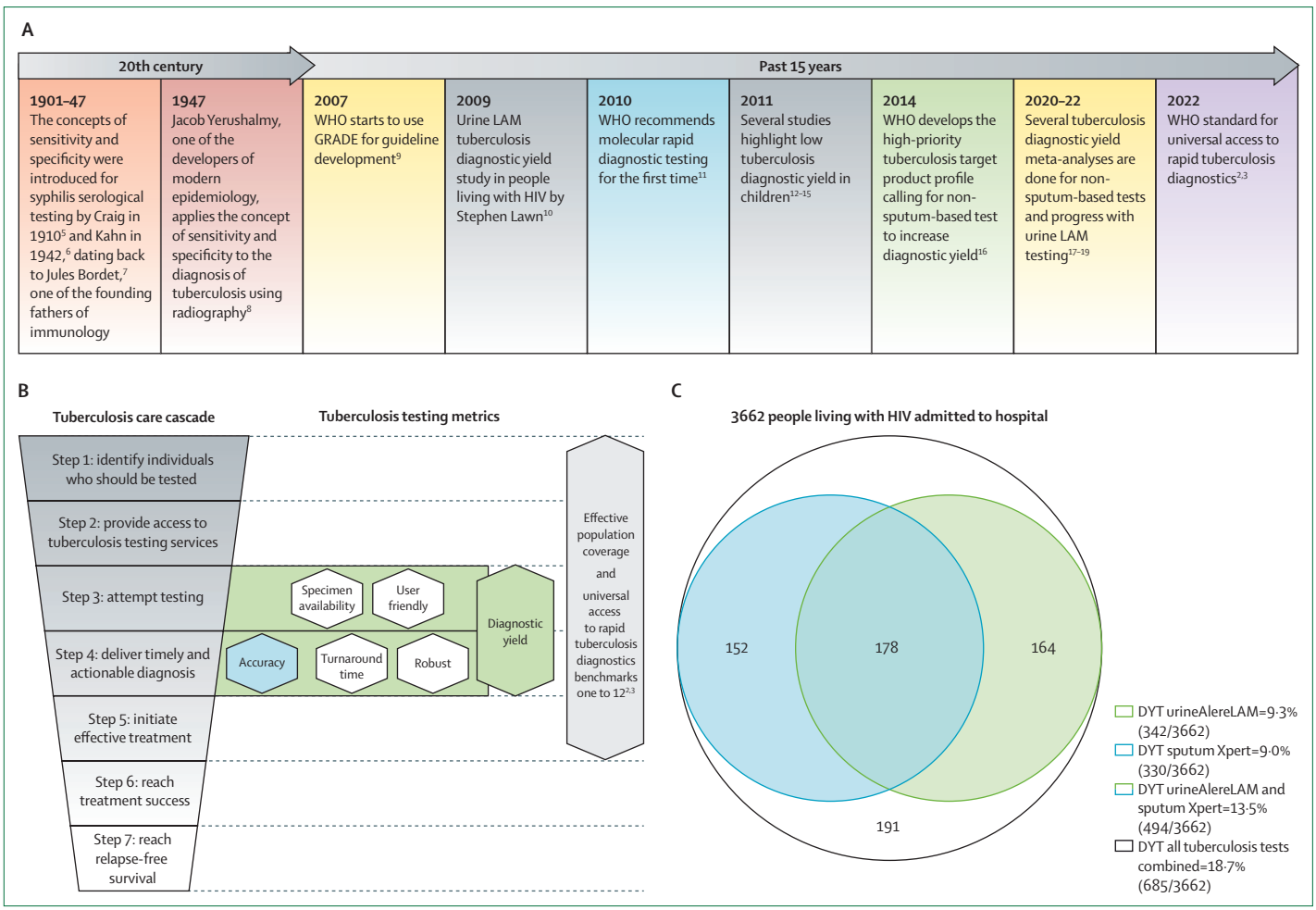

Figure: **Diagnostic yield as an important metric for the evaluation of novel tuberculosis tests**

(A) Key moments and seminal publications in the conceptualisation of diagnostic accuracy, policy development, and diagnostic yield that inform the current approach to evaluate tuberculosis diagnostics. (B) Tuberculosis testing metrics mapped to the tuberculosis care cascade. Accuracy evaluates step 4, diagnostic yield steps 3 and 4, and effective population coverage evaluates steps 1 to 5 of the care cascade. Tuberculosis testing and care metrics refer to the REASSURED criteria,[20,21] which include Ease of specimen collection, Accuracy (Sensitivity and Specificity), User-friendly, Rapid and Robust, and are considered elements of diagnostic yield. REASSURED further includes Affordable, Deliverable, Equipment-free, and Real-time connectivity, which are relevant for effective population coverage. The tuberculosis care cascade is adapted from Subbaraman and colleagues[22] and Ismail and colleagues,[3] and universal access benchmarks are described in the WHO standard.[2] (C) Comparison of tuberculosis diagnostic yield between urine lipoarabinomannan (Alere Determine TB LAM, Abbott, Chicago, IL, USA) and sputum Xpert (MTB/RIF or Ultra, Cepheid, Sunnyvale, CA, USA) among hospitalised people living with HIV from the first sample collected in the initial 2 days after enrolment.[17] The diagnostic yield was calculated as the number of patients with positive test results among all people for whom testing was attempted. DYT=Diagnostic yield among all people attempted to test. GRADE=Grading of Recommendations, Assessment, Development and Evaluations.

(H Huerga PhD); **FIND, Geneva, Switzerland** (M Kohli PhD, B E Nichols PhD, M Ruhwald PhD); **Centre for Infectious Diseases Research in Zambia, Lusaka, Zambia** (M Muyoyeta PhD); **Department of Medicine, University of Cape Town, Cape Town, South Africa** (Prof G Meintjes PhD); **Wellcome Centre for Infectious Diseases Research in Africa (CIDRI-Africa), Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa** (Prof G Meintjes); **London School of Hygiene &**

In this Health Policy diagnostic yield (DYT; panel 1) is defined as the proportion of people identified with disease using a highly specific diagnostic test (thus resulting in mostly true positives, PT), out of all people eligible to be tested (D), irrespective of adequate specimen collection. Diagnostic yield is a comprehensive measure of the performance of a test because it considers access, specimen availability, sensitivity, and test completion. Considering the diagnostic literature more broadly, the common denominator is the total number of people for whom testing is attempted, which also includes people who were unable to provide a specimen and those for whom the test failed to deliver a result (due to indeterminates, invalids, or errors, which are linked to test robustness and user friendliness). Diagnostic yield among all those diagnosed (DYD) is another definition[17–19,32–36] but only considers the diagnostic yield among those diagnosed with tuberculosis. While we subsequently use the DYT diagnostic yield definition, we still advocate for the assessment of both diagnostic yield and diagnostic yield among all those diagnosed in studies.

Usually, estimates of diagnostic yield are based on a single test attempt using a single diagnostic specimen from a single clinical encounter, as would happen in routine clinical care. Diagnostic yield could further include turnaround time such as diagnostic yield at the first clinical encounter or 24-h diagnostic yield. The concept of yield can also be extended to diagnostic algorithms involving more than one test as composite diagnostic yield.

The importance of diagnostic yield and its link to effective population coverage

Figure B maps tuberculosis testing and care metrics to the tuberculosis care cascade. The ultimate goal of a diagnostic test is to achieve population-level impact on patient outcomes, which requires effective population coverage[37–39] and universal access.[2,3] Diagnostic yield is an important part of effective population coverage as it measures a test's ability to deliver actionable positive diagnoses in those for whom testing is attempted (covering steps 3 and 4 of the tuberculosis care cascade in figure B). In addition to test accuracy, diagnostic yield covers other key aspects such as specimen availability, turnaround time, test robustness, failures, and user friendliness. All of these aspects collectively contribute to effective and timely result generation in a real-world clinical setting in high endemic countries but are not usually covered by diagnostic accuracy studies.

Panel 1: Definition of diagnostic yield

Diagnostic yield among all tested

Diagnostic yield among all of those tested is defined as the proportion of people in whom a diagnostic test identifies tuberculosis among all people for whom testing is attempted. The crucial factor is that it should mimic real-world clinical practice in a tuberculosis endemic setting.

Formula

DYT=PT/D

- DYT=Diagnostic yield among all people attempted to test
- PT=Number of people with a positive diagnosis by the test
- D=Denominator defined as the total number of people for whom tuberculosis testing is attempted

Strengths

- Simplicity in study design: the focus on positive results simplifies study design, making studies feasible even in the absence of a comprehensive reference standard that would be required to distinguish between true and false positives
- Pragmatic assessment: enables studies that replicate real-world clinical practices by considering factors such as test completion, specimen viability, and timely result availability in all people for whom testing is attempted
- Reflects real-world testing conditions: by considering a single test attempt using a single specimen from a single clinical encounter, diagnostic yield mirrors real-world testing conditions

Limitations

- Dependence on prevalence: diagnostic yield is influenced by tuberculosis prevalence, which could limit its generalisability across populations with different prevalences
- Specificity consideration: specificity requires careful consideration in light of the clinical goal, associated cost, and prevalence; diagnostic yield includes false positives in positive results; if test specificity has been well established (eg, in previous accuracy studies) and is high (ie, ≥98·5%), the effect of false positives on diagnostic yield is low, particularly if tuberculosis prevalence is high; in this case, positive results approximate the number of true positive results; diagnostic yield for tests with lower specificity should be adjusted (appendix 1) and be interpreted carefully

Diagnostic yield among all those diagnosed

Diagnostic yield among all those diagnosed is defined as the proportion of people in whom a diagnostic test identifies tuberculosis among tuberculosis positive people for whom testing is attempted. Diagnostic yield among those who are positive for tuberculosis is calculated as the number of people with a positive diagnosis by the test divided by the total number of people diagnosed with tuberculosis.

As for diagnostic yield among all those tested, diagnostic yield among all those diagnosed (DYD) could include a turnaround time component and is usually based on a single test, but can include a test series or standardised combination of tests used in real-world clinical practice.

Formula

DYD=PT/DD

- DYD=Diagnostic yield among tuberculosis positive people
- PT=Number of people with a positive diagnosis by the test
- DD=Denominator defined as the total number of people diagnosed with tuberculosis. Usually, a comprehensive microbiological reference standard that includes mycobacterial culture and a nucleic acid amplification test from any specimen including sputum, urine, blood, and other extrapulmonary samples. The participants who test positive from the index test being assessed can be included in the denominator permitted the specificity of the test is sufficiently high.

Strengths

- Not depending on prevalence: diagnostic yield among all those diagnosed is independent of tuberculosis prevalence, making it easier to compare across studies conducted in different settings or populations
- Addressing false positives: the comprehensive reference standard enables the exclusion of false positives in the count of people with a positive diagnosis
- Has been used in several tuberculosis studies

Limitations

- Requires comprehensive reference standard: diagnostic yield among all those diagnosed requires a comprehensive microbiological reference standard, with multiple tests for defining the total number of people diagnosed with tuberculosis in the denominator; this might not be feasible in all settings, particularly in pragmatic studies; even the most comprehensive reference standard could still miss people with tuberculosis

Tropical Medicine, London, UK (Prof R W Peeling PhD); **Department of Medicine, Centre for Outcomes Research & Evaluation** (N P Pai MD) **and McGill International TB Centre** (Prof M Pai MD), **McGill University, Montreal, QC, Canada; Boston Children's Hospital, Boston, MA, USA** (N R Pollock MD); **Department of Medicine, Division of Pulmonary Diseases and Critical Care Medicine, University of California Irvine, Irvine, CA, USA** (A Cattamanchi); **Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA** (Prof D W Dowdy MD); **Bill & Melinda Gates Foundation, Seattle, WA, USA** (P Dewan MD)

Correspondence to:
Prof Claudia M Denkinger, Department of Infectious Disease and Tropical Medicine, Heidelberg University Hospital, Heidelberg 69120, Germany
claudia.denkinger@uni-heidelberg.de

See **Online** for appendix 1

Specimen availability is particularly relevant for key subpopulations who are often unable to produce sputum, such as people living with HIV, children, people with extrapulmonary tuberculosis, and people with subclinical tuberculosis who do not exhibit overt signs and symptoms associated with active tuberculosis.[17,40]

Sputum induction could be necessary for patients who cannot expectorate, but this puts health-care workers at risk of infection and requires expertise, motivated staff, equipment, and time, potentially delaying specimen collection and time to diagnosis. Furthermore, sputum is a complex viscous sample requiring complex sample processing, which leads to longer turnaround times and high requirements on test robustness to avoid indeterminate and invalid tests. Despite these challenges, sputum remains the primary tuberculosis diagnostic specimen. Recognising the limitations of sputum-based diagnostics triggered the explicit inclusion of non-sputum specimens as a priority element of target product profiles for new tuberculosis diagnostics.[16,41,42]

In principle, tuberculosis tests with only moderate sensitivity, which use a more easily accessible specimen, have the potential to diagnose a higher number of people than a more sensitive molecular test reliant on sputum. For example, in an individual participant data meta-analysis[17] of tuberculosis testing among 3662 hospitalised people living with HIV, 69% (2531/3662) had a sputum specimen obtainable in the first 2 days, whereas 98% (3585/3662) had a urine specimen obtained in the first 2 days. Diagnostic yield was comparable, at 9·3% (342/3662) for urine Alere Determine TB LAM (AlereLAM, Abbott, Chicago, IL, USA) and 9·0% (330/3662) for sputum Xpert (MTB/RIF or Ultra, Cepheid, Sunnyvale, CA, USA; figure C). This result was obtained despite the lower sensitivity of the lipoarabinomannan test used in the studies (42% for urine AlereLAM) relative to the sputum assay (77% for sputum Xpert). The comparability in diagnostic yield was purely attributable to higher urine specimen availability. Although everyone can provide a urine sample, only a certain fraction of people can produce sputum samples.

Another example, albeit hypothetical, is to consider replacing Xpert Ultra on a single spot sputum (assuming specimen availability of 80%, sensitivity of 91%, and specificity of 98·5%) with tongue swab specimen and a nucleic acid amplification test (assuming specimen availability of 100%, sensitivity of 73%, and specificity of 98·5%). When applied to the same population, the same number of patients with tuberculosis would be identified by the two testing approaches, since the specimen collection advantage compensates for the loss in sensitivity when assessing tongue swab specimens.

An important consideration for the concept of diagnostic yield is test specificity. If test specificity has been well established in diagnostic accuracy studies and confirmed to be high (ie, ≥98·5%), then the number of positive results in a high incidence setting would be representative of the true positive cases. This simplifies diagnostic yield study design when highly specific tests are evaluated allowing more pragmatic studies, even in the absence of a comprehensive reference standard, to distinguish between true and false positives. For tests with lower specificity, assessment of diagnostic yield is more complex, and adjustment could be performed using Bayesian Latent Class Analysis (or other methods with similar properties) to allow explicit incorporation of additional information about an individual and their probability of having tuberculosis, eg, including chest x-ray results, whether a clinical diagnosis was made, or if follow-up is possible, response to therapy; and to include appropriate considerations of uncertainty about specificity estimates.[43] For reporting, the diagnostic yield can be adjusted by considering test specificity from other studies or meta-analyses (appendix 2 p 1).

Using more readily accessible specimens can enable higher diagnostic yield for tuberculosis. Paired with near-patient testing, including home-testing and self-testing, the higher diagnostic yield lays the basis for higher effective population coverage, including for disadvantaged and vulnerable groups. A diagnostic test with a high accuracy will have a limited effect on the population if it does not also have high diagnostic yield and is not widely available and used. Therefore,

See Online for appendix 2

Panel 2: **Example of the proposed two-step study approach and population, intervention, comparison, and outcome (PICO) questions to establish diagnostic accuracy and yield for a tongue swab-based molecular tuberculosis test**

Accuracy study

- Population: patients at risk of tuberculosis (with and without symptoms) presenting to a health facility
- Intervention: tongue swab-based molecular test
- Comparator: sputum-based molecular test
- Primary outcome: diagnostic accuracy (sensitivity and specificity) in reference to a microbiological reference standard (including a sputum culture and sputum molecular test)
- Secondary outcomes: proportion indeterminate, diagnostic accuracy against alternative reference standards (ie, clinical reference standard, extended microbiological reference standard, or latent class modelling),[43,45] negative predictive value, positive predictive value, turnaround time, and quality and proportion indeterminate of the reference standard methods

Diagnostic yield study

- Population: patients at risk of tuberculosis (with and without symptoms), ideally in health-care settings or community settings
- Intervention: tongue swab-based molecular test
- Comparator: sputum-based molecular test
- Outcomes: diagnostic yield defined as the proportion of people with a positive tongue swab among the total number of people for whom testing is attempted
- Secondary outcome: composite yield of both tests or test algorithms, time to diagnosis, sample provision, proportion indeterminate, effectiveness of implementation, and user-friendliness in programmatic setting with the intended user

not only should the diagnostic yield of a test be emphasised, but also, its coverage (figure B). Effective population coverage refers to the population at risk that is reached by a specific health intervention, such as a diagnostic test, and who are therefore able to benefit from it. As such it combines need, use, and quality.[39,44] For a test to reach high effective population coverage, it needs to be feasible to implement at scale, affordable, and cost-effective, and it should achieve high diagnostic yield within a clinically relevant turnaround time (figure B).

How to integrate diagnostic yield into diagnostic research

Panel 2 shows an example study approach to establish diagnostic accuracy and diagnostic yield for a tongue swab-based molecular tuberculosis test and a checklist for the design of a diagnostic yield study is shown in appendix 2 (pp 3–4). In general, the same criteria used for high-quality accuracy studies apply.[45–47] A two-step process might be most appropriate with the initial establishment of diagnostic accuracy (especially specificity) using a comprehensive and well validated reference standard from sputum (including methods that facilitate sputum production; detailed guidance on such studies is published elsewhere),[48] followed by a real-life diagnostic yield study. In some instances, a pragmatic combined effectiveness-implementation study might be useful, if accuracy and yield assessment can be incorporated into a single study.

For the diagnostic accuracy study, showing specificity against a comprehensive reference standard will be crucial to ensure that additional cases identified in the diagnostic yield study are true positives. A comprehensive reference standard might be less of a concern in the example of a swab-based molecular test that specifically detects *Mycobacterium tuberculosis* DNA (the caveat being presence of *M tuberculosis* DNA after cure).[49,50] However, if there is less confidence in the specificity of a test, as is the case for lipoarabinomannan-detecting urine-based tuberculosis assays, this consideration gains crucial importance and could be addressed by analytical means (eg, Bayesian latent class analysis) as outlined previously.[43]

Diagnostic yield studies should be designed with the necessary statistical power and similar statistical methods used for accurate sample size calculations.[45,51] For example, for a test with 20% diagnostic yield, 49 tuberculosis test positives would be required to achieve a 95% CI width of less than 10%. Assuming a prevalence of 25%, the study would require the enrolment of 246 participants, ideally without selection bias, as is often observed in studies that only enrol participants who are highly symptomatic and who can spontaneously expectorate sputum.

Lessons on diagnostic yield and effective population coverage

For several diseases, diagnostic innovation has improved both yield and coverage. The table lists diagnostic

	Use-case	Sample type	Test method	Accuracy	Diagnostic yield	Population coverage
HIV[52–56]	Community-based point-of-care testing and self-testing	Oral mucosal transudate	Rapid antibody testing	Reduced accuracy (sensitivity 98·7% and specificity 99·8%) against laboratory-based blood tests; confirmatory testing is recommended	High diagnostic yield due to accessible sample type and high acceptability	High population coverage; in 2020, 84% of people living with HIV knew their HIV status; expansive coverage was achieved by a successful roll-out funded by the UNITAID STAR Program and the Bill & Melinda Gates Foundation; between 60% and 90% of participants opted for self-testing depending on setting; self-testing increased uptake by 145%
Syphilis[57]	Screening and point-of-care testing	Finger-stick	Rapid antibody testing	Reduced accuracy (sensitivity 75–100% and specificity 65–100%) for rapid tests in minimal to no infrastructure areas compared with laboratory tests	High diagnostic yield; fingerprick capillary blood is feasible in community settings and health facilities during one encounter without laboratories; minimal infrastructure tests could alleviate 47% of disease burden despite low sensitivity and specificity	High population coverage; countries have begun using dual HIV and syphilis rapid diagnostic tests to increase effective population coverage for HIV and syphilis; syphilis rapid diagnostic tests were added to WHO's list of essential diagnostics; tests that require minimal infrastructure and return results in <2 h can alleviate 30–50% of disease burden
Malaria[58,59]	Point-of-care testing	Finger-stick	Rapid antigen testing	Reduced accuracy (62% sensitivity and 99% specificity) compared with laboratory-based molecular testing	High diagnostic yield: fingerprick capillary blood is feasible in community settings and health facilities during one encounter without laboratories	High population coverage; in 2021, 413 million rapid diagnostic tests were sold by manufacturers
SARS-CoV-2[60–66]	Screening and self-testing to identify people with infections, which effectively limits further spread	Nasal swab	Rapid antigen testing	Reduced accuracy (76% sensitivity and 98·9% specificity) against reference standard PCR; similar sensitivity for self-testing compared with professional testing	High diagnostic yield; accessible sample types (ie, self-performed anterior nasal swab); more than 80% of users found rapid antigen tests easy to perform	High population coverage; improved substantially, initially through large publicly-funded screening programmes; initially with assisted testing and further through lay self-testing

Table: **Diagnostic tests that achieved high effective population coverage**

examples from different diseases to illustrate the interplay of test characteristics to achieve effective population coverage, including yield.

For syphilis diagnosis, for example, one study indicated that improvement in the sensitivity of antenatal syphilis tests without a corresponding increase in patient return rate would not yield any substantial gains in health outcomes, highlighting the importance of turnaround time.[57] Nowadays, serological syphilis rapid diagnostic tests, which are less accurate but more accessible than laboratory diagnostics, are part of WHO's list of essential diagnostics.[67]

HIV self-testing also provides crucial lessons for the tuberculosis community. HIV rapid test uptake is negatively affected by the stigma and discrimination associated with the visibility of testing in health facilities. In 2012, when the US Food and Drug Administration approved oral self-tests for HIV, the concerns about lower accuracy of oral self-tests compared with laboratory-based tests were overridden because of their potential to expand diagnostic yield due to ease of sampling (eg, oral *vs* fingerprick blood), their potential for expanded access to testing, and thus increase in effective population coverage compared with laboratory tests. WHO's release of self-testing guidelines in 2016 catalysed the global availability, accessibility, and impact of these tests.[52] Improved population coverage was achieved in southern Africa by a successful roll-out of oral self-tests for screening followed by blood-based tests for confirmation, resulting in a reduction of the proportion of undiagnosed individuals without knowledge of their HIV serostatus from 40–50% in 2000 to 16% in 2020.[52]

For malaria, the development of antigen-based rapid diagnostic tests changed the landscape by offering accurate diagnosis while circumventing both venous blood collection and microscopy obstacles in peripheral health-care settings, including cost of equipment, time to result, and the need for skilled personnel. The first malaria rapid diagnostic tests emerged in the early 1990s,[68] and WHO held its first meeting on rapid diagnostic testing in 1999.[69] Initial adoption was hampered by variable field performance, which led WHO and other agencies to create an international quality control programme for malaria rapid diagnostic tests.[58] In the past 20 years, rapid diagnostic test testing has been substantially expanded around the world. In 2021, 413 million rapid diagnostic tests were sold by manufacturers and 262 million were distributed by national malaria programmes.[70]

The COVID-19 pandemic also showed the benefits of shifting attention from the narrow focus on test accuracy to diagnostic yield and effective population coverage to address diagnostic gaps.[27,60] Nasal rapid antigen tests achieved high diagnostic yield and effective population coverage despite their reduced sensitivity relative to nucleic acid amplification tests from nasopharyngeal swabs.[61,71,72]

Novel testing solutions for tuberculosis testing

Developing non-sputum-based rapid tests for tuberculosis presents substantial challenges due to the anatomical location of tuberculosis infection that usually involves the lung parenchyma with resultant lung pathobiology. Very few biomarkers progress from early research to tests with clinical use.[73] For pathogen markers (eg, antigens or DNA), abundance in non-respiratory specimens (such as blood or urine) is very low, complicating sensitive detection with low-cost point-of-care tests.[74] For host markers, detection has also proved challenging due to similarities in the immune response related to *M tuberculosis* infection (without disease) and active disease.[75] Nevertheless there are several exciting non-sputum tests based on urine, tongue swab, breath, blood, and stool on the tuberculosis testing horizon.[26,76–78] Notably, two non-sputum-based tests are already recommended by WHO and are available. These include the urine AlereLAM test to assist in tuberculosis diagnosis in people living with HIV and Xpert stool testing in children.[79] The greater appreciation of yield within the tuberculosis field will be crucial to increase acceptance and uptake of these existing tests and new non-sputum-based tests and will help to address major implementation gaps that continue to exist for these tests.

Conclusion

As part of urgent efforts to identify and treat people with tuberculosis who are missed and not receiving appropriate care, placing a stronger focus on diagnostic yield and effective population coverage is necessary. Using more readily accessible specimens together with novel near-patient tests, including home-testing and self-testing, will improve diagnostic yield and coverage, especially in disadvantaged and vulnerable groups who are at the greatest risk of disease development and spread. Thus, we propose the inclusion of diagnostic yield as an additional metric when evaluating the value of novel diagnostic tests for tuberculosis, as well as in the GRADE evidence synthesis process that informs WHO policy decisions.

Contributors

TB and CMD wrote the first draft and all other authors contributed text, revised, commented, and approved the final version of the manuscript.

Declaration of interests

TB reports patent applications in the field of tuberculosis detection and is a shareholder of Avelo. MP serves as an adviser for non-profits such as the Bill & Melinda Gates Foundation, FIND, WHO, and the Stop TB Partnership. All other authors declare no competing interests.

Acknowledgments

We thank Miriam Compton for reviewing the manuscript. This Health Policy was written without external funding. During the preparation of this work the authors used ChatGPT4 for proofreading and shortening sections of text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1 WHO. Global tuberculosis report 2023. 2023. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023 (accessed May 14, 2024).

2 WHO. WHO standard: universal access to rapid tuberculosis diagnostics. 2023. https://www.who.int/publications/i/item/9789240071315 (accessed May 14, 2024).
3 Ismail N, Nathanson CM, Zignol M, Kasaeva T. Achieving universal access to rapid tuberculosis diagnostics. *BMJ Glob Health* 2023; **8:** e012666.
4 Pai M, Kasaeva T, Swaminathan S. COVID-19's devastating effect on tuberculosis care—a path to recovery. *N Engl J Med* 2022; **386:** 1–3.
5 Craig CF. Observations upon the noguchi modification of the wassermann complement fixation test in the diagnosis of lues in the military service. *J Exp Med* 1910; **12:** 726–45.
6 Kahn RL. Serology in syphilis control: principles of sensitivity and specificity. With an appendix for health officers and industrial physicians. *JAMA* 1942; **120:** 724.
7 Bordet J, Gengou O. Sur l'existence de substances sensibilisatrices dans la plupart des sérums antimicrobiens. *Ann Inst Pasteur (Paris)* 1910; **15:** 289–302.
8 Yerushalmy J. Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray techniques. *Public Health Rep (1896)* 1947; **62:** 1432–49.
9 Hsu J, Brożek JL, Terracciano L, et al. Application of GRADE: making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implement Sci* 2011; **6:** 62.
10 Lawn SD, Edwards DJ, Kranzer K, Vogt M, Bekker LGG, Wood R. Urine lipoarabinomannan assay for tuberculosis screening before antiretroviral therapy diagnostic yield and association with immune reconstitution disease. *AIDS* 2009; **23:** 1875–80.
11 WHO. 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, 2013.
12 Nicol MP, Zar HJ. New specimens and laboratory diagnostics for childhood pulmonary TB: progress and prospects. *Paediatr Respir Rev* 2011; **12:** 16–21.
13 Zar HJ, Workman L, Isaacs W, et al. Rapid molecular diagnosis of pulmonary tuberculosis in children using nasopharyngeal specimens. *Clin Infect Dis* 2012; **55:** 1088–95.
14 Rachow A, Clowes P, Saathoff E, et al. Increased and expedited case detection by Xpert MTB/RIF assay in childhood tuberculosis: a prospective cohort study. *Clin Infect Dis* 2012; **54:** 1388–96.
15 Bates M, O'Grady J, Maeurer M, et al. Assessment of the Xpert MTB/RIF assay for diagnosis of tuberculosis with gastric lavage aspirates in children in sub-Saharan Africa: a prospective descriptive study. *Lancet Infect Dis* 2013; **13:** 36–42.
16 WHO. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva: World Health Organization, 2014.
17 Broger T, Koeppel L, Huerga H, et al. Diagnostic yield of urine lipoarabinomannan and sputum tuberculosis tests in people living with HIV: a systematic review and meta-analysis of individual participant data. *Lancet Glob Health* 2023; **11:** e903–16.
18 Dhana A, Hamada Y, Kengne AP, et al. Diagnostic accuracy of WHO screening criteria to guide lateral-flow lipoarabinomannan testing among HIV-positive inpatients: a systematic review and individual participant data meta-analysis. *J Infect* 2022; **85:** 40–48.
19 Barr DA, Lewis JM, Feasey N, et al. Mycobacterium tuberculosis bloodstream infection prevalence, diagnosis, and mortality risk in seriously ill adults with HIV: a systematic review and meta-analysis of individual patient data. *Lancet Infect Dis* 2020; **20:** 742–52.
20 Land KJ, Boeras DI, Chen XS, Ramsay AR, Peeling RW. REASSURED diagnostics to inform disease control strategies, strengthen health systems, and improve patient outcomes. *Nat Microbiol* 2019; **4:** 46–54.
21 Mabey D, Peeling RW, Ustianowski A, Perkins MD. Diagnostics for the developing world. *Nat Rev Microbiol* 2004; **2:** 231–40.
22 Subbaraman R, Nathavitharana RR, Mayer KH, et al. Constructing care cascades for active tuberculosis: a strategy for program monitoring and identifying gaps in quality of care. *PLoS Med* 2019; **16:** e1002754.
23 Schünemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; **336:** 1106–10.
24 Schünemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 22. The GRADE approach for tests and strategies-from test accuracy to patient-important outcomes and recommendations. *J Clin Epidemiol* 2019; **111:** 69–82.
25 Schünemann HJ, Mustafa R, Brozek J, et al. GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol* 2016; **76:** 89–98.
26 Yerlikaya S, Broger T, Isaacs C, et al. Blazing the trail for innovative tuberculosis diagnostics. *Infection* 2024; **52:** 29–42.
27 Pai M, Dewan PK, Swaminathan S. Transforming tuberculosis diagnosis. *Nat Microbiol* 2023; **8:** 756–59.
28 Krumholz RA. Percutaneous needle biopsy of the lung. *JAMA* 1966; **195:** 38.
29 Corral JE, Mareth KF, Riegert-Johnson DL, Das A, Wallace MB. Diagnostic yield from screening asymptomatic individuals at high risk for pancreatic cancer: a meta-analysis of cohort studies. *Clin Gastroenterol Hepatol* 2019; **17:** 41–53.
30 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:** 485–95.
31 Meggi B, Bollinger T, Mabunda N, et al. Point-of-care p24 infant testing for HIV may increase patient identification despite low sensitivity. *PLoS One* 2017; **12:** e0169497.
32 Lawn SD, Kerkhoff AD, Burton R, Meintjes G. Underestimation of the incremental diagnostic yield of HIV-associated tuberculosis in studies of the determine TB-LAM Ag urine assay. *AIDS* 2014; **28:** 1846–48.
33 Lawn SD, Kerkhoff AD, Vogt M, Wood R. High diagnostic yield of tuberculosis from screening urine samples from HIV-infected patients with advanced immunodeficiency using the Xpert MTB/RIF assay. *J Acquir Immune Defic Syndr* 2012; **60:** 289–94.
34 Huerga H, Ferlazzo G, Bevilacqua P, et al. Incremental yield of including determine-TB LAM assay in diagnostic algorithms for hospitalized and ambulatory HIV-positive patients in Kenya. *PLoS One* 2017; **12:** e0170976.
35 Boyles TH, Griesel R, Stewart A, Mendelson M, Maartens G. Incremental yield and cost of urine determine TB-LAM and sputum induction in seriously ill adults with HIV. *Int J Infect Dis* 2018; **75:** 67–73.
36 Broger T, Sossen B, du Toit E, et al. Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. *Lancet Infect Dis* 2019; **19:** 852–61.
37 Jannati A, Sadeghi V, Imani A, Saadati M. Effective coverage as a new approach to health system performance assessment: a scoping review. *BMC Health Serv Res* 2018; **18:** 886.
38 Murray CJ, Evans DB. Health systems performance assessment—debates, methods, and empiricism. Geneva: World Health Organization, 2003.
39 Shengelia B, Tandon A, Adams OB, Murray CJL. Access, utilization, quality, and effective coverage: an integrated conceptual framework and measurement strategy. *Soc Sci Med* 2005; **61:** 97–109.
40 Huerga H, Bastard M, Lubega AV, et al. Novel FujiLAM assay to detect tuberculosis in HIV-positive ambulatory patients in four African countries: a diagnostic accuracy study. *Lancet Glob Health* 2023; **11:** e126–35.
41 Denkinger CM, Kik SV, Cirillo DM, et al. Defining the needs for next generation assays for tuberculosis. *J Infect Dis* 2015; **211** (suppl 2)**:** S29–38.
42 WHO. Target product profile TB diagnostic tests for peripheral settings. https://www.who.int/news-room/articles-detail/public-consultation-for-the-target-product-profile-tb-diagnostic-tests-for-peripheral-settings. (accessed May 14, 2024).
43 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.
44 Ng M, Fullman N, Dieleman JL, Flaxman AD, Murray CJL, Lim SS. Effective coverage: a metric for monitoring universal health coverage. *PLoS Med* 2014; **11:** e1001730.
45 Drain PK, Gardiner J, Hannah H, et al. Guidance for studies evaluating the accuracy of biomarker-based nonsputum tests to diagnose tuberculosis. *J Infect Dis* 2019; **220** (suppl 3)**:** S108–15.
46 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.

47 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.

48 Denkinger CM, Schumacher SG, Gilpin C, et al. Guidance for the evaluation of tuberculosis diagnostics that meet the World Health Organization (WHO) target product profiles: an introduction to WHO process and study design principles. *J Infect Dis* 2019; **220** (suppl 3)**:** S91–98.

49 Dorman SE, Schumacher SG, Alland D, et al. Xpert MTB/RIF Ultra for detection of *Mycobacterium tuberculosis* and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *Lancet Infect Dis* 2018; **18:** 76–84.

50 Bahr NC, Nuwagira E, Evans EE, et al. Diagnostic accuracy of Xpert MTB/RIF Ultra for tuberculous meningitis in HIV-infected adults: a prospective cohort study. *Lancet Infect Dis* 2018; **18:** 68–75.

51 Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998; **17:** 857–72.

52 Giguère K, Eaton JW, Marsh K, et al. Trends in knowledge of HIV status and efficiency of HIV testing services in sub-Saharan Africa, 2000–20: a modelling study using survey and HIV testing programme data. *Lancet HIV* 2021; **8:** e284–93.

53 Hatzold K, Gudukeya S, Mutseta MN, et al. HIV self-testing: breaking the barriers to uptake of testing among men and adolescents in sub-Saharan Africa, experiences from STAR demonstration projects in Malawi, Zambia, and Zimbabwe. *J Int AIDS Soc* 2019; **22** (suppl 1)**:** e25244.

54 Zachary D, Mwenge L, Muyoyeta M, et al. Field comparison of OraQuick ADVANCE Rapid HIV-1/2 antibody test and two blood-based rapid HIV antibody tests in Zambia. *BMC Infect Dis* 2012; **12:** 183.

55 Witzel TC, Eshun-Wilson I, Jamil MS, et al. Comparing the effects of HIV self-testing to standard HIV testing for key populations: a systematic review and meta-analysis. *BMC Med* 2020; **18:** 381.

56 Pant Pai N, Balram B, Shivkumar S, et al. Head-to-head comparison of accuracy of a rapid point-of-care HIV test with oral versus whole-blood specimens: a systematic review and meta-analysis. *Lancet Infect Dis* 2012; **12:** 373–80.

57 Aledort JE, Ronald A, Rafael ME, et al. Reducing the burden of sexually transmitted infections in resource-limited settings: the role of improved diagnostics. *Nature* 2006; **444** (suppl 1)**:** 59–72.

58 Cunningham J, Jones S, Gatton ML, et al. A review of the WHO malaria rapid diagnostic test product testing programme (2008–2018): performance, procurement, and policy. *Malar J* 2019; **18:** 387.

59 Yimam Y, Mohebali M, Abbaszadeh Afshar MJ. Comparison of diagnostic performance between conventional and ultrasensitive rapid diagnostic tests for diagnosis of malaria: a systematic review and meta-analysis. *PLoS One* 2022; **17:** e0263770.

60 Mina MJ, Parker R, Larremore DB. Rethinking COVID-19 test sensitivity—a strategy for containment. *N Engl J Med* 2020; **383:** e120.

61 Pavelka M, Van-Zandvoort K, Abbott S, et al. The impact of population-wide rapid antigen testing on SARS-CoV-2 prevalence in Slovakia. *Science* 2021; **372:** 635–41.

62 Young BC, Eyre DW, Kendrick S, et al. Daily testing for contacts of individuals with SARS-CoV-2 infection and attendance and SARS-CoV-2 transmission in English secondary schools and colleges: an open-label, cluster-randomised trial. *Lancet* 2021; **398:** 1217–29.

63 Polechová J, Johnson KD, Payne P, et al. SARS-CoV-2 rapid antigen tests provide benefits for epidemic control—observations from Austrian schools. *J Clin Epidemiol* 2022; **145:** 14–19.

64 Lindner AK, Nikolai O, Kausch F, et al. Head-to-head comparison of SARS-CoV-2 antigen-detecting rapid test with self-collected nasal swab versus professional-collected nasopharyngeal swab. *Eur Respir J* 2021; **57:** 2003961.

65 Lindner AK, Nikolai O, Rohardt C, et al. Diagnostic accuracy and feasibility of patient self-testing with a SARS-CoV-2 antigen-detecting rapid test. *J Clin Virol* 2021; **141:** 104874.

66 Brümmer LE, Katzenschlager S, McGrath S, et al. Accuracy of rapid point-of-care antigen-based diagnostics for SARS-CoV-2: an updated systematic review and meta-analysis with meta-regression analyzing influencing factors. *PLoS Med* 2022; **19:** e1004011.

67 WHO. The selection and use of essential in vitro diagnostics—TRS 1031. Jan 29, 2021. https://www.who.int/publications/i/item/9789240019102 (accessed May 14, 2024).

68 Thepsamarn P, Prayoollawongsa N, Puksupa P, et al. The ICT malaria Pf: a simple, rapid dipstick test for the diagnosis of *Plasmodium falciparum* malaria at the Thai–Myanmar border. *Southeast Asian J Trop Med Public Health* 1997; **28:** 723–26.

69 WHO. New perspectives: malaria diagnosis. Dec 15, 1998. https://fctc.who.int/publications/i/item/new-perspectives-malaria-diagnosis (accessed May 14, 2024).

70 WHO. World malaria report 2021. 2021. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021 (accessed May 14, 2024).

71 WHO. Use of SARS-CoV-2 antigen-detection rapid diagnostic tests for COVID-19 self-testing. Interim guidance. March 9, 2022. https://www.who.int/publications/i/item/WHO-2019-nCoV-Ag-RDTs-Self_testing-2022.1 (accessed May 14, 2024).

72 Zhang X, Barr B, Green M, et al. Impact of community asymptomatic rapid antigen testing on COVID-19 related hospital admissions: synthetic control study. *BMJ* 2022; **379:** e071374.

73 MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. A systematic review of biomarkers to detect active tuberculosis. *Nat Microbiol* 2019; **4:** 748–58.

74 Broger T, Nicol MP, Sigal GB, et al. Diagnostic accuracy of 3 urine lipoarabinomannan tuberculosis assays in HIV-negative outpatients. *J Clin Invest* 2020; **130:** 5756–64.

75 Broger T, Basu Roy R, Filomena A, et al. Diagnostic performance of tuberculosis-specific IgG antibody profiles in patients with presumptive tuberculosis from two continents. *Clin Infect Dis* 2017; **64:** 947–55.

76 R2D2 TB Network. Product developers. https://www.r2d2tbnetwork.org/ (accessed May 14, 2024).

77 Treatment Action Group. Tuberculosis Diagnostics Pipeline Report 2023. 2023. https://www.treatmentactiongroup.org/wp-content/uploads/2023/11/2023_pipeline_TB_diagnostics_final.pdf (accessed May 14, 2024).

78 Portevin D, Moukambi F, Clowes P, et al. Assessment of the novel T-cell activation marker-tuberculosis assay for diagnosis of active tuberculosis in children: a prospective proof-of-concept study. *Lancet Infect Dis* 2014; **14:** 931–38.

79 WHO. Consolidated guidelines on tuberculosis: module 3: diagnosis: tests for TB infection. 2022. https://www.who.int/publications/i/item/9789240056084 (accessed May 14, 2024).