WHO consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection

Web Annex 2. GRADE profiles

WHO consolidated guidelines on tuberculosis. Module 3: diagnosis - rapid diagnostics for tuberculosis detection. Web Annex 2. GRADE profiles

ISBN 978-92-4-001024-6 (electronic version)

© World Health Organization 2020

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (http://www.wipo.int/amc/en/mediation/rules/).

Suggested citation. Web Annex 2. GRADE profiles. In: WHO consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see http://www.who.int/about/licensing.

Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

This publication forms part of *WHO consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection*. It is being made publicly available for transparency purposes and information, in accordance with the *WHO handbook for guideline development*, 2nd edition (2014)

Contents

Abbreviations and acronyms

AlereLAM	**Alere Determine™ TB LAM Ag**
CI	confidence interval
CRS	composite reference standard
CSF	cerebrospinal fluid
FIND	Foundation for Innovative New Diagnostics
FL-LPA	first-line line probe assay
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HIV	human immunodeficiency virus
LAM	lipoarabinomannan
LAMP	loop-mediated isothermal amplification
LF-LAM	lateral flow urine lipoarabinomannan assay
LPA	line probe assay
MDR-TB	multidrug-resistant tuberculosis
MRS	microbiological reference standard
QUADAS	quality assessment of diagnostic accuracy studies
SL-LPA	second-line line probe assay
SLID	second-line injectable drug
TB	tuberculosis
WHO	World Health Organization
XDR-TB	extensively drug-resistant tuberculosis

2.1 Grading of Recommendations Assessment, Development and Evaluation (GRADE) profiles: molecular assays

Table 1.: Xpert MTB/RIF compared to smear microscopy in adults with signs and symptoms of pulmonary tuberculosis

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Xpert MTB/RIF	smear microscopy	Relative (95% CI)	Absolute (95% CI)		
Mortality												
5 [1,2,3,4,5]	randomised trials	not serious [a]	not serious [b]	not serious	serious [c]	none	248/5265 (4.7%)	292/5144 (5.7%)	**RR 0.88** (0.73 to 1.05)	**7 fewer per 1,000** (from 15 fewer to 3 more)	⨁⨁⨁◯ MODERATE	CRITICAL
Cure												
2 [3,6,7]	randomised trials	not serious	not serious	not serious [d]	not serious	none	1786/2500 (71.4%)	1443/2080 (69.4%)	**OR 1.09** (1.02 to 1.16)	**18 more per 1,000** (from 4 more to 31 more)	⨁⨁⨁⨁ HIGH	CRITICAL
Pre-treatment loss to follow up												
3 [3,4,5]	randomised trials	not serious	serious [3,4,5,e]	not serious	not serious	none	81/642 (12.6%)	95/523 (18.2%)	**RR 0.59** (0.42 to 0.84)	**74 fewer per 1,000** (from 105 fewer to 29 fewer)	⨁⨁⨁◯ MODERATE	IMPORTANT
Time to diagnosis												
2 [2,5]	randomised trials	not serious [a]	not serious	not serious [f]	not serious [g]	none	956 participants	968 participants	**HR 1.05** (0.93 to 1.19) [Time to diagnosis]	**5 more per 1,000** (from 7 fewer to 18 more)	⨁⨁⨁⨁ HIGH	CRITICAL
							-	10.0%		**5 more per 1,000** (from 7 fewer to 18 more)		

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Xpert MTB/RIF	smear microscopy	Relative (95% CI)	Absolute (95% CI)		
Time to treatment												
4 [2,3,4,5]	randomised trials	not serious [a]	not serious	not serious [f]	serious [h]	none	4055 participants	4153 participants	**HR 1.00** (0.75 to 1.32) [Time to treatment]	**0 fewer per 1,000** (from 24 fewer to 30 more)	⨁⨁⨁◯ MODERATE	CRITICAL
							-	10.0%		**0 fewer per 1,000** (from 24 fewer to 30 more)		
Mortality in HIV-positive participants												
2	randomised trials	not serious	not serious	not serious	serious [i]	none	66/1211 (5.5%)	75/1055 (7.1%)	**RR 0.76** (0.59 to 1.00)	**17 fewer per 1,000** (from 29 fewer to 0 fewer)	⨁⨁⨁◯ MODERATE	CRITICAL
New outcome												
									not estimable		-	

CI: Confidence interval; **RR:** Risk ratio; **OR:** Odds ratio; **HR:** Hazard Ratio

Explanations

a. For all randomized trials, blinding of physicians to what test was done was impossible since knowing which test was done is part of the intervention itself. For example, the Xpert test has higher sensitivity than smear microscopy (and also produces RIF resistance results) and physicians must be allowed to take this into account when deciding about patient management. While outcomes between patients may therefore be different due to lack of blinding this was not judged to be a source of bias but rather the mechanism through which the intervention had an effect. Outcome measurement could theoretically have been influenced by the lack of blinding but this was deemed unlikely to cause bias of important magnitude. Overall, the lack of blinding was therefore judged not to put studies at increased risk of bias.Type a message

b. No evidence of inconsistency, four studies in the direction of showing benefit.

c. The 95% CI is wide likely suggesting imprecision. We caution about interpreting non-significance as no effect when the CI likely includes an effect that may be clinically important. We downgraded one level for Imprecision.

d. Cure is the outcome of interest for patient important outcome. Studies have reported treatment success which includes those cured and those completing treatment without evidence for treatment failure . However, we did not downgrade for indirectness

e. Variability in time for assessment of pre-treatment loss to follow up; Churchyard 2015 assessed within 28 days after enrolment, Cox 2014 assessed by three months after enrolment and Theron 2014 assessed by the end of the study (six months)

f. The results are from trials that directly compared the populations, interventions and outcomes of interest. We did not downgrade for imprecision

g. The results suggest that Xpert did not improve time to diagnosis compared to smear microscopy but the direction of effect is towards benefit. We did not downgrade for imprecision because the 95% CI is narrow.

h. The results suggest that Xpert did not improve the time to treatment comapred to smear microscopy. The 95% CI is wide likely suggesting imprecision

i. Similarly, the 95% CI is wide likely suggesting imprecision. We caution about interpreting non-significance as no effect when the CI likely includes an effect that may be clinically important. We downgraded one level for Imprecision.

References

1. Ngwira LG, Corbett EL,Khundi M,Barnes GL,Nkhoma A,Murowa M,et al.. Screening for tuberculosis with Xpert MTB/RIF assay versus fluorescent microscopy among adults newly diagnosed with Human Immunodeficiency Virus in rural Malawi: a cluster randomized trial (Chepetsa).. Clinical Infectious Diseases; 2019.

2. Mupfumi L, Makamure B,Chirehwa M,Sagonda T,Zinyowera S,Mason P,Metcalfe JZ,Mutetwa R. Impact of Xpert MTB/RIF on Antiretroviral Therapy-Associated Tuberculosis and Mortality: A Pragmatic Randomized Controlled Trial. Open Forum Infect Dis; 2014.

3. Cox HS, Mbhele S,Mohess N,Whitelaw A,Muller O,Zemanay W,Little F,Azevedo V,Simpson J,Boehme CC,Nicol MP.. Impact of Xpert MTB/RIF for TB diagnosis in a primary care clinic with high TB and HIV prevalence in South Africa: a pragmatic randomised trial. PLoS Med; 2014.

4. Churchyard GJ, Stevens WS,Mametja LD,McCarthy KM,Chihota V,Nicol MP,Erasmus LK,Ndjeka NO,Mvusi L,Vassall A,Sinanovic E,Cox HS,Dye C,Grant AD,Fielding KL.. Xpert MTB/RIF versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert MTB/RIF. Lancet Glob Health.; 2015.

5. Theron G, Zijenah L,Chanda D,Clowes P,Rachow A,Lesosky M,Bara W,Mungofa S,Pai M,Hoelscher M,Dowdy D,Pym A,Mwaba P,Mason P,Peter J,Dheda K, team., TB-NEAT. 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.

6. Durovni B, Saraceni V,van den Hof S,Trajman A,Cordeiro-Santos M,Cavalcante S,Menezes A,Cobelens F. Impact of replacing smear microscopy with Xpert MTB/RIF for diagnosing tuberculosis in Brazil: a stepped-wedge cluster randomized trial. PLoS Med; 2014.

7. Trajman A, Durovni B,Saraceni V,Menezes A,Cordeiro-Santos M,Cobelens F,Van den Hof S. Impact on Patients' Treatment Outcomes of XpertMTB/RIF Implementation for the Diagnosis of Tuberculosis: Follow-Up of a Stepped-Wedge Randomized Clinical Trial. PLoS One; 2015.

Table 2: Should Xpert MTB/RIF be used to diagnose pulmonary TB in adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.85 (95% CI: 0.82 to 0.88)
Specificity	0.98 (95% CI: 0.97 to 0.98)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary TB)	70 studies 10.409 patients	cross-sectional (cohort type accuracy study)	not serious [1]	not serious [a]	not serious [b]	not serious [c]	none	21 (21 to 22)	85 (82 to 88)	255 (246 to 264)	⨁⨁⨁⨁ HIGH
False negatives (patients incorrectly classified as not having pulmonary TB)								4 (3 to 4)	15 (12 to 18)	45 (36 to 54)	
True negatives (patients without pulmonary TB)	70 studies 26.828 patients	cross-sectional (cohort type accuracy study)	not serious [1]	not serious [a]	not serious	not serious	none	956 (946 to 956)	882 (873 to 882)	686 (679 to 686)	⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having pulmonary TB)								19 (19 to 29)	18 (18 to 27)	14 (14 to 21)	

Explanations

a. The median tuberculosis prevalence in the studies was 27%.

b. For individual studies, sensitivity estimates ranged from 43% to 100%. We thought that differences in enrolment criteria (different populations targeted), disease severity, and setting could in part explain heterogeneity. We did not downgrade for inconsistency.

c. There were a large number of studies and participants in this analysis. The 95% CrI around true positives and false negatives would probably not lead to different decisions depending on which credible limits are assumed. We did not downgrade for imprecision.

References

1. Horne, D. J. Kohli M. Zifodya J. S. Schiller I. Dendukuri N. Tollefson D. Schumacher,S. G. Ochodo,E. A. Pai,M. Steingart,K. R. Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database Syst Rev; 2019.

Table 3: Should Xpert Ultra be used to diagnose pulmonary tuberculosis in adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.90 (95% CI: 0.84 to 0.94)
Specificity	0.96 (95% CI: 0.93 to 0.97)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary tuberculosis)	6 studies 960 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	not serious	none	22 (21 to 23)	90 (84 to 94)	269 (253 to 281)	⨁⨁⨁⨁ HIGH
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								3 (2 to 4)	10 (6 to 16)	31 (19 to 47)	
True negatives (patients without pulmonary tuberculosis)	6 studies 1694 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	not serious	none	932 (902 to 951)	860 (833 to 878)	669 (648 to 683)	⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having pulmonary tuberculosis)								43 (24 to 73)	40 (22 to 67)	31 (17 to 52)	

Explanations

a. We considered 4/6 studies, accounting for 82.2% of the participants in this analysis, to be applicable to the review question. In Chakravorty 2017, 63% of participants had pulmonary TB; however this study accounted for only 10.4% of the total participants in this analysis. In Opota 2019, information about clinical setting and whether patients had received TB drugs for more than 7 days was not reported; however, this study accounted for only 7.4% of the total participants in this analysis. We did not downgrade for Indirectness.

Table 4: Should Xpert MTB/RIF be used to diagnose pulmonary TB in sputum in children with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.65 (95% CI: 0.55 to 0.73)
Specificity	0.99 (95% CI: 0.98 to 0.99)

Prevalences	1%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with pulmonary TB)	23 studies 493 patients	cross-sectional (cohort type accuracy study)	not serious [a]	serious [b]	not serious [c]	not serious [d]	none	6 (6 to 7)	65 (55 to 73)	129 (111 to 146)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having pulmonary TB)								4 (3 to 4)	35 (27 to 45)	71 (54 to 89)	
True negatives (patients without pulmonary TB)	23 studies 6119 patients	cross-sectional (cohort type accuracy study)	serious [e]	not serious	not serious	not serious	none	980 (971 to 985)	891 (883 to 896)	792 (785 to 796)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly classified as having pulmonary TB)								10 (5 to 19)	9 (4 to 17)	8 (4 to 15)	

Explanations

a. As assessed by QUADAS-2, 22 studies (95%) had low risk of bias.

b. Eight studies (34%) had high or unclear concern about applicability because, in these studies, patients were enrolled from inpatient tertiary care centers, which could lead to the enrollment of children with more advanced disease. Of these studies, Nhu 2013 and Singh 2016 had among the highest sensitivities. We downgraded one level for indirectness.

c. For individual studies, sensitivity estimates ranged from 27% to 100%. We thought that differences in enrolment criteria (different populations targeted), disease severity, and different ages and settings could explain the heterogeneity. We did not downgrade for inconsistency.

d. The 95% CI around true positives and false negatives would likely not lead to different decisions depending on which confidence limits are assumed. We did not downgrade for imprecision.

e. As assessed by QUADAS-2, 11 studies (47%) had unclear risk of bias based on the collection of a single culture to exclude tuberculosis. We downgraded one level for risk of bias.

Table 5: Should Xpert Ultra be used to diagnose pulmonary TB in sputum in children with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.73 (95% CI: 0.65 to 0.80)
Specificity	0.97 (95% CI: 0.96 to 0.98)

Prevalences	1%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with pulmonary TB)	3 studies 136 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	serious [b]	none	7 (6 to 8)	73 (65 to 80)	146 (129 to 159)	⊕⊕◯◯ LOW
False negatives (patients incorrectly classified as not having pulmonary TB)								3 (2 to 4)	27 (20 to 35)	54 (41 to 71)	
True negatives (patients without pulmonary TB)	3 studies 551 patients	cross-sectional (cohort type accuracy study)	not serious	not serious	not serious	not serious	none	960 (950 to 970)	873 (864 to 882)	776 (768 to 784)	⊕⊕⊕⊕ HIGH
False positives (patients incorrectly classified as having pulmonary TB)								30 (20 to 40)	27 (18 to 36)	24 (16 to 32)	

Explanations

a. Two studies (66%) had high concern about applicability because, in these studies, patients were enrolled from inpatient tertiary care centers, which could lead to the enrollment of children with more advanced disease. We downgraded one level.

b. There was a small number of children with pulmonary TB contributing to this analysis for the observed sensitivity. We downgraded one level for imprecision.

Table 6: Should Xpert MTB/RIF be used to diagnose TB meningitis in CSF in adults with signs and symptoms of TB meningitis, against a microbiological reference standard?

Sensitivity	0.70 (95% CI: 0.61 to 0.79)
Specificity	0.97 (95% CI: 0.95 to 0.98)

Prevalences	2.5%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with TB meningitis)	28 studies 521 patients	cross-sectional (cohort type accuracy study)	not serious [a]	not serious	serious [b]	not serious	none	18 (15 to 20)	70 (61 to 79)	141 (122 to 158)	⨁⨁⨁◯ MODERATE
False negatives (patients incorrectly classified as not having TB meningitis)								7 (5 to 10)	30 (21 to 39)	59 (42 to 78)	
True negatives (patients without TB meningitis)	28 studies 2582 patients	cross-sectional (cohort type accuracy study)	not serious	not serious	not serious	not serious	none	944 (928 to 956)	871 (857 to 883)	774 (762 to 785)	⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having TB meningitis)								31 (19 to 47)	29 (17 to 43)	26 (15 to 38)	

Explanations

a. We judged 79% of the studies at low risk of bias. We did not downgrade for risk of bias.

b. The sensitivity ranged from 33% to 100%. We thought that differences in CSF volume and processing could explain in part the heterogeneity, but not all. We downgraded one level for inconsistency.

Table 7: Should Xpert Ultra be used to diagnose TB meningitis in CSF in adults with signs and symptoms of TB meningitis, against a microbiological reference standard?

Sensitivity	0.87 (95% CI: 0.69 to 0.96)
Specificity	0.88 (95% CI: 0.69 to 0.95)

Prevalences	2.5%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with TB meningitis)	4 studies 40 patients	cross-sectional (cohort type accuracy study)	not serious	not serious	not serious	very serious [a]	none	22 (17 to 24)	87 (69 to 96)	174 (139 to 191)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having TB meningitis)								3 (1 to 8)	13 (4 to 31)	26 (9 to 61)	
True negatives (patients without TB meningitis)	4 studies 143 patients	cross-sectional (cohort type accuracy study)	not serious	not serious	not serious [b]	very serious [c]	none	855 (673 to 931)	789 (621 to 859)	702 (552 to 764)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having TB meningitis)								120 (44 to 302)	111 (41 to 279)	98 (36 to 248)	

Explanations

a. There were few participants in this analysis. The very wide 95% CrI around true positives and false negatives may lead to different decisions depending on which credible limits are assumed. We downgraded two levels for imprecision.

b. For individual studies, specificity estimates ranged from 43% (Chin 2019) to 100% (Perez-Risco 2018). Chin 2019 explained that they inoculated uncentrifuged CSF which could have led to low culture positivity, thus resulting in higher number of false positives. Perez-Risco 2018 contributed only 1 participant to this analysis. We did not downgrade for inconsistency.

c. The very wide 95% CrI around true negatives and false positives would likely lead to different decisions depending on which credible limits are assumed. We downgraded two levels for imprecision.

Table 8: Should Xpert MTB/RIF be used to diagnose lymph node TB in lymph node aspirates in adults with signs and symptoms of lymph node TB, against a composite reference standard?

Sensitivity	0.81 (95% CI: 0.62 to 0.92)
Specificity	0.96 (95% CI: 0.90 to 0.98)

Prevalences	2.5%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with lymph node TB)	4 studies 377 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	serious [b]	not serious [c]	none	20 (16 to 23)	81 (62 to 92)	162 (124 to 184)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having lymph node TB)								5 (2 to 9)	19 (8 to 38)	38 (16 to 76)	
True negatives (patients without lymph node TB)	4 studies 302 patients	cross-sectional (cohort type accuracy study)	serious [d]	not serious	not serious	serious [e]	none	935 (878 to 958)	863 (811 to 885)	767 (721 to 786)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having lymph node TB)								40 (17 to 97)	37 (15 to 89)	33 (14 to 79)	

Explanations

a. For indirectness, regarding applicability, for the patient selection domain, we considered most studies to have unclear concern. We were interested in how Xpert MTB/RIF performed in patients presumed to have extrapulmonary TB who were evaluated as they would be in routine practice. However, none of the studies reported this information. We downgraded one level for indirectness.

b. For individual studies, sensitivity estimates ranged from 49% to 97%. We could not explain the heterogeneity by study quality or other factors. We downgraded one level for inconsistency.

c. There were few participants contributing to this analysis for the observed sensitivity. As we had already downgraded for inconsistency, we did not downgrade further for imprecision.

d. The composite reference standard was defined by the primary study authors and therefore, was not uniform. We downgraded one level for risk of bias.

e. The very wide 95% CrI for true negatives and false positives may lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecsion.

Table 9: Should Xpert Ultra be used to diagnose lymph node TB in lymph node aspirates in adults with signs and symptoms of lymph node TB, against a microbiological reference standard?

Sensitivity	0.78 (95% CI: 0.40 to 0.97)
Specificity	0.78 (95% CI: 0.66 to 0.87)

Prevalences	2.5%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with lymph node TB)	1 studies 9 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	very serious [b]	none	20 (10 to 24)	78 (40 to 97)	156 (80 to 194)	⨁◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having lymph node TB)								5 (1 to 15)	22 (3 to 60)	44 (6 to 120)	
True negatives (patients without lymph node TB)	1 studies 64 patients	cross-sectional (cohort type accuracy study)	not serious [c]	serious [a]	not serious	very serious [d]	none	761 (644 to 848)	702 (594 to 783)	624 (528 to 696)	⨁◯◯◯ VERY LOW
False positives (patients incorrectly classified as having lymph node TB)								214 (127 to 331)	198 (117 to 306)	176 (104 to 272)	

Explanations

a. We identified only one study, which was conducted at a tertiary referral centre in South Africa, a high TB burden country. Although most participants (84%) were seen as outpatients, a high proportion had tuberculosis tests or chest radiographs prior to referral. TB prevalence in the study was 12%. Nonetheless, with only one study, applicability to other settings comes with some uncertainty. We downgraded one level for indirectness.

b. There were very few participants contributing to this analysis. The 95% CI was very wide. We downgraded two levels for imprecision.

c. In this study, the lymph node aspirates were not decontaminated before culture inoculation, which is the ideal practice for sterile specimens.

d. There were very few participants contributing to this analysis. The 95% CI was very wide. We downgraded two levels for imprecision.

Table 10: Should Xpert Ultra be used to diagnose lymph node TB in lymph node aspirates in adults with signs and symptoms of lymph node TB, against a composite reference standard?

Sensitivity	0.70 (95% CI: 0.51 to 0.85)
Specificity	1.00 (95% CI: 0.92 to 1.00)

Prevalences	2.5%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with lymph node TB)	1 studies 30 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	very serious [b]	none	17 (13 to 21)	70 (51 to 85)	140 (102 to 170)	⨁◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having lymph node TB)								8 (4 to 12)	30 (15 to 49)	60 (30 to 98)	
True negatives (patients without lymph node TB)	1 studies 43 patients	cross-sectional (cohort type accuracy study)	not serious [c]	serious [a]	not serious	serious [d]	none	975 (897 to 975)	900 (828 to 900)	800 (736 to 800)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having lymph node TB)								0 (0 to 78)	0 (0 to 72)	0 (0 to 64)	

Explanations

a. We identified only one study which was conducted at a referral centre in South Africa, a high TB burden country. Although most participants (84%) were seen as outpatients, a high proportion had tuberculosis tests or chest radiographs prior to referral. TB prevalence in the study was 41%, higher than the TB prevalences provided in the table. In some instances, prevalence may be a marker of disease spectrum, with high prevalence commonly being interpreted as indicative of more severe disease. It is possible the test will perform differently at lower prevalences. Applicability to other settings comes with some uncertainty. We downgraded one level for indirectness.

b. There were very few participants contributing to this analysis. The 95% CI was very wide. We downgraded two levels for imprecision.

c. In this study, the lymph node aspirates were not decontaminated before culture inoculation, which is the ideal practice for sterile specimens.

d. There were very few participants contributing to this analysis. In contrast to the 95% CI for sensitivity, for specificity, the interval was relatively narrow. We downgraded one level for imprecision.

Table 11: Should Xpert Ultra be used to diagnose lymph node TB in lymph node biopsies in adults with signs and symptoms of lymph node TB, against a microbiological reference standard?

Sensitivity	0.90 to 1.00
Specificity	0.38 to 0.87

Prevalences	2.5%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with lymph node TB)	2 studies 23 patients	cross-sectional (cohort type accuracy study)	serious [a]	serious [b]	not serious	very serious [c]	none	23 to 25	90 to 100	180 to 200	⨁◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having lymph node TB)								0 to 2	0 to 10	0 to 20	
True negatives (patients without lymph node TB)	2 studies 108 patients	cross-sectional (cohort type accuracy study)	serious [a]	serious [b]	serious [d]	not serious [e]	none	371 to 848	342 to 783	304 to 696	⨁◯◯◯ VERY LOW
False positives (patients incorrectly classified as having lymph node TB)								127 to 604	117 to 558	104 to 496	

Explanations

a. As assessed by QUADAS-2, we judged risk of bias as unclear because, in one study, the manner of selection not reported. We downgraded one level for risk of bias.

b. There were only two studies in this analysis. One study was conducted at a tertiary referral centre in South Africa; TB prevalence was 12%. The other study was conducted in a tertiary care hospital in China; TB prevalence was 26%. Both studies are high TB burden countries. Applicability to other settings comes with some uncertainty. We downgraded one level for indirectness.

c. There were very few participants contributing to this analysis. We downgraded two levels for imprecision.

d. The specificity estimates were variable. We could not explain the variability. We downgraded one level for inconsistency.

e. As we had already downgraded for inconsistency, we did not downgrade further for imprecision.

Table 12: Should Xpert Ultra be used to diagnose lymph node TB in lymph node biopsies in adults with signs and symptoms of lymph node TB, against a composite reference standard?

Sensitivity	0.73 (95% CI: 0.50 to 0.89)
Specificity	0.96 (95% CI: 0.88 to 1.00)

Prevalences	2.5%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 20%	
True positives (patients with lymph node TB)	1 studies 22 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	very serious [b]	none	18 (13 to 22)	73 (50 to 89)	146 (100 to 178)	⨁◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having lymph node TB)								7 (3 to 12)	27 (11 to 50)	54 (22 to 100)	
True negatives (patients without lymph node TB)	1 studies 57 patients	cross-sectional (cohort type accuracy study)	not serious [c]	serious [a]	not serious	very serious [b]	none	936 (858 to 975)	864 (792 to 900)	768 (704 to 800)	⨁◯◯◯ VERY LOW
False positives (patients incorrectly classified as having lymph node TB)								39 (0 to 117)	36 (0 to 108)	32 (0 to 96)	

Explanations

a. We identified only one study which was conducted at a referral centre in South Africa, a high TB burden country. Although most participants (84%) were seen as outpatients, a high proportion had tuberculosis tests or chest radiographs prior to referral. TB prevalence in the study was 28%, higher than the TB prevalences provided in the table. Applicability to other settings comes with some uncertainty. We downgraded one level for indirectness.

b. There were very few participants contributing to this analysis. The 95% CI was very wide. We downgraded two levels for imprecision.

c. In this study, the lymph node biopsy specimens were not decontaminated before culture inoculation, which is ideal practice for sterile specimens.

Table 13: Should Xpert MTB/RIF be used to diagnose TB meningitis in CSF in children with signs and symptoms of TB meningitis, against a microbiological reference standard?

Sensitivity	0.54 (95% CI: 0.28 to 0.78)
Specificity	0.94 (95% CI: 0.84 to 0.98)

Prevalences	1%	5%	10%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 5%	pre-test probability of 10%	
True positives (patients with TB meningitis)	6 studies 28 patients	cross-sectional (cohort type accuracy study)	serious [a]	not serious [b]	serious [c]	serious [d]	none	5 (3 to 8)	27 (14 to 39)	54 (28 to 78)	⊕◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having TB meningitis)								5 (2 to 7)	23 (11 to 36)	46 (22 to 72)	
True negatives (patients without TB meningitis)	6 studies 213 patients	cross-sectional (cohort type accuracy study)	serious [e]	not serious	not serious	serious [f]	none	929 (837 to 966)	891 (803 to 927)	844 (761 to 878)	⊕⊕◯◯ LOW
False positives (patients incorrectly classified as having TB meningitis)								61 (24 to 153)	59 (23 to 147)	56 (22 to 139)	

Explanations

a. As assessed by QUADAS-2, 3 studies (50%) had low risk of bias and the risk of bias was unclear for the remainder. We downgraded one level for risk of bias.
b. The setting was unclear or reflected a tertiary care inpatient setting in 3 studies (50%). However, this is reflective of where the target condition would typically be diagnosed and therefore we did not downgrade for indirectness.
c. For individual studies, sensitivity estimates ranged from 0% to 100%. We thought that differences in enrolment criteria (different populations targeted), disease severity, and setting could only in part explain heterogeneity. We downgraded one for inconsistency.
d. There was a low number of children with TB meningitis contributing to this analysis for the observed sensitivity. We thought the 95% CI around false negatives and true positives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.
e. The quality of the reference standard was unclear in 3 studies (50%). We downgraded one level for risk of bias.
f. We thought the 95% CI around false positives and true negatives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.

Table 14: Should Xpert Ultra repeated test be used to diagnose pulmonary TB in adults with signs and symptoms of pulmonary TB who have an initial Ultra trace result, against a microbiological reference standard?

Sensitivity	0.69 to 1.00
Specificity	0.47 to 1.00

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary TB)	3 studies 15 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	serious [b]	very serious [c]	none	17 to 25	69 to 100	207 to 300	⊕◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having pulmonary TB)								0 to 8	0 to 31	0 to 93	
True negatives (patients without pulmonary TB)	3 studies 25 patients	cross-sectional (cohort type accuracy study)	not serious	not serious	serious [b]	very serious [c]	none	458 to 975	423 to 900	329 to 700	⊕◯◯◯ VERY LOW
False positives (patients incorrectly classified as having pulmonary TB)								0 to 517	0 to 477	0 to 371	

Explanations

a. In Piersimoni 2019, >90% of participants were inpatients in a tertiary care setting. However, this study only contributed four participants (8%) to this analysis. Dorman 2018 was a multi-centre study. We did not downgrade for indirectness.

b. For individual studies, sensitivity estimates ranged from 69% to 100% and specificity from 66% to 100%. The very small number of participants in Mishra 2019a and Piersimoni 2019 (a total of four participants in each study for this analysis) may in part explain the inconsistency. We downgraded one level for inconsistency.

c. Only 3 studies, one of which Dorman 2018 contributed 42 participants and the other 2 studies contributed 4 participants each. We downgraded two levels for imprecision.

Table 15: Should more than one Xpert MTB/RIF vs. one Xpert MTB/RIF be used to diagnose pulmonary TB in sputum in children with signs and symptoms of pulmonary TB, against a microbiological reference standard?

more than one Xpert MTB/RIF		one Xpert MTB/RIF	
Sensitivity	0.59 (95% CI: 0.43 to 0.73)	Sensitivity	0.46 (95% CI: 0.35 to 0.58)
Specificity	0.99 (95% CI: 0.98 to 1.00)	Specificity	1.00 (95% CI: 0.99 to 1.00)

Prevalences	1%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested						Test accuracy CoE
								pre-test probability of 1%		pre-test probability of 10%		pre-test probability of 20%		
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	more than one Xpert MTB/RIF	one Xpert MTB/RIF	more than one Xpert MTB/RIF	one Xpert MTB/RIF	more than one Xpert MTB/RIF	one Xpert MTB/RIF	
True positives (patients with pulmonary TB)	5 studies 180 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	serious [b]	none	6 (4 to 7)	5 (3 to 6)	59 (43 to 73)	46 (35 to 58)	118 (86 to 146)	92 (70 to 116)	⊕⊕◯◯ LOW
								1 more TP in more than one Xpert MTB/RIF		**13 more TP in more than one Xpert MTB/RIF**		**26 more TP in more than one Xpert MTB/RIF**		
False negatives (patients incorrectly classified as not having pulmonary TB)								4 (3 to 6)	5 (4 to 7)	41 (27 to 57)	54 (42 to 65)	82 (54 to 114)	108 (84 to 130)	
								1 fewer FN in more than one Xpert MTB/RIF		**13 fewer FN in more than one Xpert MTB/RIF**		**26 fewer FN in more than one Xpert MTB/RIF**		
True negatives		cross-sectional	not serious	not serious	not serious	not serious	none	980 (970 to 990)	990 (980 to 990)	891 (882 to 900)	900 (891 to 900)	792 (784 to 800)	800 (792 to 800)	

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested						Test accuracy CoE
								pre-test probability of 1%		pre-test probability of 10%		pre-test probability of 20%		
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	more than one Xpert MTB/RIF	one Xpert MTB/RIF	more than one Xpert MTB/RIF	one Xpert MTB/RIF	more than one Xpert MTB/RIF	one Xpert MTB/RIF	
(patients without pulmonary TB)	5 studies 1939 patients	(cohort type accuracy study)						**10 fewer TN in more than one Xpert MTB/RIF**		**9 fewer TN in more than one Xpert MTB/RIF**		**8 fewer TN in more than one Xpert MTB/RIF**		⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having pulmonary TB)								10 (0 to 20)	0 (0 to 10)	9 (0 to 18)	0 (0 to 9)	8 (0 to 16)	0 (0 to 8)	
								10 more FP in more than one Xpert MTB/RIF		**9 more FP in more than one Xpert MTB/RIF**		**8 more FP in more than one Xpert MTB/RIF**		

Explanations

a. Two studies (40%) had high or unclear concern about applicability because, in these studies, patients were enrolled from inpatient tertiary care settings, which could lead to the enrollment of children with more advanced disease. We downgraded one level for indirectness.

b. There was a small number of children with pulmonary TB contributing to this analysis for the observed sensitivity. We thought the 95% CI around false negatives and true positives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.

Table 16: Should more than one Xpert Ultra vs. one Xpert Ultra be used to diagnose pulmonary TB in sputum in children with signs and symptoms of pulmonary TB, against a microbiological reference standard?

more than one Xpert Ultra		one Xpert Ultra	
Sensitivity	0.75 (95% CI: 0.55 to 0.89)	Sensitivity	0.64 (95% CI: 0.44 to 0.81)
Specificity	0.98 (95% CI: 0.93 to 0.99)	Specificity	1.00 (95% CI: 0.97 to 1.00)

Prevalences	1%	10%	20%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested						Test accuracy CoE
								pre-test probability of 1%		pre-test probability of 10%		pre-test probability of 20%		
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	more than one Xpert Ultra	one Xpert Ultra	more than one Xpert Ultra	one Xpert Ultra	more than one Xpert Ultra	one Xpert Ultra	
True positives (patients with pulmonary TB)	1 studies 28 patients	cross-sectional (cohort type accuracy study)	not serious	very serious [a]	not serious	very serious [b]	none	8 (6 to 9)	6 (4 to 8)	75 (55 to 89)	64 (44 to 81)	150 (110 to 178)	128 (88 to 162)	⨁◯◯◯ VERY LOW
								2 more TP in more than one Xpert Ultra		**11 more TP in more than one Xpert Ultra**		**22 more TP in more than one Xpert Ultra**		
False negatives (patients incorrectly classified as not having pulmonary TB)								2 (1 to 4)	4 (2 to 6)	25 (11 to 45)	36 (19 to 56)	50 (22 to 90)	72 (38 to 112)	
								2 fewer FN in more than one Xpert Ultra		**11 fewer FN in more than one Xpert Ultra**		**22 fewer FN in more than one Xpert Ultra**		
True negatives (patients without pulmonary TB)	1 studies 135 patients	cross-sectional (cohort type	not serious	very serious [a]	not serious	not serious	none	970 (921 to 980)	990 (960	882 (837 to 891)	900 (873	784 (744 to 792)	800 (776	

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested						Test accuracy CoE
								pre-test probability of 1%		pre-test probability of 10%		pre-test probability of 20%		
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	more than one Xpert Ultra	one Xpert Ultra	more than one Xpert Ultra	one Xpert Ultra	more than one Xpert Ultra	one Xpert Ultra	
		accuracy study)							to 990)		to 900)		to 800)	⊕⊕◯◯ LOW
								20 fewer TN in more than one Xpert Ultra		**18 fewer TN in more than one Xpert Ultra**		**16 fewer TN in more than one Xpert Ultra**		
False positives (patients incorrectly classified as having pulmonary TB)								20 (10 to 69)	0 (0 to 30)	18 (9 to 63)	0 (0 to 27)	16 (8 to 56)	0 (0 to 24)	
								20 more FP in more than one Xpert Ultra		**18 more FP in more than one Xpert Ultra**		**16 more FP in more than one Xpert Ultra**		

Explanations

a. Only one study contributed to this analysis. The results may not be applicable to other settings. We downgraded two levels for indirectness.

b. There was a low number of children with pulmonary TB contributing to this analysis for the observed sensitivity. We thought the 95% CI around false negatives and true positives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded two levels for imprecision.

Table 17: Should Xpert MTB/RIF be used to diagnose pulmonary tuberculosis in adults in the general population following a positive TB symptom screen or chest X-ray with lung abnormalities or both, against a microbiological reference standard?

Sensitivity	0.73 (95% CI: 0.62 to 0.82)
Specificity	0.99 (95% CI: 0.98 to 0.99)

Prevalences	1%	3%	7%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 3%	pre-test probability of 7%	
True positives (patients with pulmonary tuberculosis)	4 studies 867 patients	cross-sectional (cohort type accuracy study)	not serious [a]	serious [b]	serious [c]	not serious	none	7 (6 to 8)	22 (19 to 25)	51 (43 to 57)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								3 (2 to 4)	8 (5 to 11)	19 (13 to 27)	
True negatives (patients without pulmonary tuberculosis)	4 studies 48689 patients	cross-sectional (cohort type accuracy study)	not serious [a]	serious [b]	not serious	not serious	none	980 (970 to 980)	960 (951 to 960)	921 (911 to 921)	⨁⨁⨁◯ MODERATE
False positives (patients incorrectly classified as having pulmonary tuberculosis)								10 (10 to 20)	10 (10 to 19)	9 (9 to 19)	

Explanations

a. The included countries were Bangladesh, Kenya, Philippines, and Vietnam. Data from Namibia were excluded owing to inconsistencies in the diagnostic algorithm. We did not downgrade for risk of bias. This was a judgement based on an assessment of the quality of the laboratory performing the reference test.

b. The included countries were Bangladesh, Kenya, Philippines, and Vietnam. The average prevalence of tuberculosis in these countries was 1.7% (range 0.8% to 5.2%), within the range of the pre-test probabilities provided in the table. However, we noted that the populations in these prevalence surveys differed from the general population with respect to prior testing, e.g. symptom screen was limited to cough for 15 days or more, as well as the requirement for results of both symptom screen and chest radiography to be available. We downgraded one level for indirectness.

c. The sensitivity estimate for Bangladesh was 84%, higher than the sensitivity estimates for the other three countries (range, 68% to 69%). We thought we could only explain in part the inconsistency owing to lower HIV prevalence in Bangladesh. We downgraded one level for inconsistency.

Table 18: Should Xpert Ultra be used to diagnose pulmonary tuberculosis in adults in the general population following a positive TB symptom screen or chest X-ray with lung abnormalities or both, against a microbiological reference standard?

Sensitivity	0.68 (95% CI: 0.55 to 0.79)
Specificity	0.98 (95% CI: 0.97 to 0.99)

Prevalences	1%	3%	7%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 3%	pre-test probability of 7%	
True positives (patients with pulmonary tuberculosis)	4 studies 345 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	serious [b]	none	7 (6 to 8)	20 (17 to 24)	48 (39 to 55)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								3 (2 to 4)	10 (6 to 13)	22 (15 to 31)	
True negatives (patients without pulmonary tuberculosis)	4 studies 12025 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	not serious	none	970 (960 to 980)	951 (941 to 960)	911 (902 to 921)	⨁⨁⨁◯ MODERATE
False positives (patients incorrectly classified as having pulmonary tuberculosis)								20 (10 to 30)	19 (10 to 29)	19 (9 to 28)	

Explanations

a. The included countries were Myanmar, South Africa, South Africa (TREAT TB project), and Zambia. The average prevalence of tuberculosis in these countries was 2.8% (range 1.6% to 6.7%), within the range of the pre-test probabilities provided in the table. However, we noted that the populations in these prevalence surveys differed from the general population with respect to prior testing, e.g. symptom screen was limited to cough for 15 days or more, as well as the requirement for results of both symptom screen and chest radiography to be available. We downgraded one level for indirectness.

b. There were relatively few participants contributing to this analysis and a wide 95% CI. The 95% CI around true positives and false negatives may lead to different decisions depending on which limits are assumed. We downgraded one level for imprecision.

Table 19: Should two Xpert Ultra vs. one Xpert Ultra be used to diagnose pulmonary tuberculosis in adults in the general population, following a positive TB symptom screen or chest X-ray with lung abnormalities or both, against a microbiological reference standard?

two Xpert Ultra		one Xpert Ultra	
Sensitivity	0.75 (95% CI: 0.59 to 0.87)	Sensitivity	0.64 (95% CI: 0.48 to 0.79)
Specificity	0.97 (95% CI: 0.94 to 0.99)	Specificity	0.98 (95% CI: 0.95 to 0.99)

Prevalences	1%	3%	7%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested						Test accuracy CoE
								pre-test probability of 1%		pre-test probability of 3%		pre-test probability of 7%		
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	two Xpert Ultra	one Xpert Ultra	two Xpert Ultra	one Xpert Ultra	two Xpert Ultra	one Xpert Ultra	
True positives (patients with pulmonary tuberculosis)	3 studies 187 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	very serious [b]	none	8 (6 to 9)	6 (5 to 8)	23 (18 to 26)	19 (14 to 24)	53 (41 to 61)	45 (34 to 55)	⨁◯◯◯ VERY LOW
								2 more TP in two Xpert Ultra		**4 more TP in two Xpert Ultra**		**8 more TP in two Xpert Ultra**		
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								2 (1 to 4)	4 (2 to 5)	7 (4 to 12)	11 (6 to 16)	17 (9 to 29)	25 (15 to 36)	
								2 fewer FN in two Xpert Ultra		**4 fewer FN in two Xpert Ultra**		**8 fewer FN in two Xpert Ultra**		
True negatives (patients without pulmonary tuberculosis)	3 studies 4893 patients	cross-sectional (cohort type	not serious	serious [a]	not serious	not serious	none	960 (931 to 980)	970 (941 to 980)	941 (912 to 960)	951 (922 to 960)	902 (874 to 921)	911 (884 to 921)	⨁⨁⨁◯ MODERATE

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested						Test accuracy CoE
								pre-test probability of 1%		pre-test probability of 3%		pre-test probability of 7%		
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	two Xpert Ultra	one Xpert Ultra	two Xpert Ultra	one Xpert Ultra	two Xpert Ultra	one Xpert Ultra	
		accuracy study)						**10 fewer TN in two Xpert Ultra**		**10 fewer TN in two Xpert Ultra**		**9 fewer TN in two Xpert Ultra**		
False positives (patients incorrectly classified as having pulmonary tuberculosis)								30 (10 to 59)	20 (10 to 49)	29 (10 to 58)	19 (10 to 48)	28 (9 to 56)	19 (9 to 46)	
								10 more FP in two Xpert Ultra		**10 more FP in two Xpert Ultra**		**9 more FP in two Xpert Ultra**		

Explanations

a. Three countries, Myanmar, Zambia, and South Africa, contributed data to this analysis. Myanmar contributed most data. Data may not be applicable to other settings. We downgraded one level for indirectness.

b. There were few participants contributing data to this analysis. The 95% CIs for two Xpert Ultra and one Xpert Ultra were wide. We downgraded two levels for imprecision.

Table 20: Should Truenat MTB be used to diagnose pulmonary tuberculosis in adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.73 (95% CI: 0.68 to 0.78)
Specificity	0.98 (95% CI: 0.97 to 0.99)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary tuberculosis)	1 studies 258 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	serious [b]	none	18 (17 to 20)	73 (68 to 78)	220 (203 to 235)	⨁⨁⨁◯ MODERATE
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								7 (5 to 8)	27 (22 to 32)	80 (65 to 97)	
True negatives (patients without pulmonary tuberculosis)	1 studies 1078 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	not serious	none	955 (945 to 961)	881 (872 to 887)	685 (678 to 690)	⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having pulmonary tuberculosis)								20 (14 to 30)	19 (13 to 28)	15 (10 to 22)	

Explanations

a. This was a multi-centre study taking place in India, Peru, Ethiopia, and Papua New Guinea. The site in Papua New Guinea did not have a microscopy centre and thus did not contribute data to these analyses. India and Ethiopia are included in the WHO high-burden country lists for TB, TB/HIV, and MDR-TB and Peru in the high-burden country list for MDR-TB. Prevalence of tuberculosis ranged from 12.3% (Ethiopia) to 24.7% (Peru), within the range presented in the pre-test probability table.

b. The 95% CI around true positives and false negatives would probably not lead to different decisions depending on which limits are assumed. However, there were relatively few participants contributing to this analysis. We downgraded one level for imprecision.

Table 21: Should Truenat MTB be used to diagnose pulmonary tuberculosis in smear-positive adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.91 (95% CI: 0.86 to 0.94)
Specificity	-- (95% CI: -- to --)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary tuberculosis)	1 studies 174 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	serious [b]	none	23 (21 to 24)	91 (86 to 94)	272 (257 to 283)	⨁⨁⨁◯ MODERATE
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								2 (1 to 4)	9 (6 to 14)	28 (17 to 43)	
True negatives (patients without pulmonary tuberculosis)	0 studies patients							0 (0 to 0)	0 (0 to 0)	0 (0 to 0)	-
False positives (patients incorrectly classified as having pulmonary tuberculosis)								975 (975 to 975)	900 (900 to 900)	700 (700 to 700)	

Explanations

a. This was a multi-centre study taking place in India, Peru, Ethiopia, and Papua New Guinea. The site in Papua New Guinea did not have a microscopy centre and thus did not contribute data to this analysis. India and Ethiopia are included in the WHO high-burden country lists for TB, TB/HIV, and MDR-TB and Peru in the high-burden country list for MDR-TB.

b. The 95% around true positives and false negatives would probably not lead to different decisions depending on which limits are assumed. However, there were relatively few participants contributing to this analysis. We downgraded one level for imprecision.

Table 22: Should Truenat MTB be used to diagnose pulmonary tuberculosis in smear-negative adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.37 (95% CI: 0.27 to 0.48)
Specificity	0.98 (95% CI: 0.97 to 0.99)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary tuberculosis)	1 studies 84 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	serious [b]	serious [c]	none	9 (7 to 12)	37 (27 to 48)	111 (82 to 143)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								16 (13 to 18)	63 (52 to 73)	189 (157 to 218)	
True negatives (patients without pulmonary tuberculosis)	1 studies 1078 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	not serious	none	955 (944 to 961)	881 (871 to 887)	685 (678 to 690)	⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having pulmonary tuberculosis)								20 (14 to 31)	19 (13 to 29)	15 (10 to 22)	

Explanations

a. This was a multi-centre study taking place in India, Peru, Ethiopia, and Papua New Guinea. The site in Papua New Guinea did not have a microscopy centre and thus did not contribute data to these analyses. India and Ethiopia are included in the WHO high-burden country lists for TB, TB/HIV, and MDR-TB and Peru in the high-burden country list for MDR-TB.

b. Sensitivity estimates were variable, 21.1% (India), 47.4% (Peru), and 62.5% (Ethiopia), although the 95% CIs overlapped. We thought differences in patient spectrum (e.g. greater proportion of paucibacillary patients) might in part explain the lower sensitivity estimate in India. We downgraded one level for inconsistency.

c. There were few participants contributing to this analysis. As we had already downgraded one level for inconsistency, we downgraded one level for imprecision.

Table 23: Should Truenat MTB Plus be used to diagnose pulmonary tuberculosis in adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.80 (95% CI: 0.75 to 0.84)
Specificity	0.97 (95% CI: 0.95 to 0.97)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary tuberculosis)	1 studies 258 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	serious [b]	none	20 (19 to 21)	80 (75 to 84)	239 (224 to 253)	⨁⨁⨁◯ MODERATE
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								5 (4 to 6)	20 (16 to 25)	61 (47 to 76)	
True negatives (patients without pulmonary tuberculosis)	1 studies 1078 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	not serious	none	941 (928 to 950)	868 (857 to 877)	676 (666 to 682)	⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having pulmonary tuberculosis)								34 (25 to 47)	32 (23 to 43)	24 (18 to 34)	

Explanations

a. This was a multi-centre study taking place in India, Peru, Ethiopia, and Papua New Guinea. The site in Papua New Guinea did not have a microscopy centre and thus did not contribute data to the analyses. India and Ethiopia are included in the WHO high-burden country lists for TB, TB/HIV, and MDR-TB and Peru in the high-burden country list for MDR-TB. Prevalence of tuberculosis ranged from 12.3% (Ethiopia) to 24.7% (Peru), within the range presented in the pre-test probability table.

b. The 95% CI around true positives and false negatives would probably not lead to different decisions depending on which limits are assumed. However, there were relatively few participants contributing to this analysis. We downgraded one level for imprecision.

Table 24: Should Truenat MTB Plus be used to diagnose pulmonary tuberculosis in smear-positive adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.96 (95% CI: 0.92 to 0.98)
Specificity	-- (95% CI: -- to --)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary tuberculosis)	1 studies 174 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	serious [b]	none	24 (23 to 25)	96 (92 to 98)	288 (276 to 294)	⨁⨁⨁◯ MODERATE
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								1 (0 to 2)	4 (2 to 8)	12 (6 to 24)	
True negatives (patients without pulmonary tuberculosis)	0 studies patients	cross-sectional (cohort type accuracy study)						0 (0 to 0)	0 (0 to 0)	0 (0 to 0)	-
False positives (patients incorrectly classified as having pulmonary tuberculosis)								975 (975 to 975)	900 (900 to 900)	700 (700 to 700)	

Explanations

a. This was a multi-centre study taking place in India, Peru, Ethiopia, and Papua New Guinea. The site in Papua New Guinea did not have a microscopy centre and thus did not contribute data to this analysis. India and Ethiopia are included in the WHO high-burden country lists for TB, TB/HIV, and MDR-TB and Peru in the high-burden country list for MDR-TB.

b. The 95% CI around the pooled sensitivity estimate is narrow. However, there were relatively few participants contributing to this analysis. We downgraded one level for imprecision.

Table 25: Should Truenat MTB Plus be used to diagnose pulmonary tuberculosis in smear-negative adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?

Sensitivity	0.46 (95% CI: 0.36 to 0.57)
Specificity	0.97 (95% CI: 0.95 to 0.97)

Prevalences	2.5%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2.5%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with pulmonary tuberculosis)	1 studies 84 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	serious [b]	serious [c]	none	12 (9 to 14)	46 (36 to 57)	139 (108 to 171)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)								13 (11 to 16)	54 (43 to 64)	161 (129 to 192)	
True negatives (patients without pulmonary tuberculosis)	1 studies 1078 patients	cross-sectional (cohort type accuracy study)	not serious	not serious [a]	not serious	not serious	none	941 (928 to 950)	868 (857 to 877)	676 (666 to 682)	⨁⨁⨁⨁ HIGH
False positives (patients incorrectly classified as having pulmonary tuberculosis)								34 (25 to 47)	32 (23 to 43)	24 (18 to 34)	

Explanations

a. This was a multi-centre study taking place in India, Peru, Ethiopia, and Papua New Guinea. The site in Papua New Guinea did not have a microscopy centre and thus did not contribute data to these analyses. India and Ethiopia are included in the WHO high-burden country lists for TB, TB/HIV, and MDR-TB and Peru in the high-burden country list for MDR-TB.

b. Sensitivity estimates were variable, 30.8% (India), 57.9% (Peru), and 62.5% (Ethiopia), although the 95% CIs overlapped. We thought differences in patient spectrum (e.g. greater proportion of paucibacillary patients) might in part explain the lower sensitivity estimate in India. We downgraded one level for inconsistency.

c. There were few participants contributing to this analysis. The 95% CI around true positives and false negatives may lead to different decisions depending on which limits are assumed. As we had already downgraded one level for inconsistency, we downgraded one level for imprecision.

Table 26: Should Truenat MTB-RIF Dx be used to diagnose rifampicin resistance in adults with signs and symptoms of pulmonary TB, microscopy centres?

Sensitivity	0.84 (95% CI: 0.62 to 0.95)
Specificity	0.95 (95% CI: 0.91 to 0.98)

Prevalences	2%	10%	15%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 2%	pre-test probability of 10%	pre-test probability of 15%	
True positives (patients with rifampicin resistance)	1 studies 19 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	serious [b]	very serious [c]	none	17 (12 to 19)	84 (62 to 95)	126 (94 to 142)	⨁◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having rifampicin resistance)								3 (1 to 8)	16 (5 to 38)	24 (8 to 56)	
True negatives (patients without rifampicin resistance)	1 studies 167 patients	cross-sectional (cohort type accuracy study)	not serious	serious [a]	not serious	serious [d]	none	933 (889 to 956)	857 (816 to 878)	809 (771 to 830)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having rifampicin resistance)								47 (24 to 91)	43 (22 to 84)	41 (20 to 79)	

Explanations

a. This was a multi-centre study taking place in India, Peru, Ethiopia, and Papua New Guinea. Data are from microscopy centres. Papua New Guinea (reference center) did not contribute data to this analysis. India, Peru, and Ethiopia are included in the WHO high-burden country list for MDR-TB. India and Peru contributed most of the data to the determination of rifampicin resistance (in the table, true positives and false negatives) because Ethiopia contributed only one participant with rifampicin resistance. The distribution of rifampicin resistance mutations detected by the assay is unknown. These results may not be applicable to other settings. We downgraded one level for Indirectness.

b. Sensitivity estimates were variable: 100% for Peru (based on 8 RIF-resistant specimens),100% for Ethiopia (based on 1 RIF-resistant specimen), and 70% for India (based on 10 RIF-resistant specimens). We downgraded one level for inconsistency.

c. When reflexed to Truenat MTB-RIF Dx from a positive result on either Truenat MTB or Truenat MTB Plus, the proportion of non-determinate Truenat MTB-RIF Dx results was 8.8% and 15.9%, respectively. There were very few participants contributing to this analysis. The 95% CI around true positives and false negatives may lead to different decisions depending on which limits are assumed. We downgraded two levels for imprecision.

d. When reflexed to Truenat MTB-RIF Dx from a positive result on either Truenat MTB or Truenat MTB Plus, the proportion of non-determinate Truenat MTB-RIF Dx results was 8.8% and 15.9%, respectively. The 95% CI around true negatives and false positives may lead to different decisions depending on which limits are assumed. We downgraded one level for imprecision.

2.2 GRADE profiles: first-line line probe assay (FL-LPA)

Table 27. Accuracy of line probe assays (LPAs) by direct testing for detecting rifampicin resistance in patients with signs and symptoms of TB

Participants: Patients with signs and symptoms of TB
Prior testing: None
Role: Replacement test for culture-based drug-susceptibility testing
Settings: Intermediate- or central-level laboratories
Index (new) tests: GenoType MTBDR*plus* version 1 assay (Hain Lifesciences, Nehren, Germany); GenoType MTBDR*plus* version 2 assay (Hain Lifesciences, Nehren, Germany); Nipro NTM+MDRTB detection kit 2 (Nipro, Tokyo, Japan). The tests were performed by direct testing on smear-positive specimens.

Reference standard: Culture-based drug-susceptibility testing
Studies: Case–control or cohort studies comparing LPAs with a reference standard

Sensitivity	0.96 (95% CI: 0.95–0.97)
Specificity	0.98 (95% CI: 0.97–0.99)

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)		Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	
True positives (patients with rifampicin resistance)	48 studies (2 876 patients)	Cohort and case–control-type studies	Serious[a]	Not serious[b]	Not serious[c]	Not serious[d]	None	48 (47–49)	144 (142–146)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having rifampicin resistance)								2 (1–3)	6 (4–8)	
True negatives (patients without rifampicin resistance)	48 studies (7 684 patients)	Cohort and case–control-type studies	Serious[a]	Not serious[b]	Not serious[c]	Not serious[e]	None	933 (923–939)	835 (826–840)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly classified as having rifampicin resistance)								17 (11–27)	15 (10–24)	

[a] The QUADAS-2 tool was used to assess the risk of bias. The risk of bias was unclear for many studies, primarily with respect to the patient-selection domain (33/48 studies), because the method of patient sampling was unspecified (for example, consecutive or random). There was also uncertainty in the index-test and reference-test domains because many studies did not specify whether the operators of the index test and the reference test were blinded to the results of the other test (30/48 and 32/48, respectively). The risk of bias was low for the flow and timing domain. The evidence was downgraded by one point.

[b] There was low concern about applicability. Given the tests' high specificity and ability to provide results within a matter of days, the tests might improve patients' outcomes by enabling earlier initiation of appropriate therapy. The evidence was not downgraded.

[c] Although some heterogeneity was noted, this was predominantly driven by a few, small outlier studies.

[d] Imprecision was considered to be present when the pooled confidence intervals were wider than 10% in either direction.

[e] Imprecision was considered to be present when the pooled confidence intervals were wider than 5% in either direction.

Table 28. Accuracy of LPAs for detecting rifampicin resistance by indirect testing of *Mycobacterium tuberculosis* complex culture isolates

Participants: Patients with signs and symptoms of TB
Prior testing: None
Role: Replacement test for culture-based drug-susceptibility testing
Settings: Intermediate- or central-level laboratories
Index (new) tests: GenoType MTBDR*plus* version 1 assay (Hain Lifesciences, Nehren, Germany); GenoType MTBDR*plus* version 2 assay (Hain Lifesciences, Nehren, Germany); Nipro NTM+MDRTB detection kit 2 (Nipro, Tokyo, Japan). The tests were performed by indirect testing on culture isolates.

Reference standard: Culture-based drug-susceptibility testing
Studies: Case–control or cohort studies comparing LPAs with a culture-based drug-susceptibility reference test

Sensitivity	0.97 (95% CI: 0.95–0.98)
Specificity	0.99 (95% CI: 0.99–1.00)

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)		Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	
True positives (patients with rifampicin resistance)	43 studies (3 913 patients)	Cohort and case–control-type studies	Serious[a]	Not serious[b]	Not serious[c]	Not serious[d]	None	48 (48–49)	145 (143–147)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having rifampicin resistance)								2 (1–2)	5 (3–7)	
True negatives (patients without rifampicin resistance)	43 studies (6 783 patients)	Cohort and case–control-type studies	Serious[a]	Not serious[b]	Not serious[c]	Not serious[e]	None	943 (937–946)	844 (838–847)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly classified as having rifampicin resistance)								7 (4–13)	6 (3–12)	

[a] The QUADAS-2 tool was used to assess the risk of bias. The risk of bias was unclear for many studies, primarily with respect to the patient-selection domain (23/43 studies), because the method of patient sampling was unspecified (for example, consecutive or random). There was also uncertainty in the index-test and reference-test domains because many studies did not specify whether the operators of the index test and the reference test were blinded to the results of the other test (36/43 and 36/43, respectively). The risk of bias was low for the flow and timing domain. The evidence was downgraded by one point.

[b] There was low concern about applicability. Given the tests' high specificity and ability to provide results within a matter of days, the tests might improve patients' outcomes by enabling earlier initiation of appropriate therapy. The evidence was not downgraded.

[c] Although some heterogeneity was noted, this was predominantly driven by a few, small outlier studies.

[d] Imprecision was considered to be present when the pooled confidence intervals were wider than 10% in either direction.

[e] Imprecision was considered to be present when the pooled confidence intervals were wider than 5% in either direction.

Table 29. Accuracy of LPAs for detecting rifampicin resistance by indirect testing of *Mycobacterium tuberculosis* complex culture isolates compared with a composite reference standard

Participants: Patients with signs and symptoms of TB
Prior testing: None
Role: Replacement test for culture-based drug-susceptibility testing
Settings: Intermediate- or central-level laboratories
Index (new) tests: GenoType MTBDR*plus* version 1 assay (Hain Lifesciences, Nehren, Germany); GenoType MTBDR*plus* version 2 assay (Hain Lifesciences, Nehren, Germany); Nipro NTM+MDRTB detection kit 2 (Nipro, Tokyo, Japan).The tests were performed by indirect testing of *Mycobacterium tuberculosis* complex culture isolates.

Reference standard: Composite reference standard
Studies: Case–control or cohort studies comparing LPAs with a reference standard

Sensitivity	0.95 (95% CI: 0.93–0.97)
Specificity	0.99 (95% CI: 0.99–1.00)

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)		Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	
True positives (patients with rifampicin resistance)	23 studies (2 091 patients)	Cohort and case–control-type studies[a]	Serious[b]	Not serious[c]	Not serious[d]	Not serious[e]	None	48 (47–48)	143 (140–145)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having rifampicin resistance)								2 (2–3)	7 (5–10)	
True negatives (patients without rifampicin resistance)	23 studies (3 392 patients)	Cohort and case–control-type studies[a]	Serious[b]	Not serious[c]	Not serious[d]	Not serious[f]	None	945 (937–948)	846 (838–848)	⊕⊕⊕◯ MODERATE

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)		Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	
False positives (patients incorrectly classified as having rifampicin resistance)								5 (2–13)	4 (2–12)	

[a] The QUADAS-2 tool was used to assess the risk of bias. In total, 8/23 studies were cross-sectional; 8/23 were case–control; and 7 studies had an unclear design.

[b] The risk of bias was unclear for many studies, primarily with respect to the patient-selection domain (12/23 studies), because the method of patient sampling was unspecified (for example, consecutive or random). Additionally, 8/23 studies were assessed as having a high risk of bias due to the use of a case–control design. Also, there was uncertainty in the index-test and reference-test domains because many studies did not specify whether the operators of the index test and the reference test were blinded to the results of the other test (14/23 and 15/23, respectively). The risk of bias was low for the flow and timing domain. The evidence was downgraded by one point.

[c] Applicability was judged to be of low concern in the majority of studies because the population and the use of the index test matched the population of interest and the settings of intended use. The evidence was not downgraded.

[d] Although some heterogeneity was noted, this was predominantly driven by a few, small outlier studies. The evidence was not downgraded.

[e] Imprecision was considered to be present when the pooled confidence intervals were wider than 10% in either direction. The evidence was not downgraded.

[f] Imprecision was considered to be present when the pooled confidence intervals were wider than 5% in either direction. The evidence was not downgraded.

Table 30. Accuracy of LPAs by direct testing for detecting isoniazid resistance in patients with signs and symptoms of TB

Participants: Patients with signs and symptoms of TB
Prior testing: None
Role: Replacement test for culture-based drug-susceptibility testing
Settings: Intermediate- or central-level laboratories
Index (new) tests: GenoType MTBDR*plus* version 1 assay (Hain Lifesciences, Nehren, Germany); GenoType MTBDR*plus* version 2 assay (Hain Lifesciences, Nehren, Germany); Nipro NTM+MDRTB detection kit 2 (Nipro, Tokyo, Japan). The tests were performed by direct testing on smear-positive specimens.

Reference standard: Culture-based drug-susceptibility testing
Studies: Case–control or cohort studies comparing LPAs with a reference standard

Sensitivity	0.89 (95% CI: 0.86–0.92)
Specificity	0.98 (95% CI: 0.97–0.99)

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	Pre-test probability of 90%	
True positives (patients with isoniazid resistance)	46 studies (3 576 patients)	Cohort and case–control-type studies	Serious [a]	Not serious [b]	Not serious [c]	Not serious [d]	None	45 (43–46)	134 (129–138)	803 (772–827)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having isoniazid resistance)								5 (4–7)	16 (12–21)	97 (73–128)	
True negatives (patients without isoniazid resistance)	46 studies (6 896 patients)	Cross-sectional (cohort-type accuracy study)	Serious [a]	Not serious [b]	Not serious [c]	Not serious [e]	None	935 (926–940)	836 (829–841)	98 (97–99)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly								15 (10–24)	14 (9–21)	2 (1–3)	

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	Pre-test probability of 90%	
classified as having isoniazid resistance)											

[a] The QUADAS-2 tool was used to assess the risk of bias. The risk of bias was unclear for many studies, primarily with respect to the patient-selection domain (32/47 studies), because the method of patient sampling was unspecified (for example, consecutive or random). There was also uncertainty in the index-test and reference-test domains because many studies did not specify whether the operators of the index test and the reference test were blinded to the results of the other test (30/47 and 32/47, respectively). The risk of bias was low for the flow and timing domain.

[b] Applicability was judged to be of low concern in the majority of studies because the population and the use of the index test matched the population of interest and the settings of intended use.

[c] Although some heterogeneity was noted, this was predominantly driven by a few, small outlier studies.

[d] Imprecision was considered to be present when the pooled confidence intervals were wider than 10% in either direction and the number of resistant specimens tested was < 15.

[e] Imprecision was considered to be present when the pooled confidence intervals were wider than 5% in either direction and the number of sensitive specimens tested was < 15.

Table 31. Accuracy of LPAs for detecting isoniazid resistance by indirect testing of *Mycobacterium tuberculosis* complex culture isolates

Participants: Patients with signs and symptoms of TB
Prior testing: None
Role: Replacement test for culture-based drug-susceptibility testing
Settings: Intermediate- or central-level laboratories
Index (new) tests: GenoType MTBDR*plus* version 1 assay (Hain Lifesciences, Nehren, Germany); GenoType MTBDR*plus* version 2 assay (Hain Lifesciences, Nehren, Germany); Nipro NTM+MDRTB detection kit 2 (Nipro, Tokyo, Japan). The tests were performed by indirect testing on *Mycobacterium tuberculosis* complex culture isolates.

Reference standard: Culture-based drug-susceptibility testing
Studies: Case–control or cohort studies comparing LPAs with a reference standard

Sensitivity	0.91 (95% CI: 0.89–0.93)
Specificity	1.00 (95% CI: 0.99–1.00)

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	Pre-test probability of 90%	
True positives (patients with isoniazid resistance)	43 studies (4 559 patients)	Cohort and case–control-type studies[a]	Serious[b]	Not serious[c]	Not serious[d]	Not serious[e]	None	46 (44–47)	137 (133–140)	819 (797–837)	⨁⨁⨁◯ MODERATE
False negatives (patients incorrectly classified as not having isoniazid resistance)								4 (3–6)	13 (10–17)	81 (63–103)	
True negatives (patients without isoniazid resistance)	43 studies (5 903 patients)	Cohort and case–control-type studies[a]	Serious[b]	Not serious[c]	Not serious[d]	Not serious[f]	None	947 (943–950)	847 (844–850)	100 (99–100)	⨁⨁⨁◯ MODERATE
False positives (patients incorrectly								3 (0–7)	3 (0–6)	0 (0–1)	

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	Pre-test probability of 90%	
classified as having isoniazid resistance)											

[a] The QUADAS-2 tool was used to assess the risk of bias. In total, 21/43 datasets were cross-sectional; 8/43 were case–control; 2/43 datasets evaluated only strains from cases known to have MDR-TB without testing any controls; and 12/43 studies had an unclear design (for example, this includes studies in which the method of participant selection was unclear or there was uncertainty about whether specimens had been chosen for their resistance pattern).

[b] The risk of bias was unclear for many studies, primarily with respect to the patient-selection domain (21/43 studies), because the method of patient sampling was unspecified (for example, consecutive or random). There was also uncertainty in the index-test and reference-test domains because many studies did not specify whether the operators of the index test and the reference test were blinded to the results of the other test (33/43 and 33/43, respectively). The risk of bias was low for the flow and timing domain.

[c] Applicability was judged to be of low concern in the majority of studies because the population and the use of the index test matched the population of interest and the settings of intended use.

[d] Although some heterogeneity was noted, this was predominantly driven by a few, small outlier studies.

[e] Imprecision was considered to be present when the pooled confidence intervals were wider than 10% in either direction and the number of resistant specimens tested was < 15.

[f] Imprecision was considered to be present when the pooled confidence intervals were wider than 5% in either direction and the number of sensitive specimens tested was < 15.

Table 32. Accuracy of LPAs for detecting isoniazid resistance in patients with signs and symptoms of TB compared with a composite reference standard

Participants: Patients with signs and symptoms of TB
Prior testing: None
Role: Replacement test for culture-based drug-susceptibility testing
Settings: Intermediate- or central-level laboratories
Index (new) tests: GenoType MTBDR*plus* version 1 assay (Hain Lifesciences, Nehren, Germany); GenoType MTBDR*plus* version 2 assay (Hain Lifesciences, Nehren, Germany); Nipro NTM+MDRTB detection kit 2 (Nipro, Tokyo, Japan)

Reference standard: Composite reference standard
Studies: Case–control or cohort studies comparing LPAs with a composite reference standard

Sensitivity	0.85 (95% CI: 0.81–0.89)
Specificity	1.00 (95% CI: 1.00–1.00)

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	Pre-test probability of 90%f	
True positives (patients with isoniazid resistance)	24 studies (2 346 patients)	Cohort and case–control-type studies	Serious[a]	Not serious[b]	Not serious[c]	Not serious[d]	None	43 (40–44)	128 (121–133)	766 (727–797)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having isoniazid resistance)								7 (6–10)	22 (17–29)	134 (103–173)	
True negatives (patients without isoniazid resistance)	24 studies (2 170 patients)	Cohort and case–control-type studies	Serious[a]	Not serious[b]	Not serious	Not serious[e]	None	949 (946–950)	849 (847–850)	100 (100–100)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly								1 (0–4)	1 (0–3)	0 (0–0)	

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 15%	Pre-test probability of 90%f	
classified as having isoniazid resistance)											

[a] The QUADAS-2 tool was used to assess the risk of bias. The risk of bias was unclear for many studies, primarily with respect to the patient-selection domain (13/24 studies), because the method of patient sampling was unspecified (for example, consecutive or random). Also, 9/24 studies were assessed as having a high risk of bias. There was also uncertainty in the index-test and reference-test domains because many studies did not specify whether the operators of the index test and the reference test were blinded to the results of the other test (63/90 and 65/90, respectively). The risk of bias was low for the flow and timing domain. The evidence was downgraded by one point.

[b] Applicability was judged to be of low concern in the majority of studies because the population and the use of the index test matched the population of interest and the settings of intended use. The evidence was not downgraded.

[c] Although some heterogeneity was noted, this was predominantly driven by a few, small outlier studies. The evidence was not downgraded.

[d] Imprecision was considered to be present when the pooled confidence intervals were wider than 10% in either direction. The evidence was not downgraded.

[e] Imprecision was considered to be present when the pooled confidence intervals were wider than 5% in either direction. The evidence was not downgraded.

[f] A 90% prevalence was chosen to reflect the scenario in which molecular drug-susceptibility testing has already identified rifampicin resistance – that is, when the negative predictive value of this test is lower.

Table 33. Accuracy of LPAs for diagnosing MDR-TB on all specimen types by direct and indirect testing

Participants: Patients with signs and symptoms of TB
Prior testing: No
Role: Replacement test for culture-based drug-susceptibility testing
Settings: Intermediate- or central-level laboratories
Index (new) tests: GenoType MTBDR*plus* version 1 assay (Hain Lifesciences, Nehren, Germany); GenoType MTBDR*plus* version 2 assay (Hain Lifesciences, Nehren, Germany); Nipro NTM+MDRTB detection kit 2 (Nipro, Tokyo, Japan). The tests were performed on all types of specimens using direct and indirect testing.

Reference standard: Culture-based drug-susceptibility testing
Studies: Case–control or cohort studies comparing LPAs with a reference standard

Sensitivity	0.93 (95% CI: 0.90–0.95)
Specificity	0.99 (95% CI: 0.99–1.00)

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 1%	Pre-test probability of 5%	Pre-test probability of 10%	
True positives (patients with MDR-TB)	60 studies (4 248 patients)	Cohort and case–control-type studies[a]	Serious[b]	Not serious[c]	Not serious[d]	Not serious[e]	None	9 (9–9)	46 (45–47)	93 (90–95)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having MDR-TB)								1 (1–1)	4 (3–5)	7 (5–10)	
True negatives (patients without MDR-TB)	60 studies (8 785 patients)	Cohort and case–control-type studies[a]	Serious[b]	Not serious[c]	Not serious[d]	Not serious	None	983 (977–986)	943 (938–946)	894 (888–896)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly								7 (4–13)	7 (4–12)	6 (4–12)	

Outcome	Number of studies (number of patients)	Study design	Factors that may decrease the quality of evidence					Effect per 1 000 patients tested (number of patients)			Test accuracy quality of evidence
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 1%	Pre-test probability of 5%	Pre-test probability of 10%	
classified as having MDR-TB)											

[a] In total, 37/60 studies were cross-sectional; 8/60 studies used a case–control or cases-only design; and 15/60 studies had an unclear design.

[b] The QUADAS-2 tool was used to assess methodological quality. The risk of bias was unclear for many studies, primarily with respect to the patient-selection domain (34/60 studies), because the method of patient sampling was unspecified (for example, consecutive or random); the risk of bias was considered to be high for the 12 studies that used a case–control design. There was also uncertainty in the index-test and reference-test domains because many studies did not specify whether the operators of the index test and the reference test were blinded to the results of the other test (37/60 and 39/60, respectively). The risk of bias was low for the flow and timing domain. The evidence was downgraded by one point.

[c] Applicability was judged to be of low concern in the majority of studies because the population and the use of the index test matched the population of interest and the settings of intended use.

[d] Although some heterogeneity was noted for sensitivity, this was predominantly driven by a few, small outlier studies. The estimates for specificity were more homogeneous. The evidence was not downgraded.

[e] Imprecision was considered to be present when the pooled confidence intervals were wider than 10% in either direction. The evidence was not downgraded.

2.3 GRADE profiles: second-line line probe assay (SL-LPA)

Table 34. Accuracy of MTBDR*sl* by direct testing for detection of fluoroquinolone (FQ) resistance in patients with rifampicin-resistant or MDR-TB

Question: What is the diagnostic accuracy of MTBDR*sl* by direct testing for detection of FQ resistance in patients with rifampicin-resistant or MDR-TB?
Participants: patients with rifampicin-resistant or MDR-TB
Prior testing: Patients who received MTBDR*sl* testing will first have received smear microscopy, Xpert MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB detection and Xpert MTB/RIF, MTBDR*plus* (version 2.0) or an alternative line-probe assay to detect first-line drug resistance
Role: Replacement test for culture-based drug susceptibility testing
Settings: Intermediate or central level laboratories
Index (new) test: MTBDR*sl* (version 1.0).[5] The test was performed by direct testing on smear-positive specimens
Reference standard: Culture-based drug susceptibility testing
Studies: Mainly cross-sectional studies

Sensitivity	0.86 (95% CI: 0.75 to 0.93)
Specificity	0.99 (95% CI: 0.97 to 0.99)

Prevalences	5%	10%	15%

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 10%	Pre-test probability of 15%	
True positives (patients with FQ resistance)	9 studies 519 patients	cross-sectional (cohort type accuracy study) [1]	not serious [2]	not serious [3]	serious [4]	not serious	none	43 (37 to 47)	86 (75 to 93)	129 (112 to 140)	⊕⊕⊕◯ MODERATE
False negatives (patients incorrectly classified as not having FQ resistance)								7 (3 to 13)	14 (7 to 25)	21 (10 to 38)	
True negatives (patients without FQ resistance)	9 studies 1252 patients	cross-sectional (cohort type accuracy study) [1]	not serious [2]	not serious [3]	not serious	not serious	none	937 (921 to 944)	887 (872 to 895)	838 (824 to 845)	⊕⊕⊕⊕ HIGH
False positives (patients incorrectly classified as								13 (6 to 29)	13 (5 to 28)	12 (5 to 26)	

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 10%	Pre-test probability of 15%	
having FQ resistance)											

Footnotes

1. Eight studies used a cross-sectional study design and one study used a case-control study design.
2. The QUADAS-2 tool was used to assess the risk of bias. All studies used consecutive sampling. In seven studies, the reader of the index test was blinded to results of the reference standard and in two studies information about blinding to the reference standard was not reported. Several studies used critical concentrations for the phenotypic culture-based reference standard that differed from the concentrations recommended by WHO. This may have lowered specificity, but this was not observed. The evidence was not downgraded.
3. There was low concern for applicability. Given that the test's high specificity and ability to provide results within a matter of days, the test might improve patient outcomes by enabling earlier initiation of appropriate therapy. The evidence was not downgraded.
4. For individual studies, sensitivity estimates ranged from 33% to 100%. One small study with the lowest sensitivity only included three fluoroquinolone-resistant patients. However, the remaining heterogeneity could not be explained by study quality or other factors. The evidence was downgraded one point
5. This systematic review mainly evaluated MTBDR*sl* (version 1.0), which has recently been replaced with version 2.0. The addition of new probes targeting more known resistance-conferring mutations in the MTBDR*sl* (version 2.0) would be expected to yield a diagnostic accuracy at least the same as or higher than that of MTBDR*sl* (version 1.0). Therefore the findings in this review should be considered applicable to the test.

Table 35. Accuracy of MTBDR*sl* by direct testing for detection of second-line injectable drugs (SLID) resistance in patients with rifampicin-resistant or MDR-TB

Question: What is the diagnostic accuracy of MTBDR*sl* by direct testing for detection of SLID resistance in patients with rifampicin-resistant or MDR-TB?
Participants: patients with rifampicin-resistant or MDR-TB
Prior testing: Patients who received MTBDR*sl* testing will first have received smear microscopy, Xpert MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB detection and Xpert MTB/RIF, MTBDR*plus* (version 2.0) or an alternative line-probe assay to detect first-line drug resistance
Role: Replacement test for culture-based drug susceptibility testing
Settings: Intermediate or central level laboratories
Index (new) test: MTBDR*sl* (version 1.0).[5] The test was performed by direct testing on smear-positive specimens
Reference standard: Culture-based drug susceptibility testing
Studies: Mainly cross-sectional studies

Sensitivity	0.87 (95% CI: 0.38 to 0.99)
Specificity	0.99 (95% CI: 0.94 to 1.00)

Prevalences	5%	10%	15%

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 10%	Pre-test probability of 15%	
True positives (patients with SLID resistance)	8 studies 348 patients	cross-sectional (cohort type accuracy study)	serious [1]	not serious [2]	not serious [3]	serious [4]	none	44 (19 to 49)	87 (38 to 99)	131 (57 to 148)	⊕⊕◯◯ LOW
False negatives (patients incorrectly classified as not having SLID resistance)								6 (1 to 31)	13 (1 to 62)	19 (2 to 93)	
True negatives (patients without SLID resistance)	8 studies 1291 patients	cross-sectional (cohort type accuracy study)	serious [1]	not serious [2]	not serious	not serious	none	945 (889 to 950)	896 (842 to 900)	846 (796 to 850)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly								5 (0 to 61)	4 (0 to 58)	4 (0 to 54)	

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 10%	Pre-test probability of 15%	
classified as having SLID resistance)											

Footnotes

1. The QUADAS-2 was used to assess the risk of bias. All studies used consecutive or random sampling. In six studies, the reader of the index test was blinded to results of the reference standard in two studies information about blinding to the reference standard was not reported. Fifty percent of the studies used critical concentrations for the phenotypic culture-based reference standard that differed from the concentrations recommended by WHO. The evidence was downgraded by one point.
2. There was low concern for applicability. Given the test's high specificity and ability to provide results within a matter of days, the test might improve patient outcomes by enabling earlier initiation of appropriate therapy. The evidence was not downgraded.
3. For individual studies, sensitivity estimates ranged from 9% to 100%. The variability was explained in part by the use of different drugs, critical concentrations, and types of culture media in the reference standard and likely presence of *eis* resistance-conferring mutations in patients in Eastern European countries. The evidence was not downgraded and considered this in the context of other factors, in particular imprecision.
4. The wide confidence interval around true positives and false negatives may lead to different decisions depending on which confidence limits are assumed. The evidence was downgraded by one point.
5. This systematic review mainly evaluated MTBDR*sl* (version 1.0), which has recently been replaced with version 2.0. The addition of new probes targeting more known resistance-conferring mutations in the MTBDR*sl* (version 2.0) would be expected to yield a diagnostic accuracy at least the same as or higher than that of MTBDR*sl* (version 1.0). Therefore the findings in this review should be considered applicable to the test.

Table 35. Accuracy of MTBDR*sl* by indirect testing for detection of FQ resistance in patients with rifampicin-resistant or MDR-TB

Question: What is the diagnostic accuracy of MTBDR*sl* by indirect testing for detection of FQ resistance in patients with rifampicin-resistant or MDR-TB?
Participants: patients with rifampicin-resistant or MDR-TB
Prior testing: Patients who received MTBDR*sl* testing will first have received smear microscopy, Xpert MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB detection and Xpert MTB/RIF, MTBDR*plus* (version 2.0) or an alternative line-probe assay to detect first-line drug resistance
Role: Replacement test for culture-based drug susceptibility testing
Settings: Intermediate or central level laboratories
Index (new) test: MTBDR*sl* (version 1.0).[5] The test was performed by indirect testing on culture isolates
Reference standard: Culture-based drug susceptibility testing
Studies: Cross-sectional and case-control studies

Sensitivity	0.86 (95% CI: 0.79 to 0.90)
Specificity	0.99 (95% CI: 0.97 to 0.99)

Prevalences	5%	10%	15%

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 5%	Pre-test probability of 10%	Pre-test probability of 15%	
True positives (patients with FQ resistance)	19 studies 869 patients	cohort & case-control type studies [1]	not serious [2]	serious [3]	serious [4]	not serious	none	43 (40 to 45)	86 (79 to 90)	128 (119 to 136)	⊕◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having FQ resistance)								7 (5 to 10)	14 (10 to 21)	22 (14 to 31)	
True negatives (patients without FQ resistance)	19 studies 1354 patients	cohort & case-control type studies [1]	not serious [2]	serious [3]	not serious	not serious	none	937 (921 to 944)	887 (872 to 895)	838 (824 to 845)	⊕⊕◯◯ LOW
False positives (patients incorrectly classified as having FQ resistance)								13 (6 to 29)	13 (5 to 28)	12 (5 to 26)	

Footnotes

1. Thirteen studies used a cross-sectional study design and six studies used a case-control design. A sensitivity analysis that only included cross-sectional studies found sensitivity and specificity estimates similar to those for all studies.
2. The QUADAS-2 tool was used to assess the risk of bias. Fourteen studies used consecutive or random sampling. In 12 studies, the reader of the test was blinded to results of the reference standard. The majority of studies used critical concentrations for the phenotypic culture-based reference standard that differed from the concentrations recommended by WHO. The evidence was downgraded by one point.
3. Several studies included patients (such as known drug-susceptible patients) that did not match the review question. Indirectness was considered in the context of other factors, including the different critical concentrations used for culture-based drug susceptibility testing. The evidence was downgraded by one point.
4. For individual studies, sensitivity estimates ranged from 57% to 100%. Some of the variability in sensitivity might be explained by the use of different drugs, different critical concentrations, and different types of culture media in the reference standard. However, some of the variability remained unexplained. The evidence was downgraded by one point.
5. This systematic review mainly evaluated MTBDR*s*l (version 1.0), which has recently been replaced with version 2.0. The addition of new probes targeting more known resistance-conferring mutations in the MTBDR*s*l (version 2.0) would be expected to yield a diagnostic accuracy at least the same as or higher than that of MTBDR*s*l (version 1.0). Therefore the findings in this review should be considered applicable to the test.

Table 36. Accuracy of MTBDR*s*/ by indirect testing for detection of SLID resistance in patients with rifampicin-resistant or MDR-TB

Question: What is the diagnostic accuracy of MTBDR*s*/ by indirect testing for detection of SLID resistance in patients with rifampicin-resistant or MDR-TB?
Participants: patients with rifampicin-resistant or MDR-TB
Prior testing: Patients who received MTBDR*s*/ testing will first have received smear microscopy, Xpert MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB detection and Xpert MTB/RIF, MTBDR*plus* (version 2.0) or an alternative line-probe assay to detect first-line drug resistance
Role: Replacement test for culture-based drug susceptibility testing
Settings: Intermediate or central level laboratories
Index (new) test: MTBDR*s*/ (version 1.0).[5] The test was performed by indirect testing on culture isolates
Reference standard: Culture-based drug susceptibility testing
Studies: Cross-sectional and case-control studies

Sensitivity	0.77 (95% CI: 0.63 to 0.86)
Specificity	0.99 (95% CI: 0.97 to 1.00)

Prevalences	5%	10%	15%

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 5%	pre-test probability of 10%	pre-test probability of 15%	
True positives (patients with SLID resistance)	16 studies 575 patients	cohort & case-control type studies [1]	serious [2]	serious [3]	serious [4]	not serious	none	38 (32 to 43)	77 (63 to 86)	115 (95 to 129)	⊕◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having SLID resistance)								12 (7 to 18)	23 (14 to 37)	35 (21 to 55)	
True negatives (patients without SLID resistance)	16 studies 1346 patients	cohort & case-control type studies [1]	serious [2]	serious [3]	not serious	not serious	none	941 (924 to 947)	892 (876 to 897)	842 (827 to 847)	⊕⊕◯◯ LOW

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 5%	pre-test probability of 10%	pre-test probability of 15%	
False positives (patients incorrectly classified as having SLID resistance)								9 (3 to 26)	8 (3 to 24)	8 (3 to 23)	

Footnotes

1. Ten studies were cross-sectional design and six studies were case-control design. A sensitivity analysis that only included cross-sectional studies found sensitivity and specificity estimates similar to those for all studies.
2. The QUADAS-2 tool was used to assess the risk of bias. Eleven studies used consecutive or random sampling. In ten studies, the reader of the test was blinded to results of the reference standard. The majority of studies used critical concentrations for the phenotypic culture-based reference standard that differed from the concentrations recommended by WHO. The evidence was downgraded by one point.
3. Several studies included patients (drug-susceptible) that did not match the review question. Indirectness was considered in the context of other factors, including the different critical concentrations used for culture-based drug susceptibility testing. The evidence was downgraded by one point.
4. For individual studies, sensitivity estimates ranged from 25% to 100%. Some of the variability could be explained by the use of different drugs, critical concentrations, and types of culture media in the reference standard and by presence of the *eis* mutation in patients from Eastern Europe. *eis* gene is not targeted by version 1.0 of the test, which may lead to lower sensitivity among Eastern European strains. However, some of the variability remained unexplained. The evidence was downgraded by one point.
5. This systematic review mainly evaluated MTBDR*sl* (version 1.0), which has recently been replaced with version 2.0. The addition of new probes targeting more known resistance-conferring mutations in the MTBDR*sl* (version 2.0) would be expected to yield a diagnostic accuracy at least the same as or higher than that of MTBDR*sl* (version 1.0). Therefore the findings in this review should be considered applicable to the test.

Table 37. Accuracy of MTBDR*sl* by direct testing for the diagnosis of XDR-TB in patients with rifampicin-resistant or MDR-TB

Question: What is the diagnostic accuracy of MTBDR*sl* by direct testing for the diagnosis of XDR-TB in patients with rifampicin-resistant or MDR-TB?
Participants: patients with rifampicin-resistant or MDR-TB
Prior testing: Patients who received MTBDR*sl* testing will first have received smear microscopy, Xpert MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB detection and Xpert MTB/RIF, MTBDR*plus* (version 2.0) or an alternative line-probe assay to detect first-line drug resistance
Role: Replacement test for culture-based drug susceptibility testing
Settings: Intermediate or central level laboratories
Index (new) test: MTBDR*sl* (version 1.0).[5] The test was performed by indirect testing on culture isolates
Reference standard: Culture-based drug susceptibility testing
Studies: Cross-sectional and case-control studies

Sensitivity	0.69 (95% CI: 0.39 to 0.89)
Specificity	0.99 (95% CI: 0.95 to 0.99)

Prevalences	1%	5%	10%

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 1%	Pre-test probability of 5%	Pre-test probability of 10%	
True positives (patients with XDR-TB)	6 studies 143 patients	cross-sectional (cohort type accuracy study)	serious [1]	not serious [2]	not serious [3]	serious [4]	none	7 (4 to 9)	35 (19 to 45)	69 (39 to 89)	⊕⊕◯◯ LOW
False negatives (patients incorrectly classified as not having XDR-TB)								3 (1 to 6)	15 (5 to 31)	31 (11 to 61)	
True negatives (patients without XDR-TB)	6 studies 1277 patients	cross-sectional (cohort type accuracy study)	serious [1]	not serious [2]	not serious	not serious	none	980 (941 to 983)	941 (903 to 943)	891 (855 to 894)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly								10 (7 to 49)	9 (7 to 47)	9 (6 to 45)	

Outcome	Number of studies (Number of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Pre-test probability of 1%	Pre-test probability of 5%	Pre-test probability of 10%	
classified as having XDR-TB)											

Footnotes

1. The QUADAS-2 tool was used to assess the risk of bias. All studies used consecutive sampling. In four studies, the reader of the test was blinded to results of the reference standard and in two studies information about blinding was not reported. The majority of studies used critical concentrations for the phenotypic culture-based reference standard that differed from the concentrations recommended by WHO. The evidence was downgraded by one point.
2. There was low concern for applicability. Given the test's high specificity and ability to provide results within a matter of days, the test might improve patient outcomes by enabling earlier initiation of appropriate therapy. The evidence was not downgraded.
3. For individual studies, sensitivity estimates ranged from 14% to 92%. We thought variability could be explained in part by the use of different drugs, critical concentrations, and types of culture media in the reference standard and likely presence of *eis* mutation in patients in Eastern European countries. The evidence was not downgrade and considered this in the context of other factors, in particular imprecision.
4. The very wide 95% CI for true positives and false negatives may lead to different decisions depending on which confidence limits are assumed. The evidence was downgraded by one point.
5. This systematic review mainly evaluated MTBDR*sl* (version 1.0), which has recently been replaced with version 2.0. The addition of new probes targeting more known resistance-conferring mutations in the MTBDR*sl* (version 2.0) would be expected to yield a diagnostic accuracy at least the same as or higher than that of MTBDR*sl* (version 1.0). Therefore the findings in this review should be considered applicable to the test.

Table 38. Accuracy of MTBDR*sl* by indirect testing for the diagnosis of XDR-TB in patients with rifampicin-resistant or MDR-TB

Question: What is the diagnostic accuracy of MTBDR*sl* by indirect testing for the diagnosis of XDR-TB in patients with rifampicin-resistant or MDR-TB?
Participants: patients with rifampicin-resistant or MDR-TB
Prior testing: Patients who received MTBDR*sl* testing will first have received smear microscopy, Xpert MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB detection and Xpert MTB/RIF, MTBDR*plus* (version 2.0) or an alternative line-probe assay to detect first-line drug resistance
Role: Replacement test for culture-based drug susceptibility testing
Settings: Intermediate or central level laboratories
Index (new) test: MTBDR*sl* (version 1.0).[6] The test was performed by indirect testing on culture isolates
Reference standard: Culture-based drug susceptibility testing
Studies: Cross-sectional and case-control studies

Sensitivity	0.69 (95% CI: 0.39 to 0.89)
Specificity	0.99 (95% CI: 0.95 to 0.99)

Prevalences	1%	5%	10%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 5%	pre-test probability of 10%	
True positives (patients with XDR-TB)	8 studies 173 patients	cohort & case-control type studies [1]	serious [2]	serious [3]	serious [4]	not serious [5]	none	7 (4 to 9)	35 (19 to 45)	69 (39 to 89)	⊕◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having XDR-TB)								3 (1 to 6)	15 (5 to 31)	31 (11 to 61)	
True negatives (patients without XDR-TB)	8 studies 707 patients	cohort & case-control type studies [1]	serious [2]	serious [3]	not serious [4]	not serious	none	980 (941 to 983)	941 (903 to 943)	891 (855 to 894)	⊕⊕◯◯ LOW
False positives (patients incorrectly								10 (7 to 49)	9 (7 to 47)	9 (6 to 45)	

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients tested			Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 5%	pre-test probability of 10%	
classified as having XDR-TB)											

Footnotes

1. Four studies were cross-sectional design and four were case-control design.
2. The QUADAS-2 tool was used to assess the risk of bias. Six studies used consecutive sampling. In six studies, the reader of the test was blinded to results of the reference standard. All studies used critical concentrations for the phenotypic culture-based reference standard that differed from the concentrations recommended by WHO. The evidence was downgraded one point.
3. Several studies included patients (drug-susceptible) that did not match the review question. Indirectness was considered in the context of other factors, including the different critical concentrations used for culture-based drug susceptibility testing. The evidence was downgraded one point.
4. For individual studies, sensitivity estimates ranged from 20% to 100%. Some of the variability could be explained by the use of different drugs, critical concentrations, and types of culture media in the reference standard and by presence of the *eis* mutation in patients in Eastern Europe. eis gene is not targeted by version 1.0 of the test, which may lead to lower sensitivity in Eastern European strains. However, some of the variability remained unexplained. The evidence was downgraded one point.
5. The wide confidence interval around true positives and false negatives may lead to different decisions depending on which confidence limits are assumed. The evidence was not further downgraded as one point was deducted for inconsistency.
6. This systematic review mainly evaluated MTBDR*sl* (version 1.0), which has recently been replaced with version 2.0. The addition of new probes targeting more known resistance-conferring mutations in the MTBDR*sl* (version 2.0) would be expected to yield a diagnostic accuracy at least the same as or higher than that of MTBDR*sl* (version 1.0). Therefore the findings in this review should be considered applicable to the test.

2.4 GRADE profiles: lateral flow urine lipoarabinomannan assay (LF-LAM)[1]

Table 39. AlereLAM compared to no AlereLAM in HIV-positive adults to reduce mortality associated with advanced HIV disease

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AlereLAM	no AlereLAM	Relative (95% CI)	Absolute (95% CI)		
Mortality												
2	randomised trials	not serious [a]	not serious	serious [b]	not serious	none	496/2544 (19.5%)	589/2558 (23.0%)	**RR 0.85** (0.76 to 0.94)	**35 fewer per 1,000** (from 55 fewer to 14 fewer)	⨁⨁⨁◯ MODERATE	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

Explanations

a. In Gupta-Wright 2018, investigators, all study staff (other than the laboratory technician and statistician), hospital attending clinical teams, and patients were masked to the study group allocation. In Peter 2016, neither patients nor research nurses were masked to either allocation or test results. However, we doubt that the test results were biased in light of this. We did not downgrade.

b. The two trials were conducted in African countries and we do not have direct evidence of the applicability of the findings to other settings outside of Africa. We downgraded one level for indirectness.

[1] This chapter was initially published as *Web Annex B. GRADE profiles (evidence tables)* to *Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV: policy update (2019).* Geneva: World Health Organization; 2019 (https://apps.who.int/iris/bitstream/handle/10665/329511/WHO-CDS-TB-2019.18-eng.pdf). It is reproduced here for ease of reference.

Table 40. AlereLAM compared to no AlereLAM in HIV-positive adults to reduce mortality associated with advanced HIV disease, inpatient setting, CD4 ≤ 200

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AlereLAM	no AlereLAM	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: 56 weeks)												
2	randomised trials	not serious [a]	not serious	serious [b]	not serious	none	359/1449 (24.8%)	409/1437 (28.5%)	**RR 0.87** (0.77 to 0.99)	**37 fewer per 1,000** (from 65 fewer to 3 fewer)	⨁⨁⨁◯ MODERATE	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

Explanations

a. In Gupta-Wright 2018a, investigators, all study staff (other than the laboratory technician and statistician), hospital attending clinical teams, and patients were masked to the study group allocation. In Peter 2016, neither patients nor research nurses were masked to either allocation or test results. However, we doubt that the test results were biased in light of this. We did not downgrade for risk of bias.

b. The two trials were conducted in African countries and we do not have direct evidence of the applicability of the findings to other settings outside of Africa. In Gupta-Wright et al, the test was conducted in the laboratory, not at the point of care. In addition, in Gupta-Wright, the intervention was a combination of urine LAM and urine Xpert. In Peter et al, the intervention was urine LAM plus a 'nurse-informed' treatment decision. These additional considerations may not reflect how the test will be performed in routine practice. We downgraded one level for indirectness.

Table 41: Should AlereLAM be used to diagnose active TB in HIV-positive adults with TB symptoms, outpatient settings?

Sensitivity	0.29 (95% CI: 0.17 to 0.47)
Specificity	0.96 (95% CI: 0.91 to 0.99)

Prevalences	1%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with active TB)	4 studies 409 patients	cross-sectional (cohort type accuracy study)	very serious [a]	not serious [b]	not serious	not serious [c]	none	3 (2 to 5)	29 (17 to 47)	87 (51 to 141)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having active TB)								7 (5 to 8)	71 (53 to 83)	213 (159 to 249)	
True negatives (patients without active TB)	4 studies 787 patients	cross-sectional (cohort type accuracy study)	serious [d]	not serious [b]	not serious	serious [e]	none	950 (901 to 980)	864 (819 to 891)	672 (637 to 693)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having active TB)								40 (10 to 89)	36 (9 to 81)	28 (7 to 63)	

Explanations

a. As assessed by QUADAS-2, in the patient selection domain, we judged all studies at high risk of bias because they did not avoid inappropriate exclusions. We downgraded two levels for risk of bias.

b. The median TB prevalence in the studies was 43% and thus the results tend to be more applicable to settings with a higher TB prevalence. We did not downgrade for indirectness.

c. The 95% CrI around true positives and false negatives would likely not lead to different decisions depending on which credible limits are assumed. We did not downgrade for imprecision.

d. As assessed by QUADAS-2, in the reference standard domain, we judged three studies (75%) at high risk of bias because we thought the reference standard used was unlikely to correctly classify the target condition. We downgraded one level for risk of bias.

e. The 95% CrI around true negatives and false positives may lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.

Table 42: Should AlereLAM be used to diagnose active TB in HIV-positive adults irrespective of symptoms, outpatient settings, CD4 ≤ 100?

Sensitivity	0.40 (95% CI: 0.20 to 0.64)
Specificity	0.87 (95% CI: 0.68 to 0.94)

Prevalences	1%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with active TB)	2 studies 46 patients	cross-sectional (cohort type accuracy study)	very serious [a]	not serious	not serious	very serious [b]	none	4 (2 to 6)	40 (20 to 64)	120 (60 to 192)	⨁◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having active TB)								6 (4 to 8)	60 (36 to 80)	180 (108 to 240)	
True negatives (patients without active TB)	2 studies 171 patients	cross-sectional (cohort type accuracy study)	very serious [c]	not serious	not serious	very serious [d]	none	861 (673 to 931)	783 (612 to 846)	609 (476 to 658)	⨁◯◯◯ VERY LOW
False positives (patients incorrectly classified as having active TB)								129 (59 to 317)	117 (54 to 288)	91 (42 to 224)	

Explanations

a. As assessed by QUADAS-2, in the patient selection domain, we considered both studies at high risk of bias because they did not avoid inappropriate exclusions. We downgraded two levels for risk of bias.

b. There were few participants in this analysis. We downgraded two levels for imprecision.

c. As assessed by QUADAS-2, in the reference standard domain, we considered both studies at high risk of bias because we thought the reference standard used was unlikely to correctly classify the target condition. We downgraded two levels for risk of bias.

d. The very wide 95% CrIs around true negatives and false positives may lead to different decisions depending on which credible limits are assumed. We downgraded two levels for imprecision.

Table 43. Should AlereLAM be used to diagnose active TB in HIV-positive adults irrespective of symptoms, outpatient settings?

Sensitivity	0.31 (95% CI: 0.18 to 0.47)
Specificity	0.95 (95% CI: 0.87 to 0.99)

Prevalences	1%	5%	10%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 5%	pre-test probability of 10%	
True positives (patients with active TB)	6 studies 273 patients	cross-sectional (cohort type accuracy study)	serious [a]	not serious	not serious [b]	not serious [c]	none	3 (2 to 5)	16 (9 to 24)	31 (18 to 47)	⨁⨁⨁◯ MODERATE
False negatives (patients incorrectly classified as not having active TB)								7 (5 to 8)	34 (26 to 41)	69 (53 to 82)	
True negatives (patients without active TB)	6 studies 2555 patients	cross-sectional (cohort type accuracy study)	very serious [d]	not serious	not serious [e]	serious [f]	none	941 (861 to 980)	903 (827 to 941)	855 (783 to 891)	⨁◯◯◯ VERY LOW
False positives (patients incorrectly classified as having active TB)								49 (10 to 129)	47 (9 to 123)	45 (9 to 117)	

Explanations

a. As assessed by QUADAS-2, in the patient selection domain, we judged four studies (67%) at high risk of bias because they did not avoid inappropriate exclusions. We downgraded one level for risk of bias.
b. For individual studies, sensitivity ranged from 0% to 63%. We thought that the percentage of patients with TB symptoms or CD4 count could explain in part the heterogeneity. One study (LaCourse 2016) with sensitivity 0% differed from the other studies by including a) a population of exclusively pregnant women attending an antenatal care setting, b) a low proportion of symptomatic participants (19%), c) a low TB prevalence (1%), and d) a high median CD4 cell count (437 cells per µL). One study (Thit 2017) with sensitivity 63% differed from the other studies by being conducted in Myanmar, and is the only study included in this review that evaluated AlereLAM in a setting outside sub-Saharan Africa. We did not downgrade for inconsistency.
c. We thought the wide 95% CrIs around true positives and false negatives would likely not lead to different decisions depending on which credible limits are assumed. We did not downgrade for imprecision.
d. As assessed by QUADAS-2, in the reference standard domain, we judged five studies (83%) at high risk of bias because we thought the reference standard used was unlikely to correctly classify the target condition. We downgraded two levels for risk of bias.
e. For individual studies, specificity ranged from 67% to 99%. Five of the studies had specificity of 94% or higher. One study (Thit 2017) with specificity 67% differed from the other studies by being conducted in Myanmar, and is the only study included in this review that evaluated AlereLAM in a setting outside sub-Saharan Africa. We did not downgrade further for inconsistency.
f. The wide 95% CrIs around true negatives and false positives may lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.

Table 44: Should AlereLAM be used to diagnose active TB in HIV-positive adults no symptoms and no CD4 count available?

Sensitivity	0.21 (95% CI: 0.08 to 0.48)
Specificity	0.96 (95% CI: 0.89 to 0.99)

Prevalences	1%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with active TB)	0 studies patients							2 (1 to 5)	21 (8 to 48)	63 (24 to 144)	-
False negatives (patients incorrectly classified as not having active TB)								8 (5 to 9)	79 (52 to 92)	237 (156 to 276)	
True negatives (patients without active TB)	0 studies patients							950 (881 to 980)	864 (801 to 891)	672 (623 to 693)	-
False positives (patients incorrectly classified as having active TB)								40 (10 to 109)	36 (9 to 99)	28 (7 to 77)	

Table 45: Should AlereLAM be used to diagnose active TB in HIV-positive adults irrespective of symptoms, outpatient settings, CD4 ≤ 200?

Sensitivity	0.21 (95% CI: 0.08 to 0.48)
Specificity	0.96 (95% CI: 0.89 to 0.99)

Prevalences	1%	10%	30%

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1,000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 1%	pre-test probability of 10%	pre-test probability of 30%	
True positives (patients with active TB)	2 studies 65 patients	cross-sectional (cohort type accuracy study)	serious [a]	not serious	not serious [b]	very serious [c]	none	2 (1 to 5)	21 (8 to 48)	63 (24 to 144)	⊕◯◯◯ VERY LOW
False negatives (patients incorrectly classified as not having active TB)								8 (5 to 9)	79 (52 to 92)	237 (156 to 276)	
True negatives (patients without active TB)	2 studies 587 patients	cross-sectional (cohort type accuracy study)	serious [d]	not serious	not serious	serious [e]	none	950 (881 to 980)	864 (801 to 891)	672 (623 to 693)	⊕⊕◯◯ LOW
False positives (patients incorrectly classified as having active TB)								40 (10 to 109)	36 (9 to 99)	28 (7 to 77)	

Explanations

a. As assessed by QUADAS-2, in the patient selection domain, we judged one study (50%) at high risk of bias because this study did not avoid inappropriate exclusions. We downgraded one level for risk of bias.

b. We thought that differences in the percentage of patients with TB symptoms in the two studies could explain some of the heterogeneity. We did not downgrade for inconsistency.

c. The wide 95% CrI around true positives and false negatives would likely not lead to different decisions depending on which credible limits are assumed. However, there were few participants in this analysis. We downgraded two levels for imprecision.

d. As assessed by QUADAS-2, in the reference standard domain, we judged one study (50%) at high risk of bias because we thought the reference standard used was unlikely to correctly classify the target condition. We downgraded one level for risk of bias.

e. The wide 95% CrIs around true negatives and false positives would likely lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.

9789240010246