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

Web Annex C. Evidence-to-decision tables

Third edition

World Health Organization

WHO consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection, third edition. Web Annex C. Evidence-to-decision tables

ISBN 978-92-4-009046-0 (electronic version)

This publication forms part of the WHO guideline entitled *WHO consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection, third edition*. 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
COI	conflict of interest
CRS	composite reference standard
CSF	cerebrospinal fluid
DNA	deoxyribonucleic acid
DST	drug-susceptibility testing
FIND	Foundation for Innovative New Diagnostics
FL-LPA	first-line line probe assay
GDG	Guideline Development Group
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
NGS	next-generation sequencing
NTP	national TB programme
PICO	population, intervention, comparator and outcomes
PLHIV	people living with HIV
QUADAS	quality assessment of diagnostic accuracy studies
RR-TB	rifampicin-resistant tuberculosis
SL-LPA	second-line line probe assay
SLID	second-line injectable drug
STARD	Standards for Reporting Diagnostic Accuracy Studies
TB	tuberculosis
WHO	World Health Organization
XDR-TB	extensively drug-resistant tuberculosis

3.1 Evidence-to-decision tables: Xpert MTB/RIF and Xpert Ultra

PICO 1: Among adults with signs and symptoms of pulmonary TB (PTB), seeking care at health care facilities should Xpert MTB/RIF / Xpert Ultra be used as an initial test for diagnosis of PTB and rifampicin resistance (RR)?

1.1 What is impact of Xpert MTB/RIF on patient-important outcomes (cure; mortality; time to diagnosis; time to start treatment)?

Assessment

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	In 2018, tuberculosis (TB) was associated with 1.2 million deaths and a further 251,000 deaths from tuberculosis disease among people living with HIV (WHO Global tuberculosis report 2019). The absolute number of TB deaths among HIV-negative people fell by 27% between 2000 and 2018, from an estimated 1.7 million in 2000 to 1.2 million in 2018, and similarly the mortality rate fell by 42% (including 3.6% between 2017 and 2018). Of the WHO regions, Africa had the highest mortality rate (18%) (WHO Global tuberculosis report 2019). There has been progress in treatment success (cure and treatment completion). Latest data show a global success rate of 85% among new TB cases in 2017 compared to 81% in the previous year (WHO Global tuberculosis report 2019). Overall loss to follow up were high in the WHO region of the Americas accounting for 25%.	

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know		Moderate together with the RR information. RR depends on setting and Pretest probability. But many settings do still have RR. In HIV positive individuals the effect is even larger but is still considered overall moderate. RR was possibly included in the RCT evidence. But it may be an added benefit that was considered a mode additional benefit that lead the panel to move from a moderate effect resulting from the

Outcomes	With smear microscopy	With Xpert MTB/RIF	Difference	Relative effect (95% CI)
Mortality	57 per 1,000	**50 per 1,000** (41 to 60)	**7 fewer per 1,000** (15 fewer to 3 more)	**RR 0.88** (0.73 to 1.05)
Cure	694 per 1,000	**712 per 1,000** (698 to 724)	**18 more per 1,000** (4 more to 31 more)	**OR 1.09** (1.02 to 1.16)
Pre-treatment loss to follow up	182 per 1,000	**107 per 1,000** (76 to 153)	**74 fewer per 1,000** (105 fewer to 29 fewer)	**RR 0.59** (0.42 to 0.84)

Time to diagnosis	100 per 1,000	**105 per 1,000** (93 to 118)	**5 more per 1,000** (7 fewer to 18 more)	**HR 1.05** (0.93 to 1.19) [Time to diagnosis]	
Time to treatment	100 per 1,000	**100 per 1,000** (76 to 130)	**0 fewer per 1,000** (24 fewer to 30 more)	**HR 1.00** (0.75 to 1.32) [Time to treatment]	
Mortality in HIV-positive participants	71 per 1,000	**54 per 1,000** (42 to 71)	**17 fewer per 1,000** (29 fewer to 0 fewer)	**RR 0.76** (0.59 to 1.00)	

Additional desirable effect: Detection of the resistance to rifampicin: Sensitivity - 0.96, Specificity - 0.98. At 10% prevalence, 96 patients out of 1000 will be correctly diagnosed with rifampicin-resistance, and for 882 rifampicin-sensitive patients, this diagnosis will be correctly excluded. Additional undesirable effect: False detection of the resistance to rifampicin: At 10% prevalence, 18 faulse resistant to rifampicin patients will be detected out of 1000, and 4 truly resistant to rifampicin patients will be missed. resistant will be correctly diagnosed with rifampicin-resistance, and for 882 rifampicin-sensitive patients, this diagnosis will be correctly excluded.

reduction in mortality, increase in cure and

time to diagnosis.

For the subgroup of PLHIV the mortality reduction was considered large. In settings with lower MDR setting the effect may be smaller.

We have a disaggreated judgment for the desirable effects.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations

○ Large
○ Moderate
○ Small
● Trivial
○ Varies
○ Don't know

Outcomes	With smear microscopy	With Xpert MTB/RIF	Difference	Relative effect (95% CI)
Mortality	57 per 1,000	**50 per 1,000** (41 to 60)	**7 fewer per 1,000** (15 fewer to 3 more)	**RR 0.88** (0.73 to 1.05)
Cure	694 per 1,000	**712 per 1,000** (698 to 724)	**18 more per 1,000** (4 more to 31 more)	**OR 1.09** (1.02 to 1.16)
Pre-treatment loss to follow up	182 per 1,000	**107 per 1,000** (76 to 153)	**74 fewer per 1,000** (105 fewer to 29 fewer)	**RR 0.59** (0.42 to 0.84)
Time to diagnosis	100 per 1,000	**105 per 1,000** (93 to 118)	**5 more per 1,000** (7 fewer to 18 more)	**HR 1.05** (0.93 to 1.19) [Time to diagnosis]

Time to treatment	100 per 1,000	**100 per 1,000** (76 to 130)	**0 fewer per 1,000** (24 fewer to 30 more)	**HR 1.00** (0.75 to 1.32) [Time to treatment]	
Mortality in HIV-positive participants	71 per 1,000	**54 per 1,000** (42 to 71)	**17 fewer per 1,000** (29 fewer to 0 fewer)	**RR 0.76** (0.59 to 1.00)	

Additional desirable effect: Detection of the resistance to rifampicin: Sensitivity -0.96, Specificity - 0.98. At 10% prevalence, 96 patients out of 1000 will be correctly diagnosed with rifampicin-resistance, and for 882 rifampicin-sensitive patients, this diagnosis will be correctly excluded. Additional undesirable effect: False detection of the resistance to rifampicin: At 10% prevalence, 18 faulse resistant to rifampicin patients will be detected out of 1000, and 4 truly resistant to rifampicin patients will be missed. resistant will be correctly diagnosed with rifampicin-resistance, and for 882 rifampicin-sensitive patients, this diagnosis will be correctly excluded.

Certainty of evidence

What is the overall certainty of the evidence of effects?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies		

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with smear microscopy	Risk difference with Xpert MTB/RIF
Mortality	10409 (5 RCTs)[1,2,3,4,5]	⊕⊕⊕◯ MODERATE[a,b,c]	**RR 0.88** (0.73 to 1.05)	Study population	
				57 per 1,000	**7 fewer per 1,000** (15 fewer to 3 more)
Cure	4580 (2 RCTs)[3,6,7]	⊕⊕⊕⊕ HIGH[d,e]	**OR 1.09** (1.02 to 1.16)	Study population	
				694 per 1,000	**18 more per 1,000** (4 more to 31 more)
Pre-treatment loss to follow up	1165 (3 RCTs)[3,4,5]	⊕⊕⊕◯ MODERATE[3,4,5,f]	**RR 0.59** (0.42 to 0.84)	Study population	
				182 per 1,000	**74 fewer per 1,000**

					(105 fewer to 29 fewer)
Time to diagnosis	1924 (2 RCTs)[2,5]	⊕⊕⊕⊕ HIGH[a,e,g]	**HR 1.05** (0.93 to 1.19) [Time to diagnosis]	Moderate	
				100 per 1,000	**5 more per 1,000** (7 fewer to 18 more)
Time to treatmen t	8208 (4 RCTs)[2,3,4,5]	⊕⊕⊕◯ MODERATE[a,g,h]	**HR 1.00** (0.75 to 1.32) [Time to treatment]	Moderate	
				100 per 1,000	**0 fewer per 1,000** (24 fewer to 30 more)
Mortality in HIV-positive participants	2266 (2 RCTs)	⊕⊕⊕◯ MODERATE[i]	**RR 0.76** (0.59 to 1.00)	Study population	
				71 per 1,000	**17 fewer per 1,000** (29 fewer to 0 fewer)

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 randomized 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-randomized 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, randomized, 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.

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

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability	Participants assign great value to the ability of Xpert to improve the diagnosis of drug resistant TB and the impact on patients if they cannot access testing for drug resistance through Xpert. The impact on case notification and the value of Xpert for finding more TB was less clear owing to widespread clinical treatment, prolonged TATs and the challenges with feasibility and utilization of Xpert. While Xpert has eased laboratory work through convenience and automation, this preference for Xpert in the laboratory can have undesired consequences for monitoring through microscopy or for reverting back to microscopy when Xpert machines are down. While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert.	

Balance of effects

Does the balance between desirable and undesirable effects favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations

○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ● Favours the intervention ○ Varies ○ Don't know	Summary of the above: Benefits vs Harms. Probably very little variation to how people value the outcomes.	

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	From USD 9.98 in Ethiopia (Tesfaye 2017) till EUR 110.75 in Germany (Diel 2016).	Some studies used the negotiated price while studies in HIC used the regular price. This was varied in sensitivity analyses in the reviewed studies. Other cost associated with the use of the test (e.g. transportation). Unit cost varies. Median cost including implementation about US$20. Varies across countries. Just in subsaharan africa up to US$40 (unit cost). In comparison to Smear. Smear unit cost is US$3 and likely more in some settings (including drug resistant test and culture). In some countries investment for equipment is required to implement. The panel assumed resistance testing needs to be done in the comparator group.

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Systematic review by A. Zwerling: Studies employed a variety of different modelling approaches, populations and settings. Variations in costing, effectiveness and epidemiological parameters were present across included studies making direct comparisons across studies challenging.	

Cost effectiveness

Does the cost-effectiveness of the intervention favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ No included studies	Four studies were identified assessing the use of Xpert MTB/RIF among PLHIV with signs and symptoms of TB (3–5,18). Studies were conducted in countries with high HIV prevalence including South Africa, Ethiopia and Malawi. All reported Xpert would likely be cost-effective in these populations but to varying degrees and conditions of implementation. No studies assessed children specifically among these studies. Four studies among hospitalized patients were identified, 2 from the USA (21,22), 1 from Germany (20) and 1 study from China (Hong Kong Special Administrative Region (SAR)) (28). All 4 studies concluded that replacement of SSM with Xpert would results in cost-savings driven largely from high hospitalization costs associated with respiratory isolation. No studies assessed children specifically among these studies. Fifteen studies assessing cost-effectiveness of Xpert among persons presenting to primary health care facilities across Sub-Saharan Africa and Brazil. While early studies found Xpert would likely be cost-effective (albeit using a range of willingness to pay thresholds across different countries, several concerns around cost-effectiveness have been raised by subsequent analyses. Inclusion of downstream costs associated with MDR-TB and HIV treatment and care has been shown to lead to increased ICERs and increased total expenditures. Costs associated with scale-up of Xpert have been estimated to result in an important increase relative to existing TB and HIV programme budgets and in many countries may not be deemed affordable despite ICERs for Xpert approaches being under willingness to pay thresholds. Studies have highlighted the importance of implementation conditions, including existing standard of care, levels of empirical treatment, TB prevalence among presumptive patients being tested, and test volume as highly influential variables on cost-effectiveness results. Results from individual studies are summarized below. While some studies employd a population based approach no studies specifically addressed children.	GDG members suggested it probably favours the intervention and cost of treatment being considered not extra cost. In several of the HIC studies, cost savings were still realized. Panel suggested that increased use will relatively lower price. Cost may also change but the panel based their judgment on the currently available evidence about cost. Variability in cost and variability in human resources in cost-effectiveness was acknowledged. Setting and availability of the instruments may affect cost-effectiveness. Majority of studies suggsted that Xpert may be cost-effective. Judgment did not explicitly consider opportunity cost.

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations

○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	As test can be performed at all levels of the health care system, it will likely increase health care equity.	

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Report on user perspectives on Xpert MTB/RIF and Ultra testing: results from qualitative research : Test is generally described as acceptable by key stakeholders. Discordant results of repeat tests and confirmatory tests can cause confusion around what should be considered gold standard, particularly when specimen quality might be poor. Understanding and contextualizing discordant results require continuous training, experience and expertise. Establishing a thorough TB history of patients is uncommon and 'previously treated' defined differently with implications for potential of false positives results through Xpert testing. Clear parameters are needed of how to define previously treated patients, how to handle their Xpert results, and accurately capture outcomes in national databases.	Improved but not everybody who needs it can access Xpert.

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Compared to smear microscopy, users generally value the automation, convenience, higher biosafety levels and lesser human involvement that Xpert offers. The fact that it is a closed system with walk away time during the incubation (15') and machine run time (90') where lab technicians can do other testing in between was mentioned as well. As such, Xpert eased the work for lab technicians, adding a level of relief from reading hundreds of slides as well as reducing the room for errors. Persistent underutilization of Xpert machines is compounded by the challenges of delays due to sample transport, module break down, stock-out of cartridges or complicated diagnostic algorithms. Diagnostic algorithms that are simple to follow in a specific facility (f.i. test all those with presumptive TB) are more feasible and enhance utilization, but this simplicity is crucially dependent on cost and supplies.	government commitment to ensure functioning infrastructure and power; supply of cartridges, functioning laboratory services; investment in expertise to handle (discordant) results; better repair services; staff with monitoring capacities; functioning sample transport; sustainable funding models and transparent donor agreements; and simple diagnostic algorithms; à those interact and reinforce each other determining utilization

		maintance and stock outs.

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Desirable effects	Trivial	Small	**Moderate**	Large		Varies	Don't know
Undesirable effects	Large	Moderate	Small	**Trivial**		Varies	Don't know
Certainty of evidence	Very low	Low	**Moderate**	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	**Favours the intervention**	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderat e savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the interventio n	Varies	No included studies
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	**Strong recommendation for the intervention**
○	○	○	○	●

Conclusions

Recommendation

In adults with signs and symptoms of pulmonary TB, the GDG **recommends using** Xpert MTB/RIF for the diagnosis of TB (as opposed to a microbiological reference standard). (strong recommendation, moderate certainty of evidence). One member of the panel was not present.

Subgroup considerations

Applies to PLHIV (based on trial results - which alleviated concerns about the FP in low pretest probability setting). Applies to MDR TB patients. Applies to patients with prior TB (Caveat), smear neg, culture positive (high pretest prob with high FN, requiring additional testing and depending on the degree of positivity) and all other subgroups evaluated.

Implementation considerations

Manage/minimize stock out - logistical management/procurement/maintenance infrastructure set up. Treatment of detected cases.

Counselling and patient support for detected cases.

Sample transportation for both the interventioon and comparator.

Reference to implementation guides and document will be added to it. Probably similar implementation considerations.

Involvement of communities and civil societies.

Pakistan - barrier to implemenation is lack of access (given that not all patients have access leads to overall lack of use).

Assay has been available for considerable time.

Research priorities

cost effectiveness studies that use

False positive RR in low bacilary load should be investigated.

1.3 What is diagnostic accuracy of Xpert Ultra for PTB and RR, as compares with MRS?

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	To improve assay sensitivity for the detection of M. tuberculosis, the Ultra assay incorporates two different multi-copy amplification targets (IS6110 and IS1081) and a larger DNA reaction chamber than Xpert MTB/RIF (50µl PCR reaction in Ultra versus 25 µl in Xpert MTB/RIF). Ultra also incorporates fully nested nucleic acid amplification, more rapid thermal cycling, and improved fluidics and enzymes. This has resulted in Ultra having a limit of detection (LOD) of 16 bacterial colony forming units (cfu) per ml (compared to 114 cfu per ml for Xpert MTB/RIF).	
Test accuracy How accurate is the test?		

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ○ Accurate ● Very accurate ○ Varies ○ Don't know	*Test accuracy* Xpert Ultra Sensitivity: 0.90 (95% CI: 0.84 to 0.94) Specificity: 0.96 (95% CI: 0.93 to 0.97)	

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	See table below	False positives: Unsure if the reference standard is close to the gold standard. That is the reference standard is imperfect.

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 2%	Effect per 1000 patients/year for pre-test probability of 10%	Effect per 1000 patients/year for pre-test probability of 30%	Importance
True positives	cross-sectional (cohort type accuracy study)	⨁⨁⨁⨁ HIGH[a]	22 (21 to 23)	90 (84 to 94)	269 (253 to 281)	
False negatives			3 (2 to 4)	10 (6 to 16)	31 (19 to 47)	
True negatives	cross-sectional (cohort type accuracy study)	⨁⨁⨁⨁ HIGH[a]	932 (902 to 951)	860 (833 to 878)	669 (648 to 683)	
False positives			43 (24 to 73)	40 (22 to 67)	31 (17 to 52)	

a. We considered 4/6 studies, accounting for 82.2% of the participants in this analysis, to be applicable to the review question. In Chakravoty 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.

Desirable / Undesirable effects - RR

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations

○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know		False positives may not be actual false positives given the imperfect reference standard. Culture may be false negative under these circumstances which would categorize patients inappropriately as false positives using Ultra Xpert.

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 2%	Effect per 1000 patients/year for pre-test probability of 10%	Effect per 1000 patients/year for pre-test probability of 30%	Importance
True positives	cross-sectional (cohort type accuracy study)	⨁⨁⨁⨁ HIGH[a]	22 (21 to 23)	90 (84 to 94)	269 (253 to 281)	
False negatives			3 (2 to 4)	10 (6 to 16)	31 (19 to 47)	
True negatives	cross-sectional (cohort type accuracy study)	⨁⨁⨁⨁ HIGH[a]	932 (902 to 951)	860 (833 to 878)	669 (648 to 683)	
False positives			43 (24 to 73)	40 (22 to 67)	31 (17 to 52)	

a. We considered 4/6 studies, accounting for 82.2% of the participants in this analysis, to be applicable to the review question. In Chakravoty 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.

Desirable / Undesirable effects - RR

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	No adverse events were associated with Xpert testing. High quality evidence.	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High	No adverse events were associated with Xpert testing. High quality evidence. Even though, Dx studies may not capture side effects as effectively as treatment trials, in case of major side-effects would occur likely they would be reported.	Rif resistance testing and results are a benefit that is

○ No included studies		associated with Xpert Ultra.

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	Effects of treatment on TB outcomes overall comes with high certainty. Treatment of drug sensitive TB is highly effective. Treatment of MDR TB can be effective as well, if quality assured.	Assuming that the false positives are appropriately treated. May extrapolate from Xpert that there is a lower pretreatment loss to follow up.

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Discordant results of repeat tests and confirmatory tests can cause confusion around what should be considered the reference or gold standard, particularly when specimen quality might be poor. Understanding and contextualizing discordant results require continuous training, experience and expertise. Establishing a thorough TB history of patients is uncommon and 'previously treated' defined differently.	Trace results should be described. And may not always lead to treatment. Discordant results inevitably happen with all of the tests used. Version control issues should be described (under implementation considerations). This was a panel judgment. Describe Nora Engel's study.

Certainty of effects

What is the overall certainty of the evidence of effects of the test?

Judgement	Research evidence	Additional considerations

○ Very low ○ Low ○ Moderate ○ High ○ No included studies		List the certainty of the evidence separately for the elements that we described. High certainty for accuracy, direct benefits, management effects, uncertain for the link of test results to management.
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability	Participants assign great value to the ability of Xpert to improve the diagnosis of drug resistant TB and the impact on patients if they cannot access testing for drug resistance through Xpert. The impact on case notification and the value of Xpert for finding more TB was less clear owing to widespread clinical treatment, prolonged TATs and the challenges with feasibility and utilization of Xpert. While Xpert has eased laboratory work through convenience and automation, this preference for Xpert in the laboratory can have undesired consequences for monitoring through microscopy or for reverting back to microscopy when Xpert machines are down. While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert.	
Balance of effects Does the balance between desirable and undesirable effects favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the interventio n ● Favours the intervention ○ Varies ○ Don't know	Summary of the above: Benefits vs Harms. Probably very little variation to how people value the outcomes.	

Resources required How large are the resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	Same as Xpert MTB/RIF. From USD 9.98 in Ethiopia (Tesfaye 2017) till EUR 110.75 in Germany (Diel 2016).	
Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Systematic review by A. Zwerling: Studies employed a variety of different modelling approaches, populations and settings. Variations in costing, effectiveness and epidemiological parameters were present across included studies making direct comparisons across studies challenging.	copy consideration from Xpert
Cost effectiveness Does the cost-effectiveness of the intervention favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the interventio n ○ Favours the intervention ○ Varies ○ No included studies	No study directly assessing cost-effectiveness of Xpert Ultra were identified.	False positives are possibly increased.

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	As test can be performed at decentralized levels of the health care system, it will likely increase health care equity.	

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Report on user perspectives on Xpert MTB/RIF and Ultra testing: results from qualitative research : Test is generally described as acceptable by key stakeholders. Trace complicates decision-making: laboratory and clinical management of trace results is not straightforward. Study participants reported challenges with obtaining a second fresh sample when patients had left the facilities or had since been put on treatment and could not produce sputum as easily. If repeat tests are conducted after trace, they cause confusion when the second test is also trace or negative. Some laboratory managers are unsure which result to report and clinicians need expertise and experience to conduct more extensive evaluation for trace patients. This presents challenges for peripheral settings and where TATs of confirmatory tests (DST, LPA) slow down clinical decision-making. Discordant results of repeat tests and confirmatory tests can cause confusion around what should be considered gold standard, particularly when specimen quality might be poor. Understanding and contextualizing discordant results require continuous training, experience and expertise. Establishing a thorough TB history of patients is uncommon and 'previously treated' defined differently with implications for potential of false positives results through Xpert testing. Clear parameters are needed of how to define previously treated patients, how to handle their Xpert results, and accurately capture outcomes in national databases.	Clinicians may be reluctant to implement treatment based on trace results. Qualitative data was limited. lack of country specific cost-effectiveness data may reduce acceptability for implementers. Trace results are considered more difficult to act on from a laboratory standpoint. Stigmatization was raised as a concern on the basis of the trace results.

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no	Compared to smear microscopy, users generally value the automation, convenience, higher biosafety levels and lesser human involvement that Xpert offers. The fact that it is a closed	

● Probably yes ○ Yes ○ Varies ○ Don't know	system with walk away time during the incubation (15') and machine run time (90') where lab technicians can do other testing in between was mentioned as well. Specifically for Xpert Ultra, the fact that Xpert Ultra takes less time can be helpful in some situations (for instance an active case finding setting with high throughput). As such, Xpert eased the work for lab technicians, adding a level of relief from reading hundreds of slides as well as reducing the room for errors.	

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	Accurate	**Very accurate**		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of test accuracy	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the management effects	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	**Favours the intervention**	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost-effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or	**Probably favours the intervention**	Favours the intervention	Varies	No included studies

	Judgement						
			the comparison				
Equity		Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	**Strong recommendation for the intervention**
○	○	○	○	●

Conclusions

Recommendation

In adults with signs and symptoms of pulmonary TB, the GDG **recommends using** Xpert Ultra MTB/RIF for the initial diagnosis of TB (as opposed to a microbiological reference standard). (strong recommendation, **high certainty of evidence for test accuracy**).

14 in favour of strong, 2 conditional, 1 abstention. Suggested by GRC is 80% majority for a strong recommendation (87.5% result here)

Subgroup considerations

Applies to all subgroups. Same provisos.

However, in patients with prior TB, the proportion of FP increases. This may be dealt with in the interpretation of trace results. The duration since treatment and diagnosis also impacts on the degree of positivity.

Implementation considerations

Risk of false positives may be higher.

Initial test for TB

PICO 2: Among children with signs and symptoms of pulmonary TB, seeking care at health care facilities should Xpert MTB/RIF / Ultra be used as an initial test for diagnosis of pulmonary TB and RR?

2.1 What is diagnostic accuracy of Xpert MTB/RIF for PTB and RR in children, as compares with MRS and composite reference standard (CRS)[1]?

Assessment

[1] Positive culture or a clinical decision to initiate treatment for tuberculosis

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Globally, an estimated 10.0 million (range, 9.0–11.1 million) people fell ill with TB in 2018. Children (aged <15 years) accounted for 11% of this burden.	

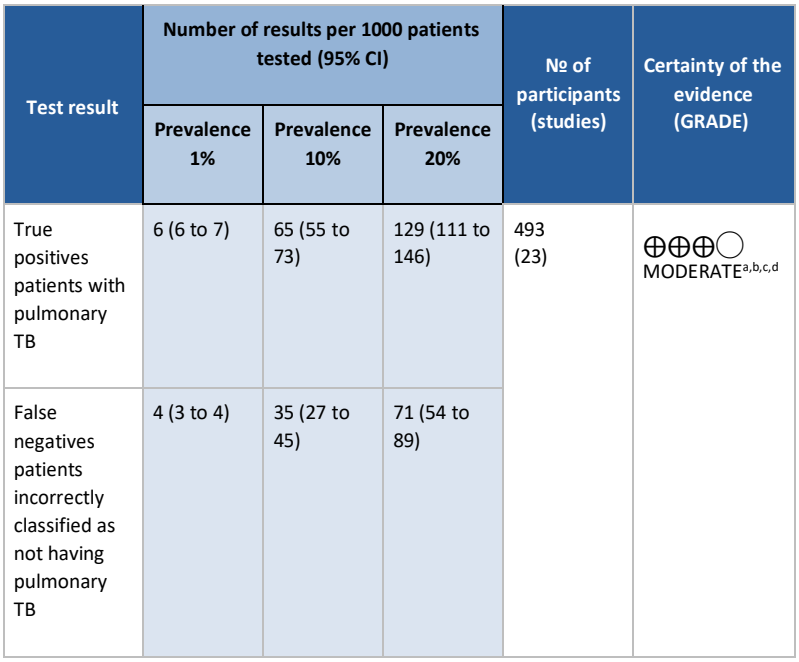
Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* Xpert MTB/RIF Sensitivity: 0.65 (95% CI: 0.55 to 0.73) Specificity: 0.99 (95% CI: 0.98 to 0.99)	

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know		

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 10%	Prevalence 20%		
True positives patients with pulmonary TB	6 (6 to 7)	65 (55 to 73)	129 (111 to 146)	493 (23)	⨁⨁⨁◯ MODERATE[a,b,c,d]
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	980 (971 to 985)	891 (883 to 896)	792 (785 to 796)	6119 (23)	⨁⨁⨁◯ MODERATE[e]	
False positives patients incorrectly classified as having pulmonary TB	10 (5 to 19)	9 (4 to 17)	8 (4 to 15)			

a. As assessed by QUADAS-2, 22 (95%) had low risk of bias.
b. As assessed by QUADAS-2, 8 studies (34%) had high or unclear concern about applicability because these patients were enrolled from tertiary care inpatient settings, which could lead to the enrollment of children with more advanced disease. Several of these studies (Nhu 2013 and Singh 2016 had among the highest sensitivities). We downgraded one level.
c. For individual studies, sensitivity estimates ranged from 27% to 100%. We thought that differences in enrolment criteria (different populations targeted), disease severity, different ages and settings could explain the heterogeneity. We did not downgrade.
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.

Rifampicin resistance detection, additional desirable effect.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence					Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	**Test result**	**Number of results per 1000 patients tested (95% CI)**			**№ of participants (studies)**	**Certainty of the evidence (GRADE)**	Subtest 108 FN in NFA considered by the panel still considered small (see EP for NFA) but FP are 0.
		Prevalence 1%	**Prevalence 10%**	**Prevalence 20%**			
	True positives patients with pulmonary TB	6 (6 to 7)	65 (55 to 73)	129 (111 to 146)	493 (23)	⨁⨁⨁◯ MODERATE[a,b,c,d]	
	False negatives patients incorrectly	4 (3 to 4)	35 (27 to 45)	71 (54 to 89)			

classified as not having pulmonary TB						
True negatives patients without pulmonary TB	980 (971 to 985)	891 (883 to 896)	792 (785 to 796)	6119 (23)	⨁⨁⨁◯ MODERATE[e]	
False positives patients incorrectly classified as having pulmonary TB	10 (5 to 19)	9 (4 to 17)	8 (4 to 15)			

a. As assessed by QUADAS-2, 22 (95%) had low risk of bias.
b. As assessed by QUADAS-2, 8 studies (34%) had high or unclear concern about applicability because these patients were enrolled from tertiary care inpatient settings, which could lead to the enrollment of children with more advanced disease. Several of these studies (Nhu 2013 and Singh 2016 had among the highest sensitivities). We downgraded one level.
c. For individual studies, sensitivity estimates ranged from 27% to 100%. We thought that differences in enrolment criteria (different populations targeted), disease severity, different ages and settings could explain the heterogeneity. We did not downgrade.
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.

Rifampicin resistance detection, additional desirable effect.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	No adverse events were associated with Xpert testing. High quality evidence. Even though, Dx trial may not capture side effects as effectively as treatment trials, in case of major side-effects would occur likely they would be reported.	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	No adverse events were associated with Xpert testing. High quality evidence. Even though, Dx trial may not capture side effects as effectively as treatment trials, in case of major side-effects would occur likely they would be reported. Additional benefit from Rif Resistance testing.	

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	Treatment of drug sensitive TB is highly effective. Treatment of MDR TB can be effective as well, if quality assured	

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert.	

Certainty of effects

What is the overall certainty of the evidence of effects of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies	Balance of the above	Label certainty by criterion

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	Participants assign great value to the ability of Xpert to improve the diagnosis of drug resistant TB and the impact on patients if they cannot access testing for drug resistance through Xpert. The impact on case notification and the value of Xpert for finding more TB was less clear owing to widespread clinical treatment, prolonged TATs and the challenges with feasibility and utilization of Xpert. While Xpert has eased laboratory work through convenience and automation, this preference for Xpert in the laboratory can have undesired consequences for monitoring through microscopy or for reverting back to microscopy when Xpert machines are down. While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert.	in absence of having reviewed available studies.

Balance of effects

Does the balance between desirable and undesirable effects favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ● Favours the intervention ○ Varies ○ Don't know	Summary of the above: Benefits vs Harms. Probably very little variation to how people value the outcomes.	

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	From USD 9.98 in Ethiopia (Tesfaye 2017) till EUR 110.75 in Germany (Diel 2016).	No studies were identified for cost effectiveness in children.

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Systematic review by A. Zwerling: Studies employed a variety of different modelling approaches, populations and settings. Variations in costing, effectiveness and epidemiological parameters were present across included studies making direct comparisons across studies challenging.	
Cost effectiveness Does the cost-effectiveness of the intervention favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	Four studies were identified assessing the use of Xpert MTB/RIF among PLHIV with signs and symptoms of TB (3–5,18). Studies were conducted in countries with high HIV prevalence including South Africa, Ethiopia and Malawi. All reported Xpert would likely be cost-effective in these populations but to varying degrees and conditions of implementation. No studies assessed children specifically among these studies. Four studies among hospitalized patients were identified, 2 from the USA (21,22), 1 from Germany (20) and 1 study from China (Hong Kong SAR) (28). All 4 studies concluded that replacement of SSM with Xpert would results in cost-savings driven largely from high hospitalization costs associated with respiratory isolation. No studies assessed children specifically among these studies. Fifteen studies assessing cost-effectiveness of Xpert among persons presenting to primary health care facilities across Sub-Saharan Africa and Brazil. While early studies found Xpert would likely be cost-effective (albeit using a range of willingness to pay thresholds across different countries, several concerns around cost-effectiveness have been raised by subsequent analyses. Inclusion of downstream costs associated with MDR-TB and HIV treatment and care has been shown to lead to increased ICERs and increased total expenditures. Costs associated with scale-up of Xpert have been estimated to result in an important increase relative to existing TB and HIV programme budgets and in many countries may not be deemed affordable despite ICERs for Xpert approaches being under willingness to pay thresholds. Studies have highlighted the importance of implementation conditions, including existing standard of care, levels of empirical treatment, TB prevalence among presumptive patients being tested, and test volume as highly influential variables on cost-effectiveness results. Results from individual studies are summarized below. While some studies employd a population based approach no studies specifically addressed children.	The panel suggest to no extrapolate to children.
Equity What would be the impact on health equity?		
Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact	This evidence has not been reviewed. As test can be performed at all levels of the health care system, it will likely increase health care equity.	This was a judgment by the panel.

● Probably increased ○ Increased ○ Varies ○ Don't know		

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Report on user perspectives on Xpert MTB/RIF and Ultra testing: results from qualitative research : Test is generally described as acceptable by key stakeholders. Discordant results of repeat tests and confirmatory tests can cause confusion around what should be considered gold standard, particularly when specimen quality might be poor. Understanding and contextualizing discordant results require continuous training, experience and expertise. Establishing a thorough TB history of patients is uncommon and 'previously treated' defined differently with implications for potential of false positives results through Xpert testing. Clear parameters are needed of how to define previously treated patients, how to handle their Xpert results, and accurately capture outcomes in national databases.	

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Compared to smear microscopy, users generally value the automation, convenience, higher biosafety levels and lesser human involvement that Xpert offers. The fact that it is a closed system with walk away time during the incubation (15') and machine run time (90') where lab technicians can do other testing in between was mentioned as well. As such, Xpert eased the work for lab technicians, adding a level of relief from reading hundreds of slides as well as reducing the room for errors. Persistent underutilization of Xpert machines is compounded by the challenges of delays due to sample transport, module break down, stock-out of cartridges or complicated diagnostic algorithms. Government commitment to ensure functioning infrastructure and power; supply of cartridges, functioning laboratory services; investment in expertise to handle (discordant) results; better repair services; staff with monitoring capacities; functioning sample transport; sustainable funding models and transparent donor agreements; and simple diagnostic algorithms; à those interact and reinforce each other determining utilizationmmaintance and stock outs. Diagnostic algorithms that are simple to follow in a specific facility (f.i. test all those with presumptive TB) are more feasible and enhance utilization, but this simplicity is crucially dependent on cost and supplies.	

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of the management effects	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	**Favours the intervention**	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost-effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	**Strong recommendation for the intervention**
○	○	○	○	●

Conclusions

Recommendation

In children with signs and symptoms of pulmonary TB, the GDG **recommends using** Xpert MTB rather than **culture** as the initial diagnostic test for TB in sputum (**moderate certainty of evidence in test accuracy**), gastric aspirate (low certainty of the evidence **in test accuracy** from children with HIV), nasopharyngeal aspirate (**moderate certainty of the evidence in test accuracy**), or stool (**low certainty of the evidence in test accuracy**) (strong recommendation).

In children with signs and symptoms of pulmonary TB, the GDG **recommends using** Xpert Ultra rather than **culture** for the initial diagnosis of TB in sputum (**low certainty of evidence in test accuracy**), nasopharyngeal aspirate (**very low certainty of the evidence in test accuracy**) (strong recommendation).

Remarks: Sputum includes induced sputum. Studies assessing the impact of Xpert on outcomes in children lacking.

The GDG felt that the choice of the test is dependent on the acceptability (for children, HCW, other stakeholders) and feasibility of conducting it in the local context. The certainty of evidence is higher for sputum and nasopharyngeal aspirates for Xpert. Describe the differential accuracy of sputum versus NPA. There was no evidence for ther specimens for Xpert Ultra.

Includes children living with HIV (for Xpert). This includes consideration about the direct benefit from RR testing in sputum samples (very low certainty) which the panel felt can be extrapolated to other samples. Explain the use of stool being not "on demand" and may be more challenging to obtain.

Justification

life threatening situation in children

Subgroup considerations

In children in whom sputum samples cannot be obtained, alternative testing should be obtained.

Test performance in children with HIV with CD4 low may be different from that observed here.

Implementation considerations

Specimen collection and their quality needs to be ensured. Sputum induction in children is challenging and requires training of staff and access to suplies may be limited.

Implementation support (including specimen transporation) for primary care settings may be particularly required.

Induced sputum collection is considered invasive in children.

Research priorities

Performance of the test in different children agegroups.

Values systematic reviews.

PICO 3: Among adults with signs and symptoms of extra-pulmonary (EP) TB, seeking care at health care facilities should Xpert MTB/RIF / Ultra used as an initial test for diagnosis of EP TB and RR?

3.1 What is diagnostic accuracy of Xpert MTB/RIF for EP TB and RR in adults, as compares with MRS and CRS?

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	EP TB is a problem	
Test accuracy How accurate is the test?		
Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* Xpert MTB/RIF Sensitivity: 0.81 (95% CI: 0.62 to 0.92) Specificity: 0.96 (95% CI: 0.90 to 0.98)	
Desirable Effects How substantial are the desirable anticipated effects?		
Judgement	Research evidence	Additional considerations

○ Trivial
○ Small
○ Moderate
● Large
○ Varies
○ Don't know

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 20%		
True positives patients with lymph node TB	20 (16 to 23)	81 (62 to 92)	162 (124 to 184)	377 (4)	⨁⨁◯◯ LOW[a,b]
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	935 (878 to 958)	863 (811 to 885)	767 (721 to 786)	302 (4)	⨁⨁◯◯ LOW[c,d]
False positives patients incorrectly classified as having lymph node TB	40 (17 to 97)	37 (15 to 89)	33 (14 to 79)		

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. The number of participants were very few. The wide 95% CrI for false negatives and true positives may lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.
c. 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.
d. As assessed by QUADAS-2, we answered the question: is the reference standard likely to correctly classify the target condition? as unclear for all studies because the composite reference standard was defined according to the primary study authors and therefore was not uniform. In addition, composite reference standards have been shown to over and under estimate diagnostic accuracy.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations

○ Large
○ Moderate
● Small
○ Trivial
○ Varies
○ Don't know

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 20%		
True positives patients with lymph node TB	20 (16 to 23)	81 (62 to 92)	162 (124 to 184)	377 (4)	⨁⨁◯◯ LOW[a,b]
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	935 (878 to 958)	863 (811 to 885)	767 (721 to 786)	302 (4)	⨁⨁◯◯ LOW[c,d]
False positives patients incorrectly classified as having lymph node TB	40 (17 to 97)	37 (15 to 89)	33 (14 to 79)		

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. The number of participants were very few. The wide 95% CrI for false negatives and true positives may lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.
c. 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.
d. As assessed by QUADAS-2, we answered the question: is the reference standard likely to correctly classify the target condition? as unclear for all studies because the composite reference standard was defined according to the primary study authors and therefore was not uniform. In addition, composite reference standards have been shown to over and under estimate diagnostic accuracy.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations

○ Very low
● Low
○ Moderate
○ High
○ No included studies

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 2%	Effect per 1000 patients/year for pre-test probability of 10%	Effect per 1000 patients/year for pre-test probability of 20%	Importance
True positives	cross-sectional		20 (16 to 23)	81 (62 to 92)	162 (124 to 184)	

False negatives	(cohort type accuracy study)	⨁⨁◯◯ LOW[a,b]	5 (2 to 9)	19 (8 to 38)	38 (16 to 76)	
True negatives	cross-sectional (cohort type accuracy study)	⨁⨁◯◯ LOW[c,d]	935 (878 to 958)	863 (811 to 885)	767 (721 to 786)	
False positives			40 (17 to 97)	37 (15 to 89)	33 (14 to 79)	

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. The number of participants were very few. The wide 95% CrI for false negatives and true positives may lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.
c. 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.
d. As assessed by QUADAS-2, we answered the question: is the reference standard likely to correctly classify the target condition? as unclear for all studies because the composite reference standard was defined according to the primary study authors and therefore was not uniform. In addition, composite reference standards have been shown to over and under estimate diagnostic accuracy.

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ● Low ○ Moderate ○ High ○ No included studies	No adverse events were associated with Xpert testing. High quality evidence. Even though, Dx trial may not capture side effects as effectively as treatment trials, in case of major side-effects would occur likely they would be reported.	GDG considered no direct harm

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert.	

Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Discordant results of repeat tests and confirmatory tests can cause confusion around what should be considered gold standard, particularly when specimen quality might be poor. Understanding and contextualizing discordant results require continuous training, experience and expertise. Establishing a thorough TB history of patients is uncommon and 'previously treated' defined differently.	
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Treatment of drug sensitive TB is highly effective. Treatment of MDR TB can be effective as well, if quality assured	list evidence separately
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability	Participants assign great value to the ability of Xpert to improve the diagnosis of drug resistant TB and the impact on patients if they cannot access testing for drug resistance through Xpert. The impact on case notification and the value of Xpert for finding more TB was less clear owing to widespread clinical treatment, prolonged TATs and the challenges with feasibility and utilization of Xpert. While Xpert has eased laboratory work through convenience and automation, this preference for Xpert in the laboratory can have undesired consequences for monitoring through microscopy or for reverting back to microscopy when Xpert machines are down. While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert.	
Balance of effects Does the balance between desirable and undesirable effects favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations

○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know		
Resources required How large are the resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderat e savings ○ Large savings ● Varies ○ Don't know	Compared to PTB samples, cost per case diagnosed using Xpert was higher among EPTB samples, with only one study identified from China.	
Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	Only one economic evaluation among extrapulmonary TB was identified, this study was conducted in a national TB referral hospital in Beijing China, and results are likely not generalizable across different countries and settings.	needs to be evaluated for quality
Cost effectiveness Does the cost-effectiveness of the intervention favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations

○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ No included studies	No studies estimated an ICER specifically for use of Xpert in EPTB compared to SSM.	
Equity What would be the impact on health equity?		
Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Access: improved but not everybody who needs it can access Xpert	
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Acceptability: generally high among patients and clinicians •Skill to clinically diagnose affected by Xpert? •The confidence into test (esp pos results) is challenged by discordant/trace results	
Feasibility Is the intervention feasible to implement?		
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes	Feasibility depends on:	No all clinicians are credentialed or able to perform the procedure,

○ Yes ○ Varies ○ Don't know	Government commitment to ensure functioning infrastructure and power; supply of cartridges, functioning laboratory services; investment in expertise to handle (discordant) results; better repair services; staff with monitoring capacities; functioning sample transport; sustainable funding models and transparent donor agreements; and simple diagnostic algorithms; à those interact and reinforce each other determining utilization **Simple to use in the lab does not automatically translate into feasibility**	referral to facility may reduce uptake (travel)

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	**Low**	Moderate	High			No included studies
Certainty of evidence of test accuracy	Very low	**Low**	Moderate	High			No included studies
Certainty of evidence of the management effects	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	Moderate	High			**No included studies**
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the interventio n	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	**Very low**	Low	Moderate	High			No included studies

	Judgement						
Cost-effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the interventio n	Varies	No included studies
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In adults with signs and symptoms of extrapulmonary TB, the panel **suggests using** XPERT MTB/RIF in LNA, LN Bx, pleural fluid, peritoneal fluid, pericardial fluid, blood, bones and joint and urine as an initial diagnostic test for the corresponding extrapulmonary TB (conditional recommendation with low certainty test accuracy for LNA and very low certainty in the test accuracy for LNBx, moderate for pleural fluid, low for peritoneal fluid, very low for pericardial, bones low, urine low, blood very low).

Remark: Clinical judgment and pretest probability should guide treatment and in a high pretest probability setting a negative test result will not rule out the condition. The GDG extrapolated that the composite reference standard would lead to similar results when LNBx is compared to LNA. High certainty in accuracy of Rif Resistance but no separate recommendation.

Blood only evaluated in PLHIV and processing specification (remark), also using third generation Xpert, very low certainty based on very few numbers.

In children the GDG the was evidence for LNA/Bx, the GDG judged the evidence to suggest the same effects and the recommendation in children is conditional as in adults (conditional recommendation, very low certainty evidence for test accuracy).

The panel was very uncertain about the use of blood - separate recommendation - in this population it may be used as an initial diagnostic test - the panel did not feel comfortable extrapolating to other patient populations.

Implementation considerations

Implementation challenges for LNBx because of lack of linkages between professionals (histopathology and access to Xpert)

Research priorities

Further studies

PICO 5: Among people with signs and symptoms of pulmonary TB, seeking care at health care facilities does repeated Xpert (Ultra) tests on subsequent samples provide any increase in sensitivity/specificity as an initial test for diagnosis of pulmonary TB and RR?

5.1 One Xpert MTB/RIF vs. more than one Xpert MTB/RIF to diagnose PTB in children with signs and symptoms of PTB, against a MRS?

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		
Test accuracy How accurate is the test?		
Judgement	Research evidence	Additional considerations
○ Very inaccurate ● Inaccurate ○ Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* one Xpert MTB/RIF Sensitivity: 0.46 (95% CI: 0.35 to 0.58) Specificity: 1.00 (95% CI: 0.99 to 1.00) more than one Xpert MTB/RIF Sensitivity: 0.59 (95% CI: 0.43 to 0.73) Specificity: 0.99 (95% CI: 0.98 to 1.00)	
Desirable Effects How substantial are the desirable anticipated effects?		
Judgement	Research evidence	Additional considerations

○ Trivial
● Small
○ Moderate
○ Large
○ Varies
○ Don't know

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%		Prevalence 10%		Prevalence 20%			
	one Xpert MTB/RIF	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		
True positives patients with pulmonary TB	5 (3 to 6)	6 (4 to 7)	46 (35 to 58)	59 (43 to 73)	92 (70 to 116)	118 (86 to 146)	180 (5)	⨁⨁◯◯ LOW[a,b]
	1 fewer TP in one Xpert MTB/RIF		13 fewer TP in one Xpert MTB/RIF		26 fewer TP in one Xpert MTB/RIF			
False negatives patients incorrectly classified as not having pulmonary TB	5 (4 to 7)	4 (3 to 6)	54 (42 to 65)	41 (27 to 57)	108 (84 to 130)	82 (54 to 114)		
	1 more FN in one Xpert MTB/RIF		13 more FN in one Xpert MTB/RIF		26 more FN in one Xpert MTB/RIF			
True negatives patients without pulmonary TB	989 (980 to 990)	980 (970 to 990)	899 (891 to 900)	891 (882 to 900)	799 (792 to 800)	792 (784 to 800)	1939 (5)	⨁⨁⨁⨁ HIGH
	9 more TN in one Xpert MTB/RIF		8 more TN in one Xpert MTB/RIF		7 more TN in one Xpert MTB/RIF			
False positives patients incorrectly classified as having pulmonary TB	1 (0 to 10)	10 (0 to 20)	1 (0 to 9)	9 (0 to 18)	1 (0 to 8)	8 (0 to 16)		
	9 fewer FP in one Xpert MTB/RIF		8 fewer FP in one Xpert MTB/RIF		7 fewer FP in one Xpert MTB/RIF			

a. As assessed by QUADAS-2, 2 studies (40%) had high or unclear concern about applicability because these patients were enrolled from tertiary care centers or exclusively inpatient settings, which could lead to the enrollment of children with more advanced disease.
b. This degree of imprecision may result in different clinical decisions at different ends of the confidence limit.

Desirable effects are more TN and fewer FP.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ○ Small ○ Trivial ● Varies ○ Don't know	<table><tr><th rowspan="3">Test result</th><th colspan="6">Number of results per 1000 patients tested (95% CI)</th><th rowspan="3">№ of participants (studies)</th><th rowspan="3">Certainty of the evidence (GRADE)</th></tr><tr><th colspan="2">Prevalence 1%</th><th colspan="2">Prevalence 10%</th><th colspan="2">Prevalence 20%</th></tr><tr><th>one Xpert MTB/RIF</th><th>more than one Xpert MTB/RIF</th><th>one Xpert MTB/RIF</th><th>more than one Xpert MTB/RIF</th><th>one Xpert MTB/RIF</th><th>more than one Xpert MTB/RIF</th></tr><tr><td rowspan="2">True positives patients with pulmonary TB</td><td>5 (3 to 6)</td><td>6 (4 to 7)</td><td>46 (35 to 58)</td><td>59 (43 to 73)</td><td>92 (70 to 116)</td><td>118 (86 to 146)</td><td rowspan="4">180 (5)</td><td rowspan="4">⨁⨁◯◯ LOW[a,b]</td></tr><tr><td colspan="2">1 fewer TP in one Xpert MTB/RIF</td><td colspan="2">13 fewer TP in one Xpert MTB/RIF</td><td colspan="2">26 fewer TP in one Xpert MTB/RIF</td></tr><tr><td rowspan="2">False negatives patients incorrectly classified as not having pulmonary TB</td><td>5 (4 to 7)</td><td>4 (3 to 6)</td><td>54 (42 to 65)</td><td>41 (27 to 57)</td><td>108 (84 to 130)</td><td>82 (54 to 114)</td></tr><tr><td colspan="2">1 more FN in one Xpert MTB/RIF</td><td colspan="2">13 more FN in one Xpert MTB/RIF</td><td colspan="2">26 more FN in one Xpert MTB/RIF</td></tr><tr><td rowspan="2">True negatives patients without pulmonary TB</td><td>989 (980 to 990)</td><td>980 (970 to 990)</td><td>899 (891 to 900)</td><td>891 (882 to 900)</td><td>799 (792 to 800)</td><td>792 (784 to 800)</td><td rowspan="4">1939 (5)</td><td rowspan="4">⨁⨁⨁⨁ HIGH</td></tr><tr><td colspan="2">9 more TN in one Xpert MTB/RIF</td><td colspan="2">8 more TN in one Xpert MTB/RIF</td><td colspan="2">7 more TN in one Xpert MTB/RIF</td></tr><tr><td rowspan="2">False positives patients incorrectly classified as having pulmonary TB</td><td>1 (0 to 10)</td><td>10 (0 to 20)</td><td>1 (0 to 9)</td><td>9 (0 to 18)</td><td>1 (0 to 8)</td><td>8 (0 to 16)</td></tr><tr><td colspan="2">9 fewer FP in one Xpert MTB/RIF</td><td colspan="2">8 fewer FP in one Xpert MTB/RIF</td><td colspan="2">7 fewer FP in one Xpert MTB/RIF</td></tr></table> a. As assessed by QUADAS-2, 2 studies (40%) had high or unclear concern about applicability because these patients were enrolled from tertiary care centers or exclusively inpatient settings, which could lead to the enrollment of children with more advanced disease.	Undesirable effects are moderate: 26 fewer TP, 26 more FN for the 10 and 20% prevalence setting. Trivial for low pretest probability.

	b. This degree of imprecision may result in different clinical decisions at different ends of the confidence limit.	

Certainty of the evidence of test accuracy What is the overall certainty of the evidence of test accuracy?		
Judgement	Research evidence	Additional considerations
○ Very low ● Low ○ Moderate ○ High ○ No included studies		

Certainty of the evidence of test's effects What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	No adverse events were associated with Xpert testing. High quality evidence. Even though, Dx trial may not capture side effects as effectively as treatment trials, in case of major side-effects would occur likely they would be reported.	Direct benefit is having only one test for sputum.

Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert. Trace complicates decision-making	

Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
Judgement	Research evidence	Additional considerations

○ Very low ○ Low ○ Moderate ○ High ● No included studies		Discordant results of repeat tests and confirmatory tests can cause confusion around what should be considered gold standard, particularly when specimen quality might be poor. Understanding and contextualizing discordant results require continuous training, experience and expertise. Establishing a thorough TB history of patients is uncommon and 'previously treated' defined differently. Assumption is that the clinicians would act the same whether or not one or two test results were obtained.
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies		separate out the evidence rating for the different criteria
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability	Participants assign great value to the ability of Xpert to improve the diagnosis of drug resistant TB and the impact on patients if they cannot access testing for drug resistance through Xpert. The impact on case notification and the value of Xpert for finding more TB was less clear owing to widespread clinical	

○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	treatment, prolonged TATs and the challenges with feasibility and utilization of Xpert. While Xpert has eased laboratory work through convenience and automation, this preference for Xpert in the laboratory can have undesired consequences for monitoring through microscopy or for reverting back to microscopy when Xpert machines are down. While clinicians' confidence in Xpert results is rather high, the challenges with feasibility and utilization mean clinicians are at times deterred from ordering Xpert.	

Balance of effects

Does the balance between desirable and undesirable effects favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the interventio n ○ Favours the intervention ● Varies ○ Don't know		moderate and high pretest probability: probably favours the comparison (in favour of two tests) low pretest probability setting: probability favours the intervention (in favour of one test)

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ● Moderate savings ○ Large savings ○ Varies ○ Don't know		No direct cost data identified. The panel assumed that twice testing is more costly. transport and parents' cost of returning for second testing.

Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies		
Cost effectiveness Does the cost-effectiveness of the intervention favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the interventio n ○ Favours the intervention ○ Varies ● No included studies	No direct research evidence identified.	
Equity What would be the impact on health equity?		
Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know		Second sample may be difficult to have patient return. Twice testing however is identifying additional cases and that means that more children

		are allowed to be treated.
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	No direct research evidence identified.	
Feasibility Is the intervention feasible to implement?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	No direct research evidence identified.	

Summary of judgements

	JUDGEMENT						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	**Inaccurate**	Accurate	Very accurate		Varies	Don't know
Desirable effects	Trivial	**Small**	Moderate	Large		Varies	Don't know
Undesirable effects	Large	Moderate	Small	Trivial		**Varies**	Don't know
Certainty of the evidence of test accuracy	Very low	**Low**	Moderate	High			No included studies
Certainty of evidence of test accuracy	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the management effects	Very low	Low	Moderate	**High**			No included studies

	Judgement						
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the interventio n	**Varies**	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	**Moderate savings**	Large savings	Varies	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost-effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	Probably yes	**Yes**		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
○	○	○	○	○

Conclusions

Recommendation

Low pretest probability setting: In children with signs and symptoms of pulmonary TB and a negative Xpert on the first initial test, the GDG suggests to not repeat testing with Xpert in sputum, gastric fluid, NPA and stool **(conditional recommendation, low certainty of evidence in test accuracy for sputum and very low for other other specimens)**.

Moderate and high pretest probability setting: In children with signs and symptoms of pulmonary TB in a moderate or high pretest probability setting and a negative Xpert on the first initial test, the GDG suggests to one repeat/total of two test(s) with Xpert in sputum, gastric fluid, NPA and stool **(conditional recommendation, low certainty of evidence in test accuracy for sputum and very low for other other specimens)**.

The GDG felt that the choice of the is dependent on the acceptability (for children, HCW, other stakeholders) and feasibility of conducting it in the local context. **The evidence reviewed evaluated repeating the same test on the same type of specimen. However, from the data reviewed on comparing single tests on different specimen, there appears to be no difference regardless of which second specimen is obtained.**

Includes children living with HIV (for Xpert). This includes consideration about the direct benefit from RR testing in sputum samples (very low certainty) which the panel felt can be extrapolated to other samples.

Applicable to: moderate or high pretest setting: If the first test is positive do not repeat the test .

In settings with moderate to high pretest probability, the incremental yield of more than two tests is unkown.

One study evaluated repeated testing on different specimen types.

Implementation considerations

Community health workers to support sputum collection in children at home (not in all settings) and subsidies can be provided

Clinician judgment is required to interpret the context in which the test is obtained (e.g. high pretest probability).

Research priorities

Testing on repeated speciment types.

Testing on pooled specimen samples

PICO 6: Among adults in a population-based TB disease prevalence survey with symptoms or chest X-ray abnormalities suggestive of pulmonary TB, should Xpert MTB/RIF/Ultra alone, be used to define the case of active TB disease[2]?

6.1 Xpert MTB/RIF to diagnose PTB in adults in general population following a positive TB symptom screen or chest X-ray with lung abnormalities or both, against a MRS.

Question

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?	
Population:	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
Intervention:	Xpert MTB/RIF
Role of the test:	replacement
Setting:	community
Conflict of interests:	Petra

Assessment

[2] Tuberculosis prevalence surveys: a handbook. Geneva: World Health Organization; 2011 (WHO/HTM/TB/2010.17, http://apps.who.int/iris/bitstream/10665/44481/1/9789241548168_eng.pdf?ua=1&ua=1, accessed 1 February 2020).

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	TB diagnosis problem is a priority	

Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* Xpert MTB/RIF Sensitivity: 0.73 (95% CI: 0.62 to 0.82) Specificity: 0.99 (95% CI: 0.98 to 0.99)	

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ○ Large ● Varies ○ Don't know		Detect cases early and with Rif Resistance Varies: moderate in low and moderate prevalence large in high prevalence Rapidity of testing is not of importance but Rif Resistance is an important

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 3%	Prevalence 7%		
True positives patients with pulmonary tuberculosis	7 (6 to 8)	22 (19 to 25)	51 (43 to 57)	867 (4)	⨁⨁⨁⨁ HIGH[a,b]
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	980 (970 to 980)	960 (951 to 960)	921 (911 to 921)	48689 (4)	⨁⨁⨁⨁ HIGH[a]	
False positives patients incorrectly classified as having pulmonary tuberculosis	10 (10 to 20)	10 (10 to 19)	9 (9 to 19)			

a. Data from Namibia were excluded owing to inconsistencies in the diagnostic algorithm. We did not downgrade for risk of bias. This was a judgement.
b. 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 explain in part the inconsistency owing to lower HIV prevalence. We did not downgrade for inconsistency.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know		In low prevalence more FP In high prevalence many FN

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 3%	Prevalence 7%		
True positives patients with pulmonary tuberculosis	7 (6 to 8)	22 (19 to 25)	51 (43 to 57)	867 (4)	⨁⨁⨁⨁ HIGH[a,b]
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	980 (970 to 980)	960 (951 to 960)	921 (911 to 921)	48689 (4)	

pulmonary tuberculosis					⊕⊕⊕⊕ HIGH[a]	
False positives patients incorrectly classified as having pulmonary tuberculosis	10 (10 to 20)	10 (10 to 19)	9 (9 to 19)			

a. Data from Namibia were excluded owing to inconsistencies in the diagnostic algorithm. We did not downgrade for risk of bias. This was a judgement.
b. 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 explain in part the inconsistency owing to lower HIV prevalence. We did not downgrade for inconsistency.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies		

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 1%	Effect per 1000 patients/year for pre-test probability of 3%	Effect per 1000 patients/year for pre-test probability of 7%	Importance
True positives	cross-sectional (cohort type accuracy study)	⊕⊕⊕⊕ HIGH[a,b]	7 (6 to 8)	22 (19 to 25)	51 (43 to 57)	
False negatives			3 (2 to 4)	8 (5 to 11)	19 (13 to 27)	
True negatives	cross-sectional (cohort type accuracy study)	⊕⊕⊕⊕ HIGH[a]	980 (970 to 980)	960 (951 to 960)	921 (911 to 921)	
False positives			10 (10 to 20)	10 (10 to 19)	9 (9 to 19)	

a. Data from Namibia were excluded owing to inconsistencies in the diagnostic algorithm. We did not downgrade for risk of bias. This was a judgement.

	b. 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 explain in part the inconsistency owing to lower HIV prevalence. We did not downgrade for inconsistency.	
Certainty of the evidence of test's effects What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies		
Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies		
Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies		Patients may not accept treatment if no symptoms loss to follow up may be high
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations

○ Very low ○ Low ○ Moderate ○ High ○ No included studies		list separately, we no information on people important outcomes
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability	No research evidence searched for.	
Balance of effects Does the balance between desirable and undesirable effects favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the interventio n ○ Favours the intervention ○ Varies		High certainty in accuracy but no information about how one will act on the test outcomes.
Resources required How large are the resource requirements (costs)?		
Judgement	Research evidence	Additional considerations

○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	No cost studies were identified.	some members of the GDG suggested there may be savings Cost of Xpert may be lower than culture - depends on setting.

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No cost effectiveness studies were identified.	

Cost effectiveness

Does the cost-effectiveness of the intervention favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	No cost effectiveness studies were identified.	

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	No research evidence searched for.	patients who are tested with Xpert (if access to treatment) are more likely to receive fast treatment. more loss to follow in culture group
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	No research evidence searched for.	patients: yes clinicians: yes payers: yes (input from GF)
Feasibility Is the intervention feasible to implement?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	No research evidence searched for.	probably yes because in some settings capacity to do Xpert might be more human resource intensive very dependent on existing infrastructure in particular for those countries that would incur high cost Some countries may not be able to procure Xpert.

		one member said it is feasible and one varies

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	Large		**Varies**	Don't know
Undesirable effects	Large	**Moderate**	Small	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of test accuracy	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the management effects	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the interventio n	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost-effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the interventio n	Varies	**No included studies**

	Judgement						
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In adults in the general population who had both a TB symptom screen and chest x-ray and either a positive chest-ray or positive symptom screen, the GDG recommends using Xpert MTB/RIF or Ultra RiF rather than culture as the initial test for pulmonary tuberculosis (**conditional recommendation, high certainty of the evidence in test accuracy for Xpert and moderate for Ultra**).

Remarks: There are concerns about loosing the capacity for culture. Trace was considered as negative in these studies. There may be slighlty more positives in Ultra Rif

Xpert positive: treat

Xpert negative: reevaluate and look at differentials

culture positive: treat

Culture negative: reevaluate and look at differentials

Implementation considerations

Scaling up Xpert would reduce the availabilty of labs conducting tests that are required in addition to initial diagnosis

Transport is easier for Xpert

Not a replacement for culture because of other DST

Research priorities

Trials - assessment of pretest probability

comparison of false positives in culture

3.2 Evidence-to-decision tables: Truenat® MTB, MTB Plus and MTB-Rif Dx

PICO 7: Among people with signs and symptoms of pulmonary TB, seeking care at health care facilities should Molbio Truenat® MTB, MTB Plus and MTB-Rif Dx used as an initial test for diagnosis of pulmonary TB and RR?

7.1 Truenat MTB to diagnose PTB in adults with signs and symptoms of PTB, against a MRS standard?

Question

Should Truenat MTB be used to diagnose pulmonary tuberculosis in adults with signs and symptoms of pulmonary TB, against a microbiological reference standard?	
Population:	adults with signs and symptoms of pulmonary TB, against a microbiological reference standard
Intervention:	Truenat MTB
Conflict of interests:	Ezio, Kumar

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Xpert MTB/RIF is WHO-recommended rapid tests that simultaneously detect tuberculosis and rifampicin resistance in people with signs and symptoms of tuberculosis and are suitable for use at lower levels of the health system. This systematic review assessed the diagnostic accuracy of Xpert MTB/RIF and Xpert Ultra for detecting tuberculosis and rifampicin resistance from pulmonary specimens in adults. There were an estimated 10 million incident cases of tuberculosis in 2018 and of the 7 million reported cases, 85% involved the lungs (WHO Global Tuberculosis Report 2019). In 2018, there were about half a million new cases of rifampicin-resistant TB, and of these, 78% had multidrug-resistant TB (WHO Global Tuberculosis Report 2019). A previous Cochrane Review found Xpert MTB/RIF sensitive and specific for pulmonary tuberculosis, although sensitivity was decreased in paucibacillary samples (Steingart 2014).	

Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* Truenat MTB Sensitivity: 0.73 (95% CI: 0.68 to 0.78) Specificity: 0.98 (95% CI: 0.97 to 0.99)	On reference level both intermediate and final results of Truenat and Xpert have correlated When compared to the preliminary data, the descrease in sensitivity lead the GDG to a judgment of accurate for this comparison.

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know		

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 30%		
True positives patients with pulmonary tuberculosis	18 (17 to 20)	73 (68 to 78)	220 (203 to 235)	258 (1)	⨁⨁⨁◯ MODERATE[a,b]
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	955 (945 to 961)	881 (872 to 887)	685 (678 to 690)	1078 (1)	

pulmonary tuberculosis					⨁⨁⨁⨁ HIGH[a]	
False positives patients incorrectly classified as having pulmonary tuberculosis	20 (14 to 30)	19 (13 to 28)	15 (10 to 22)			

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. 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.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know		

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 30%		
True positives patients with pulmonary tuberculosis	18 (17 to 20)	73 (68 to 78)	220 (203 to 235)	258 (1)	⨁⨁⨁◯ MODERATE[a,b]
False negatives patients incorrectly classified as not having	7 (5 to 8)	27 (22 to 32)	80 (65 to 97)		

pulmonary tuberculosis						
True negatives patients without pulmonary tuberculosis	955 (945 to 961)	881 (872 to 887)	685 (678 to 690)	1078 (1)	⨁⨁⨁⨁ HIGH[a]	
False positives patients incorrectly classified as having pulmonary tuberculosis	20 (14 to 30)	19 (13 to 28)	15 (10 to 22)			

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. 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.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies		

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 2%	Effect per 1000 patients/year for pre-test probability of 10%	Effect per 1000 patients/year for pre-test probability of 30%	Importance
True positives	cross-sectional (cohort type accuracy study)	⨁⨁⨁◯ MODERATE[a,b]	18 (17 to 20)	73 (68 to 78)	220 (203 to 235)	
False negatives			7 (5 to 8)	27 (22 to 32)	80 (65 to 97)	

True negatives	cross-sectional (cohort type accuracy study)	⊕⊕⊕⊕ HIGH[a]	955 (945 to 961)	881 (872 to 887)	685 (678 to 690)	
False positives			20 (14 to 30)	19 (13 to 28)	15 (10 to 22)	

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. 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.

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	No adverse events were associated with Truenat testing. Moderate quality evidence. Even though, Dx studies may not capture side effects as effectively as treatment trials, in case of major side-effects would occur likely they would be reported.	None additional

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ● High ○ No included studies	Similar to Xpert/Ultra. Effects of treatment on TB outcomes overall comes with high certainty. Treatment of drug sensitive TB is highly effective. Treatment of MDR TB can be effective as well, if quality assured.	

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low	No included studies	

○ Moderate ○ High ● No included studies		
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies		list certainty of the evidence separately
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability	There is no important uncertainty or variability in how much people value main outcomes	
Balance of effects Does the balance between desirable and undesirable effects favour the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the interventio		GDG members suggested both favours and probably favours the intervention. Suggested benefits were sens and spec were high. Direct evidence on patient outcomes is lacking.

○ Favours the intervention ○ Varies ○ Don't know		Main concerns were also related to the low certainty in the sensitivity results. 15 voting (2 COI) 13 probably favours 1 favours 1 abstention

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderat e savings ○ Large savings ● Varies ○ Don't know	On study from India was identified: Lee et al performed a budget impact analysis. Scaling up Xpert in India increased TB related healthcare expenditures by US$580 million (81% increase) over 2 years, mostly driven by increased MDR-TB treatment spending. Deploying Truenat POC increased expenditures by an additional US$100 million over Xpert (7% increase) over 2 years. cost for unit cost 13US$ in India (including equipment).	A great deal of uncertainty about longer term cost. Training requirements not included in the cost analysis. May need more training than Xpert.

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Only one study conducted in India, results depend on several important modelling assumptions including loss to follow-up prior to treatment initiation/linkage to care and Truenat sensitivity.	

Cost effectiveness

Does the cost-effectiveness of the intervention favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the interventio n ○ Favours the intervention ○ Varies ○ No included studies	Truenat was determined to be cost-effective in the Indian setting when implemented at the POC with an ICER of US$210/YLS.	Did not include cost of transmission which may lead for cost to come down. There are limits with the applicability. POC testing in different settings will have implications on cost in those settings. probably favours 8 varies 7
Equity What would be the impact on health equity?		
Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	As test can be performed at decentralized levels of the health care system, it will likely increase health care equity.	
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Data from implementation trial show assay is generally acceptable	New, more complicated test
Feasibility Is the intervention feasible to implement?		

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Data from implementation trial show assay is generally feasible	

Summary of judgements

	JUDGEMENT						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of test accuracy	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the management effects	Very low	Low	Moderate	**High**			No included studies
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the interventio n	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**

	Judgement						
Cost-effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the interventio n	Varies	No included studies
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In adults with signs and symptoms of pulmonary TB, the GDG **suggest using** Truenat MTB as initial diagnostic test for TB (**conditional recommendation, moderate certainty of evidence for test accuracy**).

We added MTP PLUS in adults with signs and symptoms of pulmonary TB, the GDG **suggest using** Truenat MTB or MTB PLUS as initial diagnostic test for TB (**conditional recommendation, moderate certainty of evidence for test accuracy**).

Remark: The recommendation includes those patients who are smear negative. There is uncertainty about PLHIV and the various subgroups of PLHIV. The sensitivity in patients in smear negatives is lower than for all adults but the TA results are still acceptable for extrapolation from smear negatives to PLHIV. This indirect data (no data in PLHIV for this version of Truenat), allowed the GDG to extrapolate to PLHIV to have this recommendation apply. However, the certainty in the test accuracy would be lowered for additional indirectness.

Children: There is no data about how the test would perform in different speciments and not enough indirect evidence in the view of the panel to extrapolate, but extrapolation to children for sputum samples was accepted.

The GDG expects the test to be less sensitive in children.

The GDG emphasized that this is a two step test.

Justification

moderate certainty of the evidence

uncertain cost effectiveness

new equipment and training increase uncertainty about acceptability and feasibility

no patient important outcomes

Implementation considerations

first TB testing than RR

enhanced quality control including contamination testing and ? lack of external quality control

Volume of waste management is not known.

Biosafety has not been assessed - will look at manufacture instructions

Monitoring and evaluation

data collection and quality assurance automatically collected

Research priorities

more cost effectiveness data from different settings

pragmatic studies, randomized trials, even accuracy studies

Studies in PLHIV

3.3 Evidence-to-decision tables: Moderate complexity automated NAATs

PICO 1. Should Moderate complexity automated NAATs on respiratory specimens be used to diagnose PTB in adults (> 15 years) with signs and symptoms of TB, MRS?	
Population:	adults (> 15 years) with signs and symptoms of TB, MRS
Intervention:	E2E solutions on respiratory specimens

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Tuberculosis (TB) causes 10 million cases and 1.5 million deaths annually and it is estimated that 3 million cases go undiagnosed each year (WHO Global Tuberculosis Report 2020). Drug-resistant TB (DR-TB) is a major threat to global TB control. Ending the global TB epidemic will be achievable over the next 20 years only if there is intensive action by all countries which have endorsed the End TB Strategy and its ambitious targets (Implementing the end TB strategy: the essentials. WHO, 2015). Early diagnosis and prompt treatment of all persons of all ages with any form of drug-susceptible or drug-resistant TB is fundamental. WHO-endorsed rapid TB diagnostics and drug susceptibility testing (DST) should be available to all persons with signs and symptoms of TB to meet the targets of the End TB Strategy.	
Test accuracy How accurate is the test?		
Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ○ Accurate ● Very accurate ○ Varies ○ Don't know	**Test accuracy** E2E solutions on respiratory specimens Sensitivity: 0.93 (95% CI: 0.91 to 0.95) Specificity: 0.98 (95% CI: 0.96 to 0.99)	
Desirable Effects How substantial are the desirable anticipated effects?		
Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	True positive result means correct TB diagnosis. True negative result will allow rapid exclusion of the TB diagnosis, decrease of stigma, better opportunities for diagnosis other diseases and likely better patient outcomes.	The assumption is that in many settings phenotypic testing may not be available or testing may not be done.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 30%		
True positives patients with PTB	23 (23 to 24)	93 (91 to 95)	279 (273 to 284)	4767 (29)	⨁⨁⨁◯ MODERATE[a,b]
False negatives patients incorrectly classified as not having PTB	2 (1 to 2)	7 (5 to 9)	21 (16 to 27)		
True negatives patients without PTB	953 (932 to 963)	879 (860 to 889)	684 (669 to 692)	9085 (29)	⨁⨁⨁⨁ HIGH[b]
False positives patients incorrectly classified as having PTB	22 (12 to 43)	21 (11 to 40)	16 (8 to 31)		

a. Of the total 29 studies, 16 (55%) had high or unclear risk of bias as they either did prior testing before including specimens in the study or used convenience sampling or the method of participant selection was not reported. We downgraded one level for risk of bias.
b. Median TB prevalence in these studies was 31% and the number of specimens for TB positive and TB negative are large, so we decided to not downgrade for indirectness.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	**False positive** result means unnecessary treatment, stigma, financial losses. **False negative** result would mean missed diagnosis, worse health outcomes, dissemination of TB infection.	a. Of the total 29 studies, 16 (55%) had high or unclear risk of bias as they either did prior testing before including specimens in the study or used convenience sampling or the method of participant selection was not reported. We downgraded one level for risk of bias.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 30%		
True positives	23 (23 to 24)	93 (91 to 95)	279 (273 to 284)	4767 (29)	

	patients with PTB					⨁⨁⨁◯ MODERATE[a,b]
	False negatives patients incorrectly classified as not having PTB	2 (1 to 2)	7 (5 to 9)	21 (16 to 27)		
	True negatives patients without PTB	953 (932 to 963)	879 (860 to 889)	684 (669 to 692)	9085 (29)	⨁⨁⨁⨁ HIGH[b]
	False positives patients incorrectly classified as having PTB	22 (12 to 43)	21 (11 to 40)	16 (8 to 31)		

b. Median TB prevalence in these studies was 31% and the number of specimens for TB positive and TB negative are large, so we decided to not downgrade for indirectness.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Overall certainty: MODERATE	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No direct evidence was considered here. Although a diagnostic study may not capture adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.	No direct evidence was reported on direct benefits or harms

Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	There are no current observational or randomized controlled studies on patient-important outcomes of using the test.	Will vary from very low to high, depending on the type of regimen that is chosen. This will require an explanation and links to recommendations
Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies	The evidence suggests that test results would be used up by clinicians and decisions will be based on the test results for both TB detection and resistance detection.	
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies		Moderate certainty in the test accuracy
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing what is wrong with me), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence)	**E2E platforms** address several preferences/values of clinicians and laboratory staff; it is faster than culture DST (like LPA or cartridge-based tests); has the advantage of being automated (unlike LPA); and gives additional clinically-relevant DR information e.g. high vs.

variability ○ No important uncertainty or variability		low resistance (unlike the current GeneXpert MTB/RIF cartridge). (Interview study)
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the interventio n ○ Varies ○ Don't know		The reference standard is culture. Clinical benefit has not been evaluated here. Clinical benefit would be superior in terms of speed of treatment. For TB diagnosis
Resources required How large are the resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	Unit test costs for BD MAX and Hain ranged from $18.52 ($13.79 - $40.70) and $15.37 ($9.61 – $37.40), with cheaper per test kit costs reported for Hain and higher operational costs associated with lab processing time. Equipment costs were strong drivers of cost variation and will vary across lab networks and operations, if equipment can be optimally placed or multiplexed to ensure high testing volume, per test cost can be minimized.	
Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Available per-test cost data while unpublished, did include overhead, equipment, building, staff and consumable costs however complete quality assessment of the study was not possible. Test cost will vary according to testing volume and laboratory operations. There is limited evidence to assess the important variability across sites, countries and implementation approaches.	

Cost effectiveness Does the cost-effectiveness of the intervention favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies ● No included studies	No studies were identified that assessed cost-effectiveness analyses for any of the E2E solutions and extrapolation was not appropriate given differences in standard of care, different care cascades and associated costs, operational conditions, testing volume and diagnostic accuracy. Implementation considerations such as test placement, lab network, and ability of program to initiate treatment quickly will all likely impact unit test cost and cost-effectiveness. Economic modelling is needed across various settings to understand the range cost-effectiveness profiles of E2E solutions and how they likely vary under different operational criteria.	
Equity What would be the impact on health equity?		
Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ○ Probably increased ○ Increased ● Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence). Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence). Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study). New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies). (interview study) The identified challenges with **E2E** utilization and accumulated delays risk compromize the added value as identified by the users, ultimately leading to	Centralization and accessibility may impact on equity. IN places where culture is not implemented. Centralized tests may provide greater access. Transport systems will impact on this equity. Very infrastructure dependent. Some members of the panel felt, therefore that equity is reduced. Differenes between the platforms supports the judgment of varies.

	underutilization and hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence)	
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Patients can be reluctant to test for TB/MDR-TB because of stigma related to MDR-TB or related to having interrupted treatment in the past, because of fears of side effects, the failure to recognize symptoms, the inability to produce sputum and the cost, distance and travel concerns related to (repeat) clinic visits. (QES: high confidence) Health workers can be reluctant to test for TB or MDR-TB because of TB associated stigma and consequences for their patients, fears of acquiring TB, fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified, fear of side effects of drugs in children, and community awareness of disease manifestations in children. (QES: high confidence) **E2E Acceptability:** The automation of E2E, which recognizes the high workload of laboratory staff, lends to the acceptability of these technologies. The physical size of the platform and how it fits into the laboratory space/workflow affect this acceptability (smaller footprint may be more acceptable). The number of samples run on the system is acceptable, if the platform is placed within a laboratory that receives a sufficient sample load to run the system. Specific (infrastructure requirements, sample quality and volumes, communication between laboratory and clinicians) and general feasibility challenges (as identified in interview study and QES respectively), and accumulated delays risk undoing the added value/benefits as identified by the users (avoiding delays, drug resistant information). (combination QES and interview study)	**E2E platforms** address several preferences/values of clinicians and laboratory staff; it is faster than culture DST (like LPA or cartridge-based tests); has the advantage of being automated (unlike LPA); and gives additional clinically-relevant DR information e.g. high vs. low resistance (unlike the current GeneXpert MTB/RIF cartridge). (Interview study) Acceptability is linked to access and some members therefore felt that acceptability may vary.
Feasibility Is the intervention feasible to implement?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	**Feasibility is challenged** by accumulation of diagnostic delays and/or underutilization at every step due to mainly health system factors: non-adherence to testing algorithms, testing for (MDR)-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor/delayed sample transport and sample quality, and result communication, delays in scheduling follow up visits and recalling patients, inconsistent result recording; lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners, and sputum containers; dusty environment, and delayed or absent local repair option); inefficient/unclear work- and patient flows (for instance inefficient organizational processes, poor links between providers, unclear follow up mechanisms or where patients need to go); and lack of data-driven and inclusive national implementation processes. These challenges lead to delays and/or underutilization. (QES: high confidence) The feasibility of E2E platforms is challenged by how/if the platform fits into the physical space of the laboratory (considering bench size and weight of the platform). A poorly functioning sample network challenges feasibility of implementing E2E and laboratory technicians voiced concerns over the quality of	An **efficient sample transportation system, with sustainable funding mechanisms** is crucial for feasibility, especially if an algorithm requires multiple samples at different times, from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, the sample risks being contaminated and yielding inconclusive results on molecular diagnostics. Here, participants cited good personnel skill, standardized operating procedures, and

	samples. Additional feasibility considerations for this method include ensuring clinicians and laboratory staff have time to communicate effectively regarding diagnostic results if the platform is centralized, while also ensuring the laboratory where it is placed is central enough to receive adequate numbers of samples to make the machine worth running. (interview study).	significant laboratory infrastructure as essential in reducing sample contamination in their laboratory. (interview study) Implementation of new diagnostics must be accompanied with **training for clinicians**, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of Xpert MTB/RIF this has been a challenge (QES: high confidence and interview study). Furthermore, introduction of new diagnostics must be accompanied by **guidelines and algorithms, which support clinicians and laboratories in communicating with each other**, such that they can discuss discordant results, and interpret laboratory results in the context of drug availability, patient history, and patient progress on a current drug regimen.(Interview study) **E2E platforms** address several preferences/values of clinicians and laboratory staff; it is faster than culture DST (like LPA or cartridge-based tests); has the advantage of being automated (unlike LPA); and gives additional clinically-relevant DR information e.g. high vs. low resistance (unlike the current GeneXpert MTB/RIF cartridge). (Interview study)

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	Accurate	**Very accurate**		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know

	Judgement						
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of test accuracy	Very low	Low	Moderate	High			**No included studies**
Certainty of evidence of the management effects	Very low	Low	Moderate	High			**No included studies**
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			No included studies
Certainty effects	Very low	Low	**Moderate**	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the interventio n	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost-effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	**Varies**	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In people with signs and symptoms of pulmonary TB, moderate complexity automated NAATs for detection of pulmonary TB may be used on respiratory samples rather than culture (Conditional recommendation; moderate certainty of evidence for diagnostic accuracy)

Remark: limited data EPTB

Subgroup considerations

Children: all studies that reported age included only adults

PLHIV: recommendations apply to PLHIV

Implementation considerations

- requires well established laboratories

- laboratory specifications (machine size) requires appropriate infrastructure

- specimen transport standardization

- not complex tests, means that high level technical staff may not be required

- can be used with other tests, some laboratories may already have existint systems

- will depend on the number of specimens being tested - if few, tests will be come relatively more expensive.

- maintenance and support for equipment

- in resource limited settings, implementation shoudl be balanced with simpler NAATs that are less centralized

Research priorities

Implementation and operational research

Some of this equipment is already used for TB/HIV equipment and comparison of advantages against other technologies will be informative for future re

commendations

how can the use of these tests be optimized in the overal landscape of TB testing and care, in particular setting specific.

COVID context

data for children

in the context of pathways and algoritms

Throughput dependent - strategy

collaboration between programs

PICO 2. Should Moderate complexity automated NAATS on respiratory specimens be used to diagnose rifampicin resistance in adults (> 15 years) with microbiologically confirmed PTB, MRS?

Population:	adults (> 15 years) with microbiologically confirmed PTB, MRS

Intervention:	Moderate complexity automated NAATs on respiratory specimens

Assessment

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Drug-resistant TB continues to be a public health threat. Worldwide in 2019, close to half a million people developed rifampicin-resistant TB (RR-TB), of which 78% had multidrug-resistant TB (MDR-TB) (WHO Global TB report, 2020)	

Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ○ Accurate ● Very accurate ○ Varies ○ Don't know	**Test accuracy** Moderate complexity automated NAATs on respiratory specimens Sensitivity: 0.97 (95% CI: 0.93 to 0.98) Specificity: 0.99 (95% CI: 0.97 to 0.99)	

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	True positive result means correct detection of Rifampicin resistance. True negative result will allow rapid exclusion of the rifampicin resistance, decrease of stigma, better opportunities for diagnosis other diseases and likely better patient outcomes.	The assumption is that in many settings phenotypic testing may not be available or testing may not be done.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 15%		
True positives patients with	19 (19 to 20)	97 (93 to 98)	145 (140 to 148)	702 (18)	⨁⨁⨁◯ MODERATE[a,b]

a. There were 8 (44%) out of 18 studies that had high or unclear risk of bias as the participant selection was not reported or there was prior testing done for the specimens included in the study. We downgraded one level for risk of bias.

b. The median prevalence of

	rifampicin resistance					rifampicin resistance in these studies was 15%, which is representative of drug resistance in most countries for pulmonary TB. We did not downgrade for indirectness
	False negatives patients incorrectly classified as not having rifampicin resistance	1 (0 to 1)	3 (2 to 7)	5 (2 to 10)		
	True negatives patients without rifampicin resistance	969 (956 to 975)	890 (878 to 896)	841 (829 to 846)	2172 (18)	⨁⨁⨁⨁ HIGH
	False positives patients incorrectly classified as having rifampicin resistance	11 (5 to 24)	10 (4 to 22)	9 (4 to 21)		

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ○ Small ● Trivial ○ Varies ○ Don't know	**False positive** result means unnecessary treatment, stigma, financial losses. **False negative** result would mean missed diagnosis, worse health outcomes, dissemination of TB infection.	a. There were 8 (44%) out of 18 studies that had high or unclear risk of bias as the participant selection was not reported or there was prior testing done for the specimens included in the study. We downgraded one level for risk of bias. b. The median prevalence of rifampicin resistance in these studies was 15%, which is representative of drug resistance in most countries for pulmonary TB. We

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 15%		
True positives patients with rifampicin resistance	19 (19 to 20)	97 (93 to 98)	145 (140 to 148)	702 (18)	⨁⨁⨁◯ MODERATE[a,b]

	False negatives patients incorrectly classified as not having rifampicin resistance	1 (0 to 1)	3 (2 to 7)	5 (2 to 10)		did not downgrade for indirectness
	True negatives patients without rifampicin resistance	969 (956 to 975)	890 (878 to 896)	841 (829 to 846)	2172 (18)	⨁⨁⨁⨁ HIGH
	False positives patients incorrectly classified as having rifampicin resistance	11 (5 to 24)	10 (4 to 22)	9 (4 to 21)		

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Overall certainty: MODERATE	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No direct evidence was considered here. Although a diagnostic study may not capture adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.	

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies	There are no current observational or randomized controlled studies on patient-important outcomes of using the test.	- link relvant recommendations varies from very low to high (e.g. drug sensi TB)
Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	The evidence suggests that test results would be used up by clinicians and decisions will be based on the test results for both TB detection and resistance detection.	
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies	This is the summary of the preceding judgements 5-8	moderate for test accuracy
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing what is wrong with me), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence) **E2E platforms** address several preferences/values of clinicians and laboratory staff; it is faster than culture DST (like LPA or cartridge-based tests); has the advantage of being automated (unlike LPA); and gives additional clinically-relevant DR information e.g. high vs. low resistance (unlike the current GeneXpert MTB/RIF cartridge). (Interview study)	

variability		
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ● Favors the intervention ○ Varies ○ Don't know		The reference standard is PHENOTYPIC (the comparator) Clinical benefit has not been evaluated here. Clinical benefit would be superior in terms of speed of treatment.
Resources required How large are the resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	Unit test costs for BD MAX and Hain ranged from $18.52 ($13.79 - $40.70) and $15.37 ($9.61 – $37.40), with cheaper per test kit costs reported for Hain and higher operational costs associated with lab processing time. Equipment costs were strong drivers of cost variation and will vary across lab networks and operations, if equipment can be optimally placed or multiplexed to ensure high testing volume, per test cost can be minimized.	
Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Available per-test cost data while unpublished, did include overhead, equipment, building, staff and consumable costs however complete quality assessment of the study was not possible. Test cost will vary according to testing volume and laboratory operations. There is limited evidence to assess the important variability across sites, countries and implementation approaches.	
Cost effectiveness Does the cost-effectiveness of the intervention favor the intervention or the comparison?		

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies ● No included	No studies were identified that assessed cost-effectiveness analyses for any of the E2E solutions and extrapolation was not appropriate given differences in standard of care, different care cascades and associated costs, operational conditions, testing volume and diagnostic accuracy. Implementation considerations such as test placement, lab network, and ability of program to initiate treatment quickly will all likely impact unit test cost and cost-effectiveness. Economic modelling is needed across various settings to understand the range cost-effectiveness profiles of E2E solutions and how they likely vary under different operational criteria.	

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence). Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence). Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study). New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies). (interview study) The identified challenges with **E2E** utilization and accumulated delays risk compromize the added value as identified by the users, ultimately leading to underutilization and hamper access to prompt and accurate testing and treatment particularly for vulnerable groups**.** (QES: High confidence)	

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Patients can be reluctant to test for TB/MDR-TB because of stigma related to MDR-TB or related to having interrupted treatment in the past, because of fears of side effects, the failure to recognize symptoms, the inability to produce sputum and the cost, distance and travel concerns related to (repeat) clinic visits. (QES: high confidence) Health workers can be reluctant to test for TB or MDR-TB because of TB associated stigma and consequences for their patients, fears of acquiring TB, fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified, fear of side effects of drugs in children, and community awareness of disease manifestations in children. (QES: high confidence) **E2E Acceptability:** The automation of E2E, which recognizes the high workload of laboratory staff, lends to the acceptability of these technologies. The physical size of the platform and how it fits into the laboratory space/workflow affect this acceptability (smaller footprint may be more acceptable). The number of samples run on the system is acceptable, if the platform is placed within a laboratory that receives a sufficient sample load to run the system. Specific (infrastructure requirements, sample quality and volumes, communication between laboratory and clinicians) and general feasibility challenges (as identified in interview study and QES respectively), and accumulated delays risk undoing the added value/benefits as identified by the users (avoiding delays, drug resistant information). (combination QES and interview study)	

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The feasibility of **E2E platforms is challenged by how/if the platform fits into the physical space of the laboratory** (considering bench size and weight of the platform). **A poorly functioning sample network challenges feasibility of implementing E2E** and laboratory technicians voiced concerns over the quality of samples. Additional feasibility considerations for this method include ensuring clinicians and laboratory staff have time to communicate effectively regarding diagnostic results if the platform is centralized, while also ensuring the laboratory where it is placed is central enough to receive adequate numbers of samples to make the machine worth running. (interview study)	**Feasibility is challenged** by accumulation of diagnostic delays and/or underutilization at every step due to mainly health system factors: non-adherence to testing algorithms, testing for (MDR)-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor/delayed sample transport and sample quality, and result communication, delays in scheduling follow up visits and recalling patients, inconsistent result recording; lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners, and sputum containers; dusty environment, and delayed or absent local

		repair option); inefficient/unclear work- and patient flows (for instance inefficient organizational processes, poor links between providers, unclear follow up mechanisms or where patients need to go); and lack of data-driven and inclusive national implementation processes. These challenges lead to delays and/or underutilization. (QES: high confidence) An **efficient sample transportation system, with sustainable funding mechanisms** is crucial for feasibility, especially if an algorithm requires multiple samples at different times, from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, the sample risks being contaminated and yielding inconclusive results on molecular diagnostics. Here, participants cited good personnel skill, standardized operating procedures, and significant laboratory infrastructure as essential in reducing sample contamination in their laboratory. (interview study) Implementation of new diagnostics must be accompanied with **training for clinicians**, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of Xpert MTB/RIF this has been a challenge (QES: high confidence and interview study). Furthermore, introduction of new diagnostics must be accompanied by **guidelines and algorithms, which support clinicians and laboratories in communicating with each other**, such that they can discuss discordant results, and interpret laboratory results in the context of drug availability, patient history, and patient progress on a current drug regimen.(Interview study)

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	Accurate	**Very accurate**		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	Small	**Trivial**		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of test accuracy	Very low	Low	Moderate	High			**No included studies**
Certainty of evidence of the management effects	Very low	Low	Moderate	High			No included studies
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	**Favors the intervention**	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost-effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In adults with signs and symptoms of pulmonary TB, moderate complexity automated NAATs for detection of rifampicin resistance may be used on respiratory samples (rather than culture based phenotypic DST) (Conditional recommendation; moderate certainty of evidence for diagnostic accuracy)

Justification

Despite large benefits, trivial harms (and high certainty evidence in managment for some populations) the panel decided on a conditional recommendation because of uncertainty about cost, feasibility and acceptability

Subgroup considerations

Children and PLHIV - same as for INH

Implementation considerations

Same as of INH

Monitoring and evaluation

Same as of INH but low and high INH resistance not applicable here.

Research priorities

Same as of INH

Position in overall diagnostic flow

PICO 3. Moderate complexity automated NAATs on respiratory specimens be used to diagnose isoniazid resistance in adults (> 15 years) with microbiologically confirmed PTB, MRS?

Population:	adults (> 15 years) with microbiologically confirmed PTB, MRS
Intervention:	Moderate complexity automated NAATs on respiratory specimens

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Isoniazid-resistant TB is present in 8% of TB cases worldwide and reduces treatment success in patients treated with the standard 6-month first-line regimen (WHO treatment guidelines for isoniazid-resistant tuberculosis, 2018). Further, as countries continue to be faced with a significant burden of TB disease, there is an increased need to rapidly test higher volumes (or numbers) of specimens. Using new laboratory technologies that allow for testing of different conditions using disease-specific tests on the same platform can provide significant system efficiencies and cost savings, increase patient access, and ultimately improve quality of care (Information note. Global TB Programme and Department of HIV/AIDS). Emerging data suggest that, in some settings, **RR** testing has suboptimal specificity for MDR-TB (WHO Global tuberculosis report 2020). This means that testing for resistance to **isoniazid** is increasingly important. For instance, a study in DRC found one in five **RR** patients to be **isoniazid** susceptible (Bismwa 2020), and the most recent South African National Survey of Drug Resistance found hotspots of rifampicin mono-resistance, where the prevalence ratio of such cases exceeded that of MDR-TB by as much as 30% (NICD 2016). Conversely, **isoniazid** resistance in the presence of rifampicin susceptibility (isoniazid mono-resistance) is also increasingly recognised as another emerging challenge in managing tuberculosis as it is an important enabler of MDR-TB (Sulis 2020).	
Test accuracy How accurate is the test?		
Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	**Test accuracy** Moderate complexity automated NAATs on respiratory specimens Sensitivity: 0.86 (95% CI: 0.83 to 0.89) Specificity: 0.99 (95% CI: 0.98 to 1.00)	
Desirable Effects How substantial are the desirable anticipated effects?		
Judgement	Research evidence	Additional considerations

○ Trivial
○ Small
○ Moderate
● Large
○ Varies
○ Don't know

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 15%		
True positives patients with isoniazid resistance	17 (17 to 18)	86 (83 to 89)	130 (124 to 134)	854 (18)	⨁⨁⨁◯ MODERATE[a,b,c]
False negatives patients incorrectly classified as not having isoniazid resistance	3 (2 to 3)	14 (11 to 17)	20 (16 to 26)		
True negatives patients without isoniazid resistance	972 (961 to 977)	893 (883 to 897)	843 (834 to 847)	1904 (18)	⨁⨁⨁⨁ HIGH[b]
False positives patients incorrectly classified as having isoniazid resistance	8 (3 to 19)	7 (3 to 17)	7 (3 to 16)		

True positive result means rapid extended drug resistance profiling allows for early initiation of optimized therapy and likely better patient outcomes. Amplification of drug resistance would be less likely. Information on inhA promotor mutations could also guide high dose isoniazid therapy.

True negative result will allow rapid exclusion of the TB diagnosis, decrease of stigma, better opportunities for diagnosis other diseases and likely better patient outcomes.

a. There were 8 (44%) out of 18 studies that had high or unclear risk of bias as the participant selection was not reported or there was prior testing done for the specimens included in the study. We downgraded one level for risk of bias.
b. The median prevalence in these studies was 19.7%. With high number of specimens being evaluated in these studies, we did not downgrade for indirectness.
c. Sensitivity for INH resistance ranges from 58% to 100%. There was one study with low sensitivity, however, overlapping confidence intervals were seen. We did not downgrade for inconsistency.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations

○ Large
○ Moderate
● Small
○ Trivial
○ Varies
○ Don't know

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 15%		
True positives patients with isoniazid resistance	17 (17 to 18)	86 (83 to 89)	130 (124 to 134)	854 (18)	⨁⨁⨁◯ MODERATE[a,b,c]
False negatives patients incorrectly classified as not having isoniazid resistance	3 (2 to 3)	14 (11 to 17)	20 (16 to 26)		
True negatives patients without isoniazid resistance	972 (961 to 977)	893 (883 to 897)	843 (834 to 847)	1904 (18)	⨁⨁⨁⨁ HIGH[b]
False positives patients incorrectly classified as having isoniazid resistance	8 (3 to 19)	7 (3 to 17)	7 (3 to 16)		

False positive result means unnecessary treatment, stigma, financial losses.

False negative result would mean missed diagnosis, worse health outcomes, dissemination of TB infection.

a. There were 8 (44%) out of 18 studies that had high or unclear risk of bias as the participant selection was not reported or there was prior testing done for the specimens included in the study. We downgraded one level for risk of bias.
b. The median prevalence in these studies was 19.7%. With high number of specimens being evaluated in these studies, we did not downgrade for indirectness.
c. Sensitivity for INH resistance ranges from 58% to 100%. There was one study with low sensitivity, however, overlapping confidence intervals were seen. We did not downgrade for inconsistency.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Overall certainty: MODERATE	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No direct evidence was considered here. Although a diagnostic study may not capture adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.	no direct evidence was reported

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	There are no current observational or randomized controlled studies on patient-important outcomes of using the test.	very low certainty - link to the treatment recommendations

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	The evidence suggests that test results would be used up by clinicians and decisions will be based on the test results for both TB detection and resistance detection.	

Certainty of effects

What is the overall certainty of the evidence of effects of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Summary of the points 5-9	moderate certainty in test accuracy

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing what is wrong with me), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence) **E2E platforms** address several preferences/values of clinicians and laboratory staff; it is faster than culture DST (like LPA or cartridge-based tests); has the advantage of being automated (unlike LPA); and gives additional clinically-relevant DR information e.g. high vs. low resistance (unlike the current GeneXpert MTB/RIF cartridge). (Interview study)	

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the interventio n ○ Varies		The reference standard is phenotypic DST (the comparator) Clinical benefit has not been evaluated here. Clinical benefit would be superior in terms of speed of treatment.

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	Unit test costs for BD MAX and Hain ranged from $18.52 ($13.79 - $40.70) and $15.37 ($9.61 – $37.40), with cheaper per test kit costs reported for Hain and higher operational costs associated with lab processing time. Equipment costs were strong drivers of cost variation and will vary across lab networks and operations, if equipment can be optimally placed or multiplexed to ensure high testing volume, per test cost can be minimized.	

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Available per-test cost data while unpublished, did include overhead, equipment, building, staff and consumable costs however complete quality assessment of the study was not possible. Test cost will vary according to testing volume and laboratory operations. There is limited evidence to assess the important variability across sites, countries and implementation approaches.	
Cost effectiveness Does the cost-effectiveness of the intervention favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies ● No included	No studies were identified that assessed cost-effectiveness analyses for any of the E2E solutions and extrapolation was not appropriate given differences in standard of care, different care cascades and associated costs, operational conditions, testing volume and diagnostic accuracy. Implementation considerations such as test placement, lab network, and ability of program to initiate treatment quickly will all likely impact unit test cost and cost-effectiveness. Economic modelling is needed across various settings to understand the range cost-effectiveness profiles of E2E solutions and how they likely vary under different operational criteria.	
Equity What would be the impact on health equity?		
Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence). Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence). Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study).	Some variability across countries, current limited culture/pDST testing access and lmited INH DST before these test were availables

	New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies). (interview study) The identified challenges with **E2E** utilization and accumulated delays risk compromize the added value as identified by the users, ultimately leading to underutilization and hamper access to prompt and accurate testing and treatment particularly for vulnerable groups**.** (QES: High confidence)	

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Patients can be reluctant to test for TB/MDR-TB because of stigma related to MDR-TB or related to having interrupted treatment in the past, because of fears of side effects, the failure to recognize symptoms, the inability to produce sputum and the cost, distance and travel concerns related to (repeat) clinic visits. (QES: high confidence) Health workers can be reluctant to test for TB or MDR-TB because of TB associated stigma and consequences for their patients, fears of acquiring TB, fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified, fear of side effects of drugs in children, and community awareness of disease manifestations in children. (QES: high confidence) **E2E Acceptability:** The automation of E2E, which recognizes the high workload of laboratory staff, lends to the acceptability of these technologies. The physical size of the platform and how it fits into the laboratory space/workflow affect this acceptability (smaller footprint may be more acceptable). The number of samples run on the system is acceptable, if the platform is placed within a laboratory that receives a sufficient sample load to run the system. Specific (infrastructure requirements, sample quality and volumes, communication between laboratory and clinicians) and general feasibility challenges (as identified in interview study and QES respectively), and accumulated delays risk undoing the added value/benefits as identified by the users (avoiding delays, drug resistant information). (combination QES and interview study)	

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes	**Feasibility is challenged** by accumulation of diagnostic delays and/or underutilization at every step due to mainly health system factors: non-adherence to testing algorithms, testing for (MDR)-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages,	Similar considerations to TB detections

○ Varies ○ Don't know	poor/delayed sample transport and sample quality, and result communication, delays in scheduling follow up visits and recalling patients, inconsistent result recording; lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners, and sputum containers; dusty environment, and delayed or absent local repair option); inefficient/unclear work- and patient flows (for instance inefficient organizational processes, poor links between providers, unclear follow up mechanisms or where patients need to go); and lack of data-driven and inclusive national implementation processes. These challenges lead to delays and/or underutilization. (QES: high confidence) The feasibility of **E2E platforms is challenged by how/if the platform fits into the physical space of the laboratory** (considering bench size and weight of the platform). **A poorly functioning sample network challenges feasibility of implementing E2E** and laboratory technicians voiced concerns over the quality of samples. Additional feasibility considerations for this method include ensuring clinicians and laboratory staff have time to communicate effectively regarding diagnostic results if the platform is centralized, while also ensuring the laboratory where it is placed is central enough to receive adequate numbers of samples to make the machine worth running. (interview study)	An **efficient sample transportation system, with sustainable funding mechanisms** is crucial for feasibility, especially if an algorithm requires multiple samples at different times, from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, the sample risks being contaminated and yielding inconclusive results on molecular diagnostics. Here, participants cited good personnel skill, standardized operating procedures, and significant laboratory infrastructure as essential in reducing sample contamination in their laboratory. (interview study) Implementation of new diagnostics must be accompanied with **training for clinicians**, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of Xpert MTB/RIF this has been a challenge (QES: high confidence and interview study). Furthermore, introduction of new diagnostics must be accompanied by **guidelines and algorithms, which support clinicians and laboratories in communicating with each other**, such that they can discuss discordant results, and interpret laboratory results in the context of drug availability, patient history, and patient progress on a current drug regimen.(Interview study)

Summary of judgements

	JUDGEMENT						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know

	Judgement						
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of test accuracy	Very low	Low	Moderate	High			**No included studies**
Certainty of evidence of the management effects	Very low	Low	Moderate	High			**No included studies**
Certainty of evidence of the test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty effects	Very low	Low	**Moderate**	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the interventio n	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost-effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either	**Conditional recommendation for the intervention**	Strong recommendation for the intervention

○	○	the intervention or the comparison ○	●	○

Conclusions

Recommendation

In adults with signs and symptoms of pulmonary TB, moderate complexity automated NAATs for detection isoniazid resistance may be used on respiratory sample (rather than culture based phenotypic DST) (Conditional recommendation; moderate certainty of evidence for diagnostic accuracy)

Subgroup considerations

children: Once TB is detected, detection of isoniazide resistance can be extrapolated to (nature of paucibacillary disease in children should be kept in mind and resistance results may not available even if detected in the first place) actionable results may differ- (applies to all diagnostic molecular assays)

PLHIV: extrapolation fine

Implementation considerations

Training on how to interpret results

otherwise same as for detection

Monitoring and evaluation

Monitoring of relapses and appropriate treatment based on test results

Monitoring of indeterminate result rates

Research priorities

same as for detection

data for children

research the impact of this testing on low and high level INH

3.4 Evidence-to-decision tables: loop-mediated isothermal amplification for the detection of *M. tuberculosis* (TB-LAMP)

PICO 1. Diagnostic accuracy of TB-LAMP vs. smear microscopy to diagnose pulmonary tuberculosis in all adults with presumptive pulmonary TB

Assessment

	Judgement	Research evidence	Additional considerations
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Currently, sputum smear microscopy is the most common diagnostic method used to detect TB as it is inexpensive, rapid and relatively simple to perform. However, the sensitivity of microscopy is poor, ranging from 30-70% depending on the setting, and is particularly poor among children and people living with HIV. It is in this context that the WHO has identified the development and evaluation of new diagnostic tools as a necessary part of further efforts in TB control.	
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	In this review using data from the 1,810 TB suspects in whom the most stringent reference standard was available (Standard 1), TB-LAMP had a pooled sensitivity 15% higher than smear microscopy (78% vs 63%, see below). While specificity was 2% lower (98% vs 100%), this may be partly explained by the identification of TB cases that were misclassified as TB negative by the gold standard (TB culture) as all of the studies were considered to have high risk of bias in the gold standard (see comment). *Test accuracy* TB-LAMP Sensitivity: 0.78 (95% CI: 0.71 to 0.83) Specificity: 0.98 (95% CI: 0.96 to 0.99) Smear Microscopy Sensitivity: 0.63 (95% CI: 0.56 to 0.69) Specificity: 1.00 (95% CI: 0.97 to 1.00)	Risk of bias in reference standard: 1 study performed only LJ culture (Madagascar RFA); 6 studies that performed MGIT had culture contamination rate <5%. LJ culture is less sensitive for TB diagnosis than liquid culture and low culture contamination rates suggest over-decontamination which can lower TB culture yield. Both of these bias the reference standard towards misclassifying TB cases as negative and lowering the calculated specificity of the index test.
Desirable effects	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small ● Moderate	The anticipated desirable effect is the diagnosis of additional TB positive cases that would be missed by smear microscopy (TP). TB-LAMP would correctly identify 7 more cases per 1000 individuals tested if the pre-test probability of TB is 5% and 22 more cases per 1000 individuals test if the pre-test probability of TB is 15% (see table below). Correct identification of additional TB cases should lead to higher cure rates, less sequelae to the individual patient, and less transmission in the community. The anticipated undesirable effect is the incorrect identification of an individual as a TB case when they are actually TB negative (FP). In this	Desirable effect: There is following reason to think this finding would be **more marked** in a real-world setting: TB-LAMP may have correctly identified some cases that were incorrectly

	○ Large ○ Varies ○ Don't know	pooled data TB-LAMP had inferior performance to smear microscopy leading to an estimate of 16 more cases misclassified per 1000 individuals tested if the pre-test probability of TB is 5% and 14 more cases per 1000 individuals test if the pre-test probability of TB is 15% (see table below). Incorrect identification of an individual as TB positive would lead to inappropriate treatment with potential medication toxicities to the individual, possible negative effect of stigmatization of the individual, negative economic effects for the individual and society.	misclassified as TB negative by the gold standard for reasons described above.
Undesirable effects	**How substantial are the undesirable anticipated effects?** ○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know		NOTE 1: Desirable effect may be even more present in a real-world setting: TB-LAMP may have correctly identified some TB cases that were incorrectly misclassified as TB negative due to the above-said limitations of the gold standard. NOTE 2: Patients with non-tuberculous mycobacteria (NTM) were excluded from this analysis, but will be present in reality, being detected as FP by smear microscopy, thus decreasing specificity, of smear microscopy. NOTE 3: Patients with past history of TB were excluded from the analysis. If they are tested by TB-LAMP, which may happen in endemic settings, this would further increase detection of FP due to TB-LAMP detecting nonviable bacteria (TB-LAMP specificity decreased).

Outcome	Study design	Test accuracy QoE	Effect per 1000 patients/year for pre-test probability of 5%		Effect per 1000 patients/year for pre-test probability of 15%		Importance
			TB-LAMP	smear microscopy	TB-LAMP	smear microscopy	
True positives	cross-sectional (cohort type accuracy study)	⨁◯◯◯ VERY LOW	39 (36 to 42)	32 (28 to 35)	117 (107 to 124)	95 (84 to 104)	IMPORTANT
TP absolute difference			7 more TP in TB-LAMP		22 more TP in TB-LAMP		
False negatives			11 (8 to 14)	18 (15 to 22)	33 (26 to 43)	55 (46 to 66)	IMPORTANT
FN absolute difference			7 fewer FN in TB-LAMP		22 fewer FN in TB-LAMP		
True negatives	cross-sectional (cohort type accuracy study)	⨁◯◯◯ VERY LOW	932 (909 to 942)	948 (923 to 950)	834 (813 to 843)	848 (826 to 850)	IMPORTANT
TN absolute difference			16 fewer TN in TB-LAMP		14 fewer TN in TB-LAMP		
False positives			18 (8 to 41)	2 (0 to 27)	16 (7 to 37)	2 (0 to 24)	IMPORTANT
FP absolute difference			16 more FP in TB-LAMP		14 more FP in TB-LAMP		

Certainty of the evidence of test accuracy	**What is the overall certainty of the evidence of test accuracy?** • Very low ○ Low ○ Moderate ○ High ○ No included studies	In this review the risk of bias was considered very serious for all 7 studies included in the analysis of TB-LAMP vs smear microscopy for the following reasons: 1) 1 study performed only LJ culture (Madagascar RFA) 2) 6 studies that performed MGIT had culture contamination rate <5% (5-10% is considered an acceptable range) 3) 2 studies (Uganda RFA, Haiti Unpublished) did not exclude all participants with prior TB (thus potentially causing false positive TB-LAMP results since DNA assays such as TB-LAMP can detect nonviable bacteria) 4) 3 studies (Madagascar RFA, Uganda RFA, Haiti unpublished) did not clearly report the number of patients enrolled. Indirectness was considered serious for all studies: No studies were conducted in peripheral microscopy centers (4 were done at reference laboratories and 3 done at hospital-/university-affiliated outpatient clinics) Inconsistency was considered very serious for test sensitivity: There was considerable heterogeneity in sensitivity estimates across individual studies Inconsistency was considered serious for test specificity: There was moderate heterogeneity in specificity estimates across individual studies Imprecision was considered not serious for all studies. Publication bias: n/a	
Certainty of the evidence of tests effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?** ○ Very low ○ Low ○ Moderate ○ High • No included studies	The test is relatively labour-intensive and presents certain burden for the health worker. The burden and adverse effects are potentially insignificant for the patient.	
Certainty	**What is the overall certainty if**	The effect of the test result on the patient management (including cure, death, treatment initiation time) was not covered in the studies included in the review.	

	the evidence of effects of the management that is guided by the test results? ○ Very low ○ Low ○ Moderate ○ High • No included studies		
Certainty of the evidence of test results/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High • No included studies	The link between test results and management decisions may be uncertain in various settings. In some occasions clinicians use empirical treatment for TB. In others capacity of health system may be insufficient to provide the patient with necessary treatment.	
Certainty of effects	**What is the overall certainty of the evidence of effects of the test?** ○ Very low ○ Low ○ Moderate ○ High • No included studies	This question is intended to summarize previous four questions on the certainty of the evidence.	
Values	**Is there important uncertainty about or variability in how much people value the**	No important uncertainty or variability in how people value the main outcomes.	

	main outcomes? ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability ○ No known undesirable outcomes		
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison ?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know	The significant increase in sensitivity and likely equivalent specificity (when the above mentioned study limitations are taken into account) indicate that TB-LAMP is a more accurate overall test than smear microscopy.	

Resources required	**How large are the resource requirements (costs)?** ● Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	Weighted average per-test cost of TB-LAMP if used as routine diagnostic test was US$14.43 for Viet Nam and US$15.92 for Malawi. First year expenditure required for implementation at medium workload peripheral laboratory for TB-LAMP in Viet Nam was US$26,917. This cost was approximately US$3000 lower in Malawi, attributable to lower operating and staff costs. Complete roll-out of the TB-LAMP assay in all of the peripheral microscopy laboratories in Malawi and Viet Nam would constitute 17% and 9% of the total NTP budget reported to the WHO for 2014 fiscal year, respectively.	
Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ● Very low ○ Low ○ Moderate ○ High ○ No included studies	The basis for the analysis is one cost, affordability, and cost-effectiveness study conducted in Viet Nam (low HIV) and Malawi (high HIV), both of which are low MDR-TB burden settings.	
Cost effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the	In cost-effectiveness analysis, both of the TB-LAMP scenarios improved case detection rates to between 74-76% and 88-90%, respectively, compared to the base-case scenario rates of 59% and 82%. The incremental cost per disability adjusted life years (DALY) for the TB-LAMP replacement for SSM strategy was between US$41 and US$131, which was higher than that of the add-on scenario at US$39 and US$123 in Malawi and Viet Nam, respectively. Both strategies were cost-effective when comparing to the World Health Organization (WHO) willingness-to-pay (WTP) threshold levels. These conclusion did not change in a range of sensitivity analysis performed.	

	comparison • Probably favours the intervention ○ Favours the intervention ○ Varies ○ No included studies		
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact • Probably increased ○ Increased ○ Varies ○ Don't know	Patient accessing lower levels of the health systems may have easier access to this test.	
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ○ Probably yes ○ Yes • Varies ○ Don't know	The test may be acceptable to be implemented for individuals at low risk of MDR-TB and/or HIV associated TB prevalence. The test will require strengthening of human resources, as it is relatively labor-intensive. May be implemented in settings where Xpert is not available. Patient are stakeholders as well.	Sustainability concerns.
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ○ Probably	Implementation of the test would require additonal funding and technical support for the training of staff and procuring the equipment. Quality assurance is not exists for TB-LAMP as of now. Additional staff is probably required. Limited data on implementation up to date. Low volume (workload) settings. Scaleability is a challenge.	Short shelf life is a limitation. In Tanzania study bigger number of FP was observed comparing to the other studies.

	yes ○ Yes ● Varies ○ Don't know	Contamination issue, related to long-term effect of the test (impact).	

Summary of judgements

	Judgement							Implications
Problem	No	Probably no	Probably yes	Yes		Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	Accurate	Very accurate		Varies	Don't know	
Desirable effects	Trivial	Small	Moderate	Large		Varies	Don't know	
Undesirable effects	Large	Moderate	Small	Trivial		Varies	Don't know	
Certainty of the evidence of test accuracy	Very low	Low	Moderate	High		No included studies		
Certainty of the evidence of test effects	Very low	Low	Moderate	High		No included studies		
Certainty of the evidence of management's effects	Very low	Low	Moderate	High		No included studies		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High		No included studies		
Certainty of effects	Very low	Low	Moderate	High		No included studies		

	Judgement							Implications
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability			No known undesirable outcomes	
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know	
Certainty of evidence of required resources	Very low	Low	Moderate	High			No included studies	
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	No included studies	
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know	
Acceptability	No	Probably no	Probably yes	Yes		Varies	Don't know	
Feasibility	No	Probably no	Probably yes	Yes		Varies	Don't know	

Conclusions

Should TB-LAMP vs. smear microscopy be used to diagnose pulmonary tuberculosis in all adults with presumptive pulmonary TB?

Type of recommendation	Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the	Conditional recommendation for the intervention	Strong recommendation for the intervention

	○	○	**intervention or the comparison** ○	●	○
RECOMMENDATION	TB-LAMP may be used rather than sputum smear microscopy for the diagnosis of pulmonary TB in adults not at risk for MDR-TB or HIV associated TB (Conditional recommendations, Very low quality of evidence).				

PICO 2. Diagnostic accuracy of TB-LAMP vs. smear microscopy to diagnose pulmonary tuberculosis in HIV positive adults with presumptive pulmonary TB

Assessment

	Judgement	Research evidence	Additional consideration
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Currently, sputum smear microscopy is the most common diagnostic method used to detect TB as it is inexpensive, rapid and relatively simple to perform. However, the sensitivity of microscopy is poor, ranging from 30-70% depending on the setting, and is particularly poor among children and people living with HIV.	
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ○ Accurate ○ Very accurate ○ Varies ● Don't know	Based on the data from 4 studies (271 patients) with HIV included in the review, TB-LAMP had a sensitivity of 64% (Reference Standard 2). However, in the patients in this review smear microscopy had an unexpectedly high sensitivity (62%) raising the question of whether the sensitivity of TB-LAMP is artificially inflated due to an unexpectedly high percentage of smear positive cases. **Very limited number of patients (271) included in the analysis.** *Test accuracy* TB-LAMP Sensitivity: 0.64 (95% CI: 0.49 to 0.76) Specificity: 0.99 (95% CI: 0.85 to 1.00) Smear microscopy Sensitivity: 0.62 (95% CI: 0.34 to 0.89) Specificity: 0.99 (95% CI: 0.95 to 1.00)	*TB-LAMP sensitivity is only marginally higher than one of smear microscopy in the HIV positive population, but CI are wide and overlapping.*
Desirable effects	**How substantial are the desirable**	Based on the limited dataset from this review, very similar numbers of true positive, false negative, false positive, and true negative results would be obtained with TB-LAMP compared to smear microscopy.	Available for analysis dataset is small. Since TB-LAMP performed

	anticipated effects? ○ Trivial ○ Small ○ Moderate ○ Large ○ Varies • Don't know	No demonstrated increase of sensitivity and specificity in TB-LAMP compared with smear microscopy	better in overall study which had a much larger sample size (see PICO question 1a), there is possibility for a better performance in the HIV positive population should a larger sample size be analyzed.
Undesirable effects	**How substantial are the undesirable anticipated effects?** ○ Large ○ Moderate ○ Small ○ Trivial ○ Varies • Don't know		Available for analysis dataset is small.

Outcome	**Study design**	**Test accuracy QoE**	**Effect per 1000 patients/year for pre-test probability of 5%**		**Effect per 1000 patients/year for pre-test probability of 15%**		**Importance**
			TB-LAMP	**smear microscopy**	**TB-LAMP**	**smear microscopy**	
True positives	cross-sectional (cohort type accuracy study)	⊕◯◯◯ VERY LOW	32 (25 to 38)	31 (17 to 45)	96 (74 to 114)	93 (51 to 134)	IMPORTANT
TP absolute difference			1 more TP in TB-LAMP		3 more TP in TB-LAMP		
False negatives			18 (12 to 25)	19 (5 to 33)	54 (36 to 76)	57 (16 to 99)	IMPORTANT
FN absolute difference			1 fewer FN in TB-LAMP		3 fewer FN in TB-LAMP		
True negatives	cross-sectional (cohort type accuracy study)	⊕◯◯◯ VERY LOW	939 (808 to 949)	941 (903 to 950)	840 (722 to 849)	842 (808 to 850)	IMPORTANT
TN absolute difference			2 fewer TN in TB-LAMP		2 fewer TN in TB-LAMP		
False positives			11 (1 to 142)	9 (0 to 47)	10 (1 to 128)	8 (0 to 42)	IMPORTANT
FP absolute difference			2 more FP in TB-LAMP		2 more FP in TB-LAMP		

Certainty of the evidence of test accuracy	**What is the overall certainty of the evidence of test accuracy?** • Very low ○ Low ○ Moderate ○ High ○ No included studies	In this review the risk of bias was considered very serious for all 4 studies included in the analysis of TB-LAMP vs smear microscopy for the following reasons: 1) 2 studies (SA and Uganda) that performed MGIT had culture contamination rate <5% (5-10% is considered an acceptable range) 2) 1 study (Uganda RFA) did not exclude all participants with prior TB (thus potentially causing false positive TB-LAMP results since DNA assays such as TB-LAMP can detect nonviable bacteria) 3) 1 study (Uganda RFA) did not clearly report the number of patients enrolled. **Indirectness** was considered serious for all studies: No studies were conducted in peripheral microscopy centers (1 was done at reference laboratories and 3 done at district hospital outpatient clinics) **Inconsistency** was considered not serious for test sensitivity: There was considerable heterogeneity in sensitivity estimates across individual studies Inconsistency was considered not serious for test specificity: There was moderate heterogeneity in specificity estimates across individual studies **Imprecision** was considered serious for all studies (small sample size and wide confidence intervals for pooled estimates). **Publication bias**: n/a	
Certainty of the evidence of test effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?** • Very low ○ Low ○ Moderate ○ High ○ No included studies	The test is relatively labour-intensive and presents certain burden for the health worker. The burden and adverse effects are potentially insignificant for the patient.	
Certainty of the evidence of management effects	**What is the overall certainty if the evidence of effects of the management that is guided by the test results?**	The effect of the test result on the patient management (including cure, death, treatment initiation time) was not covered in the studies included in the review.	

	• Very low ○ Low ○ Moderate ○ High ○ No included studies		
CERTAINTY OF THE EVIDENCE OF TEST RESULT/MANAGEMENT	**How certain is the link between test results and management decisions?** • Very low ○ Low ○ Moderate ○ High ○ No included studies	The link between test results and management decisions may be uncertain in various settings. In some occasions clinicians use empirical treatment for TB. In others capacity of health system may be insufficient to provide the patient with necessary treatment.	
CERTAINTY OF EFFECTS	**What is the overall certainty of the evidence of effects of the test?** • Very low ○ Low ○ Moderate ○ High ○ No included studies	This question is intended to summarize previous four questions on the certainty of the evidence.	
VALUES	**Is there important uncertainty about or variability in how much people value the main outcomes?**	No important uncertainty or variability in how people value the main outcomes.	

	○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability ○ No known undesirable outcomes		
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ● Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies	Although we expect that TB-LAMP has a higher sensitivity than smear microscopy in this population, this was not seen in the data from the 271 patients evaluated for this review. As a result, it is difficult to conclusively balance the desirable vs undesirable effects of the intervention although we suspect that with a larger sample size the balance would favour the intervention (TB-LAMP) over the comparison (smear microscopy).	

	○ Don't know		
Resources required	**How large are the resource requirements (costs)?** ● Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	Weighted average per-test cost of TB-LAMP if used as routine diagnostic test was US$14.43 for Viet Nam and US$15.92 for Malawi. First year expenditure required for implementation at medium workload peripheral laboratory for TB-LAMP in Viet Nam was US$26,917. This cost was approximately US$3000 lower in Malawi, attributable to lower operating and staff costs. Complete roll-out of the TB-LAMP assay in all of the peripheral microscopy laboratories in Malawi and Viet Nam would constitute 17% and 9% of the total NTP budget reported to the WHO for 2014 fiscal year, respectively. There was no cost estimations done separately for HIV+ and HIV- patients.	
Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ● Low ○ Moderate ○ High ○ No included studies	The basis for the analysis is one cost and cost-effectiveness study conducted in Viet Nam and Malawi.	

Cost effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ● Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ○ No included studies	In cost-effectiveness analysis, both of the TB-LAMP scenarios improved case detection rates to between 74-76% and 88-90%, respectively, compared to the base-case scenario rates of 59% and 82%. The incremental cost per disability adjusted life years (DALY) for the TB-LAMP replacement for SSM strategy was between US$41 and US$131, which was higher than that of the add-on scenario at US$39 and US$123 in Malawi and Viet Nam, respectively. Both strategies were cost-effective when comparing to the World Health Organization (WHO) willingness-to-pay (WTP) threshold levels. These conclusion did not change in a range of sensitivity analysis performed. The cost-effectiveness estimation was not separately done for a HIV+ comparing to HIV- patient populations.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ● Probably no impact ○ Probably increased ○ Increased ○ Varies ○ Don't know	No added benefit	

Acceptability	**Is the intervention acceptable to key stakeholders?** ● No ○ Probably no ○ Probably yes ○ Yes ○ Varies ○ Don't know	**No evidence of additional yield of TB-LAMP among HIV positive patients compared to smear microscopy.** The test may be acceptable to be implemented in settings with low MDR-TB prevalence. The test will require strengthening of human resources, as it is relatively labor-intensive. May be implemented at lower levels of the health systems.	
Feasibility	**Is the intervention feasible to implement?** ○ No ● Probably no ○ Probably yes ○ Yes ○ Varies ○ Don't know	Implementation of the test would require additonal funding and technical support for the training of staff and procuring the equipment.	

Summary of judgements

	Judgement							Implications
Problem	No	Probably no	Probably yes	Yes		Varies	Don't know	Favours TB-LAMP
Test accuracy	Very inaccurate	Inaccurate	Accurate	Very accurate		Varies	Don't know	Probably favours smear microscopy
Desirable effects	Trivial	Small	Moderate	Large		Varies	Don't know	Favours smear microscopy
Undesirable effects	Large	Moderate	Small	Trivial		Varies	Don't know	Probably favours smear microscopy

	JUDGEMENT							IMPLICATIONS
Certainty of the evidence of test accuracy	Very low	Low	Moderate	High		No included studies		Probably favours smear microscopy
Certainty of the evidence of test effects	Very low	Low	Moderate	High		No included studies		Favours smear microscopy
Certainty of the evidence of management's effects	Very low	Low	Moderate	High		No included studies		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High		No included studies		
Certainty of effects	Very low	Low	Moderate	High		No included studies		Favours smear microscopy
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability		No known undesirable outcomes		Favours neither intervention
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	Don't know	Favours neither intervention
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know	Probably favours smear microscopy
Certainty of evidence of required resources	Very low	Low	Moderate	High			No included studies	Probably favours smear microscopy
Cost effectiveness	Favours the	Probably favours the	Does not favour either the	Probably favours the	Favours the	Varies	No include	Favours neither intervention

	Judgement							Implications
	comparison	comparison	intervention or the comparison	intervention	intervention		d studies	
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know	Favours neither intervention
Acceptability	No	Probably no	Probably yes	Yes		Varies	Don't know	Favours smear microscopy
Feasibility	No	Probably no	Probably yes	Yes		Varies	Don't know	Probably favours smear microscopy

Conclusions

Should TB-LAMP vs. smear microscopy be used to diagnose pulmonary tuberculosis in HIV positive adults with presumptive pulmonary TB?

Type of recommendation	**Strong recommendation against the intervention**	**Conditional recommendation against the intervention**	**Conditional recommendation for either the intervention or the comparison**	**Conditional recommendation for the intervention**	**Strong recommendation for the intervention**
	○	○	●	○	○
Recommendation	There is no additional value of TB-LAMP over microscopy in HIV positive adults with presumptive pulmonary TB				

PICO 3. Diagnostic accuracy of TB-LAMP to diagnose pulmonary tuberculosis in adults with presumptive pulmonary TB and negative sputum smears

Assessment

	Judgement	Research evidence	Additional consideration
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes	Currently, sputum smear microscopy is the most common diagnostic method used to detect TB as it is inexpensive, rapid and relatively simple to perform. However, the sensitivity of microscopy is poor, ranging from 30-70% depending on the setting, and is particularly poor among children and people living with HIV.	

<table>
<tr><td></td><td>○ Varies
○ Don't know</td><td></td><td></td></tr>
<tr><td>Test accuracy</td><td>How accurate is the test?

○ Very inaccurate
○ Inaccurate
• Accurate
○ Very accurate

○ Varies
○ Don't know</td><td>In this review using data from the 1,349 TB suspects in whom the most stringent reference standard was available (Standard 1), across 7 studies, TB-LAMP had a pooled sensitivity 42% in those patients with presumed TB, who were negative by smear microscopy.

TPP for smear negative patients is Se 70%. Neither for Xpert this standard is met for smear negative (68%)

Test accuracy
TB-LAMP Sensitivity: 0.42 (95% CI: 0.30 to 0.55)
Specificity: 0.98 (95% CI: 0.96 to 0.99)</td><td>The low specificity may be partly explained by the identification of TB cases that were misclassified as TB negative by the gold standard (TB culture) as all of the studies were considered to have high risk of bias in the gold standard (see comment).</td></tr>
<tr><td>Desirable effects</td><td>How substantial are the desirable anticipated effects?

○ Trivial
○ Small
• Moderate
○ Large

○ Varies
○ Don't know</td><td>Under pre-test probability of 5% number of true positive only slightly exceeds number of false-positive (21 vs 19 per 1000 patients tested).

Should TB-LAMP be used to diagnose pulmonary tuberculosis in adults with presumptive pulmonary TB and negative sputum smears?</td><td>Inacurate</td></tr>
<tr><td>Undesirable effects</td><td>How substantial are the undesirable anticipated effects?

○ Large
• Moderate
○ Small
○ Trivial

○ Varies
○ Don't know</td><td></td><td></td></tr>
</table>

Test result	Number of results per 1000 patients tested (95% CI)		Number of participants (studies)	Quality of the Evidence (GRADE)
	Prevalence 5%	Prevalence 15%		
True positives (patients with pulmonary tuberculosis)	21 (15 to 28)	63 (45 to 83)	1349 (7)	⨁◯◯◯ VERY LOW [1,2,3]
False negatives (patients incorrectly classified as not having pulmonary tuberculosis)	29 (22 to 35)	87 (67 to 105)		
True negatives (patients without pulmonary tuberculosis)	931 (912 to 941)	833 (816 to 842)	1349 (7)	⨁◯◯◯ VERY LOW [1,2,4]
False positives (patients	19 (9 to 38)	17 (8 to 34)		

		incorrectly classified as having pulmonary tuberculosis)	
Certainty of the evidence of test accuracy	**What is the overall certainty of the evidence of test accuracy?** • Very low ○ Low ○ Moderate ○ High ○ No included studies	In this review the **risk of bias was considered very serious** for all 7 studies included in the analysis of TB-LAMP vs smear microscopy for the following reasons: 1) 1 study performed only LJ culture (Madagascar RFA) 2) 6 studies that performed MGIT had culture contamination rate <5% (5-10% is considered an acceptable range) 3) 2 studies (Uganda RFA, Haiti Unpublished) did not exclude all participants with prior TB (thus potentially causing false positive TB-LAMP results since DNA assays such as TB-LAMP can detect nonviable bacteria) 4) 3 studies (Madagascar RFA, Uganda RFA, Haiti unpublished) did not clearly report the number of patients enrolled. **Indirectness** was considered serious for all studies: No studies were conducted in peripheral microscopy centers (4 were done at reference laboratories and 3 done at hospital-/university-affiliated outpatient clinics) **Inconsistency** was considered very serious **for test sensitivity**: There was considerable heterogeneity in sensitivity estimates across individual studies **Inconsistency** was considered serious **for test specificity**: There was moderate heterogeneity in specificity estimates across individual studies **Imprecision** was considered not serious for all studies. *Publication bias – n/a*	
Certainty of the evidence of tests effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?** ○ Very low ○ Low ○ Moderate ○ High	The test is relatively labour-intensive and presents certain burden for the health worker. The burden and adverse effects are potentially insignificant for the patient.	

	● No included studies		
Certainty of the evidence of management effects	**What is the overall certainty if the evidence of effects of the management that is guided by the test results?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The effect of the test result on the patient management (including cure, death, treatment initiation time) was not covered in the studies included in the review.	
Certainty of the evidence of test result/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The link between test results and management decisions may be uncertain in various settings. In some occasions clinicians use empirical treatment for TB. In others capacity of health system may be insufficient to provide the patient with necessary treatment.	
Certainty of effects	**What is the overall certainty of the evidence of effects of the test?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	This question is intended to summarize previous four questions on the certainty of the evidence.	
Values	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	There is no important uncertainty about or variability in how much people value the main outcomes.	

	○ No known undesirable outcomes		
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison • Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know	Probably favours the intervention, especially at higher prevalence of TB. TB-LAMP Sensitivity: 0.42 (95% CI: 0.30 to 0.55). Specificity: 0.98 (95% CI: 0.96 to 0.99)	Major undesirable effect is a need to do 2 tests
Resources required	**How large are the resource requirements (costs)?** • Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	Weighted average per-test cost of TB-LAMP if used as routine diagnostic test was US$14.43 for Viet Nam and US$15.92 for Malawi. First year expenditure required for implementation at medium workload peripheral laboratory for TB-LAMP in Viet Nam was US$26,917. This cost was approximately US$3000 lower in Malawi, attributable to lower operating and staff costs. Complete roll-out of the TB-LAMP assay in all of the peripheral microscopy laboratories in Malawi and Viet Nam would constitute 17% and 9% of the total NTP budget reported to the WHO for 2014 fiscal year, respectively.	
Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low • Moderate ○ High ○ No included studies	The basis for the analysis is one cost, affordability, and cost-effectiveness study conducted in Viet Nam (low HIV) and Malawi (high HIV), both of which are low MDR-TB burden settings.	

Cost effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ No included studies	In cost-effectiveness analysis, both of the TB-LAMP scenarios improved case detection rates to between 74-76% and 88-90%, respectively, compared to the base-case scenario rates of 59% and 82%. The incremental cost per disability adjusted life years (DALY) for the TB-LAMP replacement for SSM strategy was between US$41 and US$131, which was higher than that of the add-on scenario at US$39 and US$123 in Malawi and Viet Nam, respectively. Both strategies were cost-effective when comparing to the World Health Organization (WHO) willingness-to-pay (WTP) threshold levels. These conclusion did not change in a range of sensitivity analysis performed. The cost-effectiveness estimation was not separately done for a HIV+ comparing to HIV- patient populations.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Patient accessing lower levels of the health systems may have easier access to this test, which potentially would improve their access to the quality diagnosis.	
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable to be implemented in settings with low MDR-TB and/or low HIV prevalence. The test will require strengthening of human resources, as it is relatively labor-intensive. May be implemented at lower levels of the health systems.	
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ○ Probably yes ○ Yes ● Varies ○ Don't know	Implementation of the test would require additonal funding and technical support for the training of staff and procuring the equipment. Quality assurance for the technology is not available as of yet.	Current demostrational studies were implemented with extensive technical and financial support of Eiken and FIND.

Summary of judgements

	Judgement							Implications
Problem	No	Probably no	Probably yes	Yes		Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	Accurate	Very accurate		Varies	Don't know	
Desirable effects	Trivial	Small	Moderate	Large		Varies	Don't know	
Undesirable effects	Large	Moderate	Small	Trivial		Varies	Don't know	
Certainty of the evidence of test accuracy	Very low	Low	Moderate	High		No included studies		
Certainty of the evidence of test effects	Very low	Low	Moderate	High		No included studies		
Certainty of the evidence of management's effects	Very low	Low	Moderate	High		No included studies		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High		No included studies		
Certainty of effects	Very low	Low	Moderate	High		No included studies		
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability		No known undesirable outcomes		

	Judgement							Implications
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know	
Certainty of evidence of required resources	Very low	Low	Moderate	High			No included studies	
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	No included studies	
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know	
Acceptability	No	Probably no	Probably yes	Yes		Varies	Don't know	
Feasibility	No	Probably no	Probably yes	Yes		Varies	Don't know	

Conclusions

Should TB-LAMP be used to diagnose pulmonary tuberculosis in adults with presumptive pulmonary TB and negative sputum smears?

Type of recommendation	Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
	○	○	○	●	○

Recommendation	TB-LAMP may be used as a follow-on test to smear microscopy in adults suspected of having pulmonary TB in adults presumed to have TB, not at risk for MDR-TB or HIV associated TB, especially when further testing of sputum smear-negative specimens is necessary (Conditional recommendation, Very low quality of evidence).

3.5 Evidence-to-decision tables: lateral flow urine lipoarabinomannan assay (LF- LAM)

PICO1: Should AlereLAM vs. no AlereLAM be used for HIV-positive adults to reduce mortality associated with advanced HIV disease, inpatient setting?	
Population:	HIV-positive adults to reduce mortality associated with advanced HIV disease, inpatient setting
Intervention:	AlereLAM
Comparison:	no AlereLAM
Main outcomes:	Mortality;
Setting:	inpatient

Assessment

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Tuberculosis (TB) remains the leading cause of hospitalization and in-hospital deaths among people living with HIV despite the increased access to antiretroviral treatment (ART) (Ford 2016). A systematic review of the prevalence of TB identified at autopsy suggests that, in resource-limited settings, TB is responsible for around 40% of all HIV-related deaths and that TB often was disseminated and undiagnosed at the time of death (Gupta 2015). Globally in 2017, only 51% of the estimated 10.0 million TB cases were notified among people living with HIV (WHO Global Report 2018). However, most death from TB is preventable if TB is detected early and effectively treated.	Non-sputum-based point-of-care TB diagnostic tests are highly desired to narrow the diagnostic gap and ensure timely treatment. Detection of mycobacterial antigen in urine is promising, as this would allow for a TB diagnosis that is non-site specific. Urine is easy to collect and store, and lacks the infection control risks associated with sputum collection. The lateral flow assay, Alere Determine™ TB LAM Ag assay 'AlereLAM', was developed as a simple point-of-care test for diagnosis of active TB in people living with HIV. AlereLAM is commercially available, does not require access to special laboratory equipment, and produces a result after 25 minutes, meeting many desired target product profile requirements.

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know Voting: Moderate - 7 (including a chair); Large - 6.		*35 saved lives per 100 admitted patients is a large effect.* *It will be further augmented by a reduction of transmission.* *From TB/HIV community perspective, every saved life is a large effect.*

Outcomes	With no AlereLAM	With AlereLAM	Difference	Relative effect (95% CI)
Mortality	230 per 1,000	**196 per 1,000** (175 to 216)	**35 fewer per 1,000** (55 fewer to 14 fewer)	**RR 0.85** (0.76 to 0.94)

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know		Patient important outcomes: missed cases overtreatment empirical treatment was approximately balanced - 6% difference in treatment which will translate into some side effects (which will be balanced against the benefits of a mortality reduction). Doing TB treatment in presumably false-positive cases also assumes doing TB preventive therapy, which partially withraws negative effect of avertreatment.

Outcomes	With no AlereLAM	With AlereLAM	Difference	Relative effect (95% CI)
Mortality	230 per 1,000	**196 per 1,000** (175 to 216)	**35 fewer per 1,000** (55 fewer to 14 fewer)	**RR 0.85** (0.76 to 0.94)

Certainty of evidence

What is the overall certainty of the evidence of effects?

Judgement	Research evidence	Additional considerations

○ Very low ○ Low ● Moderate ○ High ○ No included studies	<table><tr><th rowspan="2">Outcomes</th><th colspan="2">Anticipated absolute effects* (95% CI)</th><th rowspan="2">Relative effect (95% CI)</th><th rowspan="2">№ of participants (studies)</th><th rowspan="2">Certainty of the evidence (GRADE)</th><th rowspan="2">Comments</th></tr><tr><th>Risk with no AlereLAM</th><th>Risk with AlereLAM</th></tr><tr><td>Mortality</td><td colspan="2">Study population</td><td>RR 0.85 (0.76 to 0.94)</td><td>5102 (2 RCTs)</td><td>⨁⨁⨁◯ MODERATE[a,b]</td><td></td></tr><tr><td></td><td>230 per 1,000</td><td>196 per 1,000 (175 to 216)</td><td></td><td></td><td></td><td></td></tr></table> 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. 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.	

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability		*It is likely that no important variability exists in how much people value following important outcomes:* *Mortality.* *Cure from (TB).* *Treatment side effects (in false positives).* *Drug resistance.*

Balance of effects

Does the balance between desirable and undesirable effects favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ● Favours the intervention ○ Varies ○ Don't know	Summary of the above: Large benefits, Small harms. Probably very little variation to how people value the outcomes	

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ● Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	Systematic review by A. Zwerling: No detailed micro-costing of AlereLAM implementation was performed in inpatient settings. Using published costing data for South Africa limited to unit test cost, Boyles 2018 calculated the cost per patient for each AlereLAM containing algorithm among inpatients. Cost per patient screened by each algorithm generally increased with increasing diagnostic yield and ranged from US$10.5 for Xpert/Culture and AlereLAM/Xpert, US$12.5 for the AlereLAM/Xpert/culture, US$37.2 for the AlereLAM/Xpert SI, US$49.6 for Xpert SI/culture, and US$42 for AlereLAM/Xpert SI/culture approach. Boyles 2018 did not perform a cost-effectiveness analysis or calculate incremental cost-effectiveness ratios.	*While implemented in hospitals presents relatively low incremental cost. Implementation in a weak health system would cost more.* *Cost of avoided transmission need to be equated.* *Cost will vary depending if the LAM alone or Dx cascade are implemented?* *Cost will differ by context.*

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Systematic review by A. Zwerling: Models found cost-effectiveness of AlereLAM to be robust across a variety of sensitivity analyses, variations in key parameters and across different country settings and scenarios. Key parameters that are likely influential on cost-effectiveness include: TB prevalence, target population, and AlereLAM specificity, cost of treating TB and HIV and life expectancy post TB survival, and time horizon. However, one detailed micro-costing study published in 2018 estimates unit test costs for AlereLAM implementation several fold higher (US$23) than most current models (US$2-4). Modeling studies may contribute to the certainty of the evidence in this domain	*Modeling studies may contribute to the certainty of the evidence in this domain* *High variability* *Only one trial* *Variety of models* *No empirical studies*

Cost effectiveness

Does the cost-effectiveness of the intervention favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations

○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ● Varies ○ No included studies	Systematic review by A. Zwerling: Reddy et al assessed cost-effectiveness of AlereLAM algorithms in unselected hospitalized PLHIV. Using the modified CEPAC-I model calibrated to STAMP trial results, Reddy 2019 found Xpert + AlereLAM + urine Xpert to be cost-effective among unselected hospitalized HIV patients with ICERs of $450/YLS (Years of life saved, YLS) in Malawi and $840/YLS in South Africa compared with standard of care (Xpert alone). The modified intervention of Xpert + AlereLAM was even more cost-effective with ICERs of $420/YLS in Malawi and $810/YLS in South Africa compared with standard of care. Increased ICERs are due to inclusion of downstream costs associated with lifelong ART and HIV care.	*Only data for Africa are available* *SA results more definitive and Malawi results are less definitive*

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ○ Probably increased ● Increased ○ Varies ○ Don't know	As test can be performed at all levels of the health care system, it will likely increase health care equity.	*Universal test* *Potential to reduce inequity* *Because extrapulmonary TB are already disadvantaged it is potential to improve care for them*

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Report on user perspectives on TB LAM testing: results from qualitative research : Test is generally described as acceptable by key stakeholders.	Patients: Providers: Policy-makers/programs: Payers: Others: In children: Urine collection was more cumbersome especially in younger and sicker children as it requires both the child's and the caregiver's cooperation and may be affected by medical causes such as dehydration (Kroidl 2015).

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no	MSF study (H. Huerga) Advantages of using LAM:	In children: Urine collection was more cumbersome especially in

○ Probably yes ● Yes ○ Varies ○ Don't know Voting: Probably yes – 5, Yes – 6, Varies – 1, Abstained - 1.	• LAM implementation required little increase in clinician workload and no additional workspace • Test successfully performed at the point of care, no need to transport samples, no need of laboratory, no additional equipment • Test was perceived as easy to use with good inter-reader agreement • Most patients were able to submit a urine sample in contrast to sputum samples • LAM results available in very short time and allowed TB treatment initiation on the same day Challenges of using LAM: • There maybe challenges with reading grade 1 and interpreting faint bands. • It is important to train on the interpretation of results and ensure the use of the reading card. • CD4 to select patients is problematic because not always immediately available. • Alternative clinical criteria such as seriously ill alone would miss a lot of patients who could benefit from LAM. Qualitative study (N. Engel) Advantages of using LAM: • Urine sample is easily available, less stigmatized & safe • Minimal user skills • Low maintenance/equipment requirements • Short TAT of 25' Challenges of using LAM: • Not everybody can produce, or collect urine samples • Visibility of faint results • Stockouts of urine containers, micropipettes unavailable, no running water/toilets for patients • Delays in Rx initiation	younger and sicker children as it requires both the child's and the caregiver's cooperation and may be affected by medical causes such as dehydration (Kroidl 2015). *What seems simple in the actual scale up may be difficult.* *Everything is feasible, giving proper resources, but millions are spent already for this test without much progress.*

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Desirable Effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable Effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of evidence	Very low	Low	**Moderate**	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	**Favours the intervention**	Varies	Don't know
Resources required	Large costs	**Moderate costs**	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**

Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the interventio n	**Varies**	No included studies
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	**Increased**	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	Probably yes	**Yes**		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ○	**Strong recommendation for the intervention** ●

Conclusions

Recommendation

In inpatient settings, WHO recommends using AlereLAM to assist in the diagnosis of active TB **in HIV-positive adults, adolescents and children with signs or symptoms of TB (pulmonary and extrapulmonary) or advanced HIV disease or who are seriously ill** (strong recommendation; moderate certainty in the evidence about the intervention effects).

Remark: AlereLAM should not be used as a replacement or triage test. It should be used as add on to clincial judgment in combination with other tests.

*** as per prior definition**

The recommendation for seriously ill PLHIV also applies to outpatient settings.

Implementation considerations

in many settings sequential testing during multiple visits may be challenging to implement

Needs to be done in the context of an algorithm that considers other testing

Quality control of the assay will have to be undertaken.

Use of the reading card when applying the test

For children in particular hygenic sample conditions

Monitoring and evaluation

Data collection and linkage to other assays

Research priorities

Extrapulmonary disease data

Algorithms

Data in Children

LAM positive has higher mortality risk - is this a different group?

Global data

Implementation studies pragmatic and operational studies

PICO 2. Should AlereLAM vs. no AlereLAM be used for HIV-positive adults to reduce mortality associated with advanced HIV disease, inpatient setting, CD4 ≤ 200?

Population:	HIV-positive adults to reduce mortality associated with advanced HIV disease, inpatient setting, CD4 ≤ 200
Intervention:	AlereLAM
Comparison:	no AlereLAM
Main outcomes:	Mortality;
Setting:	inpatient

Assessment

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Tuberculosis (TB) remains the leading cause of hospitalization and in-hospital deaths among people living with HIV despite the increased access to antiretroviral treatment (ART) (Ford 2016). A systematic review of the prevalence of TB identified at autopsy suggests that, in resource-limited settings, TB is responsible for around 40% of all HIV-related deaths and that TB often was disseminated and undiagnosed at the time of death (Gupta 2015). Globally in 2017, only 51% of the estimated 10.0 million TB cases were notified among people living with HIV (WHO Global Report 2018). However, most death from TB is preventable if TB is detected early and effectively treated.	Non-sputum-based point-of-care TB diagnostic tests are highly desired to narrow the diagnostic gap and ensure timely treatment. Detection of mycobacterial antigen in urine is promising, as this would allow for a TB diagnosis that is non-site specific. Urine is easy to collect and store, and lacks the infection control risks associated with sputum collection. The lateral flow assay, Alere Determine™ TB LAM Ag assay 'AlereLAM', was developed as a simple point-of-care test for diagnosis of active TB in people living with HIV. AlereLAM is commercially available, does not require access to special laboratory equipment, and produces a result after 25 minutes,

		meeting many desired target product profile requirements.

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	**Outcomes** / **With no AlereLAM** / **With AlereLAM** / **Difference** / **Relative effect (95% CI)** Mortality follow up: 56 weeks / 285 per 1,000 / **248 per 1,000** (219 to 282) / **37 fewer per 1,000** (65 fewer to 3 fewer) / **RR 0.87** (0.77 to 0.99)	*37 saved lives per 100 admitted patients is a large effect.* *It will be further augmented by a reduction of transmission.* *From TB/HIV community perspective, every saved life is a large effect.*

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	**Outcomes** / **With no AlereLAM** / **With AlereLAM** / **Difference** / **Relative effect (95% CI)** Mortality follow up: 56 weeks / 285 per 1,000 / **248 per 1,000** (219 to 282) / **37 fewer per 1,000** (65 fewer to 3 fewer) / **RR 0.87** (0.77 to 0.99)	*Patient important outcomes:* *missed cases* *overtreatment* *empirical treatment was approximately balanced - 6% difference in treatment which will translate into some side effects (which will be balanced against the benefits of a mortality reduction).* *TB treatment in presumably false-positive cases would have positive external effect of TB preventive therapy, which partially compensate for the negative effect of overtreatment.*

Certainty of evidence

What is the overall certainty of the evidence of effects?

Judgement	Research evidence	Additional considerations

○ Very low ○ Low ● Moderate ○ High ○ No included studies	<table><tr><th>Outcomes</th><th colspan="2">Anticipated absolute effects* (95% CI)</th><th>Relative effect (95% CI)</th><th>№ of participants (studies)</th><th>Certainty of the evidence (GRADE)</th><th>Comments</th></tr><tr><td></td><td>Risk with no AlereLAM</td><td>Risk with AlereLAM</td><td></td><td></td><td></td><td></td></tr><tr><td>Mortality follow up: 56 weeks</td><td colspan="2">Study population</td><td>RR 0.87 (0.77 to 0.99)</td><td>2886 (2 RCTs)</td><td>⨁⨁⨁◯ MODERATE[a,b]</td><td></td></tr><tr><td></td><td>285 per 1,000</td><td>248 per 1,000 (219 to 282)</td><td></td><td></td><td></td><td></td></tr></table>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. 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.	

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	*It is likely that no important variability exists in how much people value following important outcomes:* *Mortality.* *Cure from (TB).* *Treatment side effects (in false positives).* *Drug resistance.*	

Balance of effects

Does the balance between desirable and undesirable effects favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations

○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ● Favours the intervention ○ Varies ○ Don't know	*Summary of the above: Large benefits, Small harms.* *Probably very little variation to how people value the outcomes*	

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ● Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	**Systematic review by A. Zwerling:** No detailed micro-costing of AlereLAM implementation was performed in inpatient settings. Using published costing data for South Africa limited to unit test cost, Boyles 2018 calculated the cost per patient for each AlereLAM containing algorithm among inpatients. Cost per patient screened by each algorithm generally increased with increasing diagnostic yield and ranged from $10.5 for Xpert/Culture and AlereLAM/Xpert, $12.5 for the AlereLAM/Xpert/culture, $37.2 for the AlereLAM/Xpert SI, $49.6 for Xpert SI/culture, and $42 for AlereLAM/Xpert SI/culture approach. Boyles 2018 did not perform a cost-effectiveness analysis or calculate incremental cost-effectiveness ratios.	*While implemented in hospitals presents relatively low incremental cost. Implementation in a weak health system would cost more.* *Cost of avoided transmission need to be equated.* *Cost will vary depending if the LAM alone or Dx cascade are implemented?* *Cost will differ by context.*

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	**Systematic review by A. Zwerling:** Models found cost-effectiveness of AlereLAM to be robust across a variety of sensitivity analyses, variations in key parameters and across different country settings and scenarios. Key parameters that are likely influential on cost-effectiveness include: TB prevalence, target population, and AlereLAM specificity, cost of treating TB and HIV and life expectancy post TB survival, and time horizon. However, one detailed micro-costing study published in 2018 estimates unit test costs for AlereLAM implementation several fold higher ($23) than most current models ($2-4). Modeling studies may contribute to the certainty of the evidence in this domain	*Modeling studies may contribute to the certainty of the evidence in this domain* *High variability* *Only one trial* *Variety of models* *No empirical studies*

Cost effectiveness

Does the cost-effectiveness of the intervention favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations

○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ● Varies ○ No included studies	**Systematic review by A. Zwerling:** Reddy et al assessed cost-effectiveness of AlereLAM algorithms in unselected hospitalized PLHIV. Using the modified CEPAC-I model calibrated to STAMP trial results, Reddy 2019 found Xpert + AlereLAM + urine Xpert to be cost-effective among unselected hospitalized HIV patients with ICERs of $450/YLS (Years of life saved, YLS) in Malawi and $840/YLS in South Africa compared with standard of care (Xpert alone). The modified intervention of Xpert + AlereLAM was even more cost-effective with ICERs of $420/YLS in Malawi and $810/YLS in South Africa compared with standard of care. Increased ICERs are due to inclusion of downstream costs associated with lifelong ART and HIV care.	*Only data for Africa are available* *SA results more definitive and Malawi results are less definitive*

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ○ Probably increased ● Increased ○ Varies ○ Don't know	As test can be performed at all levels of the health care system, it will likely increase health care equity.	*Universal test* *Potential to reduce inequity* *Because extrapulmonary TB are already disadvantaged it is potential to improve care for them*

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Report on user perspectives on TB LAM testing: results from qualitative research : Test is generally described as acceptable by key stakeholders.	**Patients:** **Providers:** **Policy-makers/programs:** **Payers:** **Others:** In children: Urine collection was more cumbersome especially in younger and sicker children as it requires both the child's and the caregiver's cooperation and may be affected by medical causes such as dehydration (Kroidl 2015).

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations

○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	**MSF study (H. Huerga)** **Advantages of using LAM:** • LAM implementation required little increase in clinician workload and no additional workspace • Test successfully performed at the point of care, no need to transport samples, no need of laboratory, no additional equipment • Test was perceived as easy to use with good inter-reader agreement • Most patients were able to submit a urine sample in contrast to sputum samples • LAM results available in very short time and allowed TB treatment initiation on the same day **Challenges of using LAM**: • There maybe challenges with reading grade 1 and interpreting faint bands. • It is important to train on the interpretation of results and ensure the use of the reading card. • CD4 to select patients is problematic because not always immediately available. • Alternative clinical criteria such as seriously ill alone would miss a lot of patients who could benefit from LAM. **Qualitative study (N. Engel)** **Advantages of using LAM:** • Urine sample is easily available, less stigmatized & safe • Minimal user skills • Low maintenance/equipment requirements • Short TAT of 25' **Challenges of using LAM**: • Not everybody can produce, or collect urine samples • Visibility of faint results • Stockouts of urine containers, micropipettes unavailable, no running water/toilets for patients • Delays in Rx initiation	In children: Urine collection was more cumbersome especially in younger and sicker children as it requires both the child's and the caregiver's cooperation and may be affected by medical causes such as dehydration (Kroidl 2015). *What seems simple in the actual scale up may be difficult.* *Everything is feasible, giving proper resources, but millions are spent already for this test without much progress.*

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Desirable Effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable Effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of evidence	Very low	Low	**Moderate**	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	**Favours the intervention**	Varies	Don't know
Resources required	Large costs	**Moderate costs**	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know

Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the interventio n	**Varies**	No included studies
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	**Increased**	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	Probably yes	**Yes**		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	**Strong recommendation for the intervention**
○	○	○	○	●

Conclusions

Recommendation

In inpatient settings, WHO suggests using AlereLAM to assist in the diagnosis of active TB in HIV-positive adults irrespective of TB symptoms with a CD4 count < 200 (strong recommendation; moderate certainty in the evidence about the intervention effects).

PICO3. Should AlereLAM be used to diagnose active TB in HIV-positive adults with TB symptoms, outpatient settings?

Population:	HIV-positive adults with TB symptoms, outpatient settings
Intervention:	AlereLAM
Setting:	outpatient

Assessment

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Tuberculosis (TB) remains the leading cause of hospitalization and in-hospital deaths among people living with HIV despite the increased access to antiretroviral treatment (ART) (Ford 2016). A systematic review of the prevalence of TB identified at autopsy suggests that, in resource-limited settings, TB is responsible for around 40% of all HIV-related deaths and that TB often was disseminated and undiagnosed at the time of death (Gupta 2015). Globally in 2017, only 51% of the estimated 10.0 million TB cases were notified among people living with HIV (WHO Global Report 2018). However, most death from TB is preventable if TB is detected early and effectively treated.	Non-sputum-based point-of-care TB diagnostic tests are highly desired to narrow the diagnostic gap and ensure timely treatment. Detection of mycobacterial antigen in urine has attracted great attention over time. Urine-based antigen testing would allow for a TB diagnosis that is non-site specific. Urine is further easy to collect and store, and lacks the infection control risks associated with sputum collection. Multiple platforms have been developed to detect lipoarabinomannan (LAM), initially as enzyme-linked immunosorbent (ELISA) assays that were evaluated in several clinical settings. Later, the lateral flow assay, Alere Determine™ TB LAM Ag assay 'AlereLAM', was developed as a simple point-of-care test for diagnosis of active TB in people living with HIV. AlereLAM is commercially available, does not require access to special laboratory equipment, and produces a result

		after 25 minutes, meeting many desired target product profile requirements.

Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurat ○ Inaccurat e ● Accurate ○ Very accurate ○ Varies ○ Don't know	**Test accuracy** AlereLAM Sensitivity: 0.29 (95% CI: 0.17 to 0.47) Specificity: 0.96 (95% CI: 0.91 to 0.99) Prevalence (Pre-testing probability) 10% Typically seen in symptomatic persons in outpatient settings	The review only included studies with a microbiological reference standard (culture or Xpert). The review, **does not** assess performance against a c**omposite reference standard** that uses microbiological or clinical information to classify TB. This was done in the original WHO and Cochrane Review (WHO Lipoarabinomannan Policy Guidance 2015; Shah 2016), but found little difference against a microbiological reference standard. A substantial number of TB cases may not be verified by microbiological testing if only sputum is tested and when patients with advanced HIV are assessed, which may lead to underestimation of **sensitivity** and

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 10%	Prevalence 30%		
True positives patients with active TB	3 (2 to 5)	29 (17 to 47)	87 (51 to 141)	409 (4)	⨁⨁◯◯ LOW[a,b,c]
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	950 (901 to 980)	864 (819 to 891)	672 (637 to 693)	787 (4)	⨁⨁◯◯ LOW[a,d,e]
False positives patients incorrectly classified as having active TB	40 (10 to 89)	36 (9 to 81)	28 (7 to 63)		

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.

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.

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.

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.

	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.	increase of number of **FN.** Furthermore, while CD4 counts decrease, the sickest patients may not be able to produce a sputum specimen or they have extrapulmonary TB. Exclusion of latter patient groups (individuals without sputa) may also lead to underestimation of sensitivity. Impact of nontuberculous mycobacteria and other environmental factors on test specificity remains unclear, but may possibly lead to **FP** results. Only a single study outside of Africa

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	In outpatients settings, (**10% prevalence**) out of 1000 patients with signs and symptoms of TB, for **29** patients the TB diagnosis will be correctly established. Out of **100** patients with positive test result, **only 45** would actually have active TB, and thus benefit from rapid diagnosis and early treatment initiation Furthermore, in outpatients settings, out of 1000 patients with signs and symptoms of TB, for **864** patients the TB diagnosis will be correctly excluded. Out of **100** patients with negative test result, **92** would actually not have active TB, and thus benefit from sparing the unnecessary treatment; and also benefit of reassurance and alternative diagnosis. However, out of 1000 patients with signs and symptoms of TB, for **36** patients the TB diagnosis will be falsely established. Out of 100 patients with positive test result, **55** would not have active TB, and thus would have risk of unnecessary treatment and stigma. Furthermore, out of 1000 patients with signs and symptoms of TB, for 71 patients the TB diagnosis will be missed. Out of 100 patients with negative test result, 8 would actually not have active TB, and thus will be exposed to increased risk of morbidity and mortality, delayed treatment initiation and pose the continued risk of transmission.	As urine LAM does not provide information about drug resistance, thus a positive result (both TP and FP) will necessitate additional testing (Xpert, culture) in order to identify evidence for phenotypic or molecular drug resistance. As urine LAM sensitivity does not allow

	Change of the PTT would mostly affect number of FN: 10% - 71; 30% - 213.	identification of all cases of MTB, additional testing may be required following a negative result (TN and FN). As the test can be performed on an easy to collect urine sample outside a laboratory, the time to diagnosis can be reduced The desirable effect of the test may be further augmented by the fact that in low-resource settings, certain proportion of TB patients may be diagnosed by LF-LAM and not by WHO recommended rapid TB diagnostic test (Xpert) due to the following reasons: 1) sputum Xpert has lower sensitivity in HIV-positive than HIV-negative people; 2) patients may not be able to produce sputum; 3) patients may not have access to Xpert.

Undesirable Effects

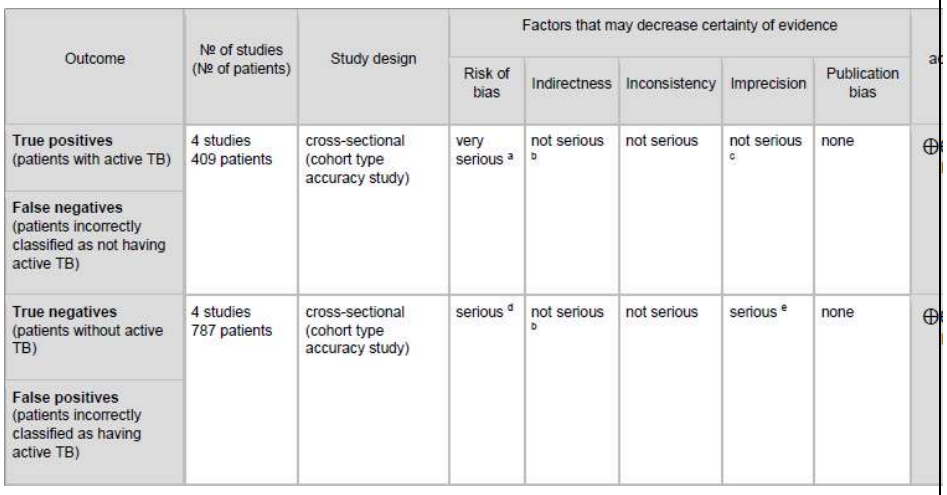
How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations

○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know	In outpatients settings, (**10% prevalence**) out of 1000 patients with signs and symptoms of TB, for **29** patients the TB diagnosis will be correctly established. Out of **100** patients with positive test result, **only 45** would actually have active TB, and thus benefit from rapid diagnosis and early treatment initiation Furthermore, in outpatients settings, out of 1000 patients with signs and symptoms of TB, for **864** patients the TB diagnosis will be correctly excluded. Out of **100** patients with negative test result, **92** would actually not have active TB, and thus benefit from sparing the unnecessary treatment; and also benefit of reassurance and alternative diagnosis. However, out of 1000 patients with signs and symptoms of TB, for **36** patients the TB diagnosis will be falsely established. Out of 100 patients with positive test result, **55** would not have active TB, and thus would have risk of unnecessary treatment and stigma. Furthermore, out of 1000 patients with signs and symptoms of TB, for 71 patients the TB diagnosis will be missed. Out of 100 patients with negative test result, 8 would actually not have active TB, and thus will be exposed to increased risk of morbidity and mortality, delayed treatment initiation and pose the continued risk of transmission. *Change of the PTT would mostly affect number of FN: 10% - 71; 30% - 213.*	As CD4 counts decrease, the sickest patients may be not able to produce a sputum specimen or would have EPTB. Thus their **TP** results would not be confirmed by microbiological reference standard and will be misclassified as **FP**. NTM and other environmental factors may possibly lead to **FP** results. *Undesirable effects maybe partially compensated by the use of other tests in an algorithm.*

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ● Low ○ Moderate ○ High ○ No included studies		

Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					a
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
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	⊕
False negatives (patients incorrectly classified as not having active TB)								
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	⊕
False positives (patients incorrectly classified as having active TB)								

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.

	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. 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. 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. 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.	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	No adverse events were associated with LAM testing (Peter 2016). High quality evidence. Even though, Dx trial may not capture side effects as effectively as treatment trials, in case of major side-effects would occur likely they would be reported.	Similarly as for inpatients (see 1.1) POC benefits for the test. We assume that there are no adverse effects associated with the test. The AlereLAM assay can be performed at the patient bedside, in a clinic or a laboratory with minimal training. There was no difficulty in urine collection, no discussion on patient harms, it was deemed as easy to perform. Direct benefit – being quickly diagnosed. Risk - Not following the result, if test is not part of an algorithm.

		Burden is considered unimportant .

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	2 RCTs in hospitalized HIV-positive adults showed a decrease in mortality associated with use of the AlereLAM comparing to the non-use of AlereLAM. Pooled risk ratio was 0.85 (95% CI 0.76 to 0.94) i.e. study participants undergoing AlereLAM testing had 0.85 times the risk or 15% lower risk of mortality than participants undergoing routine TB diagnostic testing without AlereLAM. The absolute effect was 35 fewer deaths per 1,000 (from 14 fewer to 55 fewer) (high-certainty evidence).	In additional analyses, the review authors demonstrated that within diagnostic accuracy studies that included follow-up for clinical outcomes, without using AlereLAM results for clinical decision making, there appeared to be an association between AlereLAM positivity among both participants with and without confirmed TB (by microbiological and/or clinical study reference standards) and mortality. These data must be interpreted

		cautiously as they represent secondary analyses within observational cohorts, are limited in size, and may not control for important biases or other factors. It is likely that these findings may represent the effect of missed diagnoses (that could be averted through earlier diagnosis using rapid AlereLAM testing) and/or that there is a biological association between disease severity resulting in AlereLAM excretion in urine.

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Report on user perspectives on TB LAM testing: results from qualitative research: testing makes a critical difference for many patients from a hard to diagnose group. In one RCT, > 95% of clinicians acted on a positive test result (Peter 2016).	The tests can be performed at the patient bedside, in a clinic or a laboratory with minimal training. Patients in trials do promptly receive anti-TB therapy after testing and there is little concern that this would not happen outside of trials. For a negative LAM test result, clinicians may use empirical treatment for TB.

Certainty of effects

What is the overall certainty of the evidence of effects of the test?

Judgement	Research evidence	Additional considerations
○ Very low ● Low ○ Moderate ○ High ○ No included studies	Summary of the above conclusions	

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	It is likely that no important variability exists in how much people value following important outcomes: Mortality. Cure from (TB). Treatment side effects (in false positives). Drug resistance.	

Balance of effects

Does the balance between desirable and undesirable effects favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations

○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know	Summary of the above	Given the high mortality in persons living with HIV, acting on all positive LAM results likely balances any possible adverse effects associated with unnecessary treatment with reducing mortality.

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ● Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	**Systematic review by A. Zwerling:** Mukora 2018 employed a detailed micro-costing approach among outpatient clinics testing PLHIV with CD4 ⩽150 cells/μL including costs from both the clinic level and above clinic level, across non-governmental organizations (NGO) and department of health (DoH) implementers/clinics and included costs from both start-up and implementation periods. Mukora 2018 estimated a total unit cost of AlereLAM testing at $23.55 (NGO clinics) and $22.72 (department of health (DOH) operated clinics). Unit costs were higher than have been reported in other studies from South Africa (~$3-4.00) largely driven by the inclusion of both clinic level ($11.49 NGO & $10.85 DOH) and above clinic level costs ($12.06 NGO & $11.87 DOH).	Impacts modelling data on reduction and transmission Cost is more significant for outpatient, exceeding 3.5 USD, assuming investment in outpatient activities. Use of LF-LAM should be seen as part of algorithm. Better diagnostics always include additional cost

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	**Systematic review by A. Zwerling:** Models found cost-effectiveness of AlereLAM to be robust across a variety of sensitivity analyses, variations in key parameters and across different country settings and scenarios. Key parameters that are likely influential on cost-effectiveness include: TB prevalence, target population, and AlereLAM specificity, cost of treating TB and HIV and life expectancy post TB survival, and time horizon. However, one detailed micro-costing study published in 2018 estimates unit test costs for AlereLAM implementation several fold higher ($23) than most current models ($2-4). Modeling studies may contribute to the certainty of the evidence in this domain	

Cost effectiveness

Does the cost-effectiveness of the intervention favour the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ● Varies ○ No included studies	**Systematic review by A. Zwerling:** Models consistently demonstrated AlereLAM containing approaches could be cost-effective among African HIV positive adults across a range of settings and parameters evaluated despite heterogeneous diagnostic approaches evaluated.	

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations

○ Reduced ○ Probably reduced ○ Probably no impact ○ Probably increased ● Increased ○ Varies ○ Don't know	As test can be performed at all levels of the health care system, it will likely increase health care equity.	

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Report on user perspectives on TB LAM testing: Test is generally described as acceptable by key stakeholders.	**Patients:** **Providers:** **Policy-makers/programs:** **Payers:** **Others:** In children: Urine collection was more cumbersome especially in younger and sicker children as it requires both the child's and the caregiver's cooperation and may be affected by medical causes such as dehydration (Kroidl 2015).

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no	**MSF study (H. Huerga)** **Advantages of using LAM:** • LAM implementation required little increase in clinician workload and no additional workspace	At scale implementation may be tricky.

○ Probably yes ● Yes ○ Varies ○ Don't know	• Test successfully performed at the point of care, no need to transport samples, no need of laboratory, no additional equipment • Test was perceived as easy to use with good inter-reader agreement • Most patients were able to submit a urine sample in contrast to sputum samples • LAM results available in very short time and allowed TB treatment initiation on the same day **Challenges of using LAM**: • There maybe challenges with reading grade 1 and interpreting faint bands. • It is important to train on the interpretation of results and ensure the use of the reading card. • CD4 to select patients is problematic because not always immediately available. • Alternative clinical criteria such as seriously ill alone would miss a lot of patients who could benefit from LAM. **Qualitative study (N. Engel)** **Advantages of using LAM:** • Urine sample is easily available, less stigmatized & safe • Minimal user skills • Low maintenance/equipment requirements • Short TAT of 25' **Challenges of using LAM**: • Not everybody can produce, or collect urine samples • Visibility of faint results • Stockouts of urine containers, micropipettes unavailable, no running water/toilets for patients • Delays in Rx initiation	Concerns were raised by the panel about quality control that needs to be implemented. Outpatient setting will add additional challenges. In children: Urine collection was more cumbersome especially in younger and sicker children as it requires both the child's and the caregiver's cooperation and may be affected by medical causes such as dehydration (Kroidl 2015).

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable Effects	Trivial	Small	**Moderate**	Large		Varies	Don't know
Undesirable Effects	Large	**Moderate**	Small	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	**Low**	Moderate	High			No included studies
Certainty of the evidence of test's effects	Very low	Low	**Moderate**	High			No included studies
Certainty of the evidence of management's effects	Very low	Low	**Moderate**	High			No included studies

	Judgement						
Certainty of the evidence of test result/management	Very low	Low	**Moderate**	High			No included studies
Certainty of effects	Very low	**Low**	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the interventio n	Varies	Don't know
Resources required	Large costs	**Moderate costs**	Negligible costs and savings	Moderat e savings	Large savings	Varies	Don't know
Certainty of evidence of required resources	**Very low**	Low	Moderate	High			No included studies
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the interventio n	Favours the interventio n	**Varies**	No included studies
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	**Increased**	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	Probably yes	**Yes**		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation
1.2. In outpatient settings, WHO suggests using AlereLAM to assist in the diagnosis of active TB in HIV-positive adults, adolescents and children with TB with signs or symptoms of TB (pulmonary and extrapulmonary) (conditional recommendation; low certainty in the evidence about test accuracy).

3.6 Evidence-to-decision tables: Low complexity automated NAATs

PICO 4. Should Low complexity automated NAATs on sputum be used to diagnose INH resistance in patients with microbiologically confirmed pulmonary TB, irrespective of resistance to RIF, MRS?	
Population:	patients with microbiologically confirmed pulmonary TB, irrespective of resistance to RIF, MRS
Intervention:	MTB/XDR assay on sputum
Setting:	outpatient and inpatient
Subgroups:	

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Emerging data suggest that, in some settings, **RR** testing has suboptimal specificity for MDR-TB (WHO Global tuberculosis report 2020). This means that testing for resistance to **isoniazid** is increasingly important. For instance, a study in DRC found one in five **RR** patients to be **isoniazid** susceptible (Bismwa 2020), and the most recent South African National Survey of Drug Resistance found hotspots of rifampicin mono-resistance, where the prevalence ratio of such cases exceeded that of MDR-TB by as much as 30% (NICD 2016). Conversely, **isoniazid** resistance in the presence of rifampicin susceptibility (isoniazid mono-resistance) is also increasingly recognised as another emerging challenge in managing tuberculosis as it is an important enabler of MDR-TB (Sulis 2020).	
Test accuracy How accurate is the test?		
Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ○ Accurate ● Very accurate	**Test accuracy**	

○ Varies ○ Don't know	MTB/XDR assay on sputum Sensitivity: 0.94 (95% CI: 0.89 to 0.97) Specificity: 0.98 (95% CI: 0.95 to 0.99)	

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	Rapid extended drug resistance profiling allows for early initiation of optimized therapy and likely better patient outcomes. Amplification of drug resistance would be less likely. Information on i*nhA* promotor mutations could also guide high dose isoniazid therapy.	the assumption is that in many settings phenotypic testing may not be available or testing may not be done.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 5%	Prevalence 10%		
True positives patients with INH resistance	9 (9 to 10)	47 (45 to 49)	94 (89 to 97)	994 (3)	⨁⨁⨁◯ MODERATE[a,b]
False negatives patients incorrectly classified as not having INH resistance	1 (0 to 1)	3 (1 to 5)	6 (3 to 11)		
True negatives patients without INH resistance	970 (942 to 982)	931 (904 to 942)	882 (857 to 893)	611 (3)	⨁⨁⨁◯ MODERATE[a]
False positives patients incorrectly classified as having INH resistance	20 (8 to 48)	19 (8 to 46)	18 (7 to 43)		

a. The median prevalence of isoniazid resistance in the included studies was 67.2% (range, 26.8% (DIAMA, Benin) to 93.9% (FIND, Moldova), higher than the three prevalences in the GRADE table. Applicability to settings with a lower prevalence of isoniazid resistance comes with some uncertainty. Although the population for this PICO question is 'irrespective of rifampicin resistance,' owing to enrollment criteria in the studies, we note that most participants were rifampicin resistant. We did not downgrade for indirectness.

b. Sensitivity estimates ranged from 81% (FIND, New Delhi) to 100% (DIAMA, Rwanda). Regarding the low sensitivity estimate in New Delhi, the study authors reported that sequencing did not show the presence of variants typically associated with resistance in many phenotypically isoniazid-resistant samples suggesting that variants not analyzed by Xpert MTB/XDR might play a role. We did not downgrade for inconsistency. This was a judgement.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	There is uncertainty about test performance in patients with paucibacillary disease. If only used as a reflex test following an Xpert MTB/RIF or Xpert Ultra positive rifampicin resistant result, then the test will not detect non-MDR isoniazid resistance. The test detects a subset of all known INH resistance. A false positive in a non-RR patient would lead to the regimen being changed.	Cepheid 2020 Of 530 specimens tested, 512 had pDST results available. Of these 512 specimens with pDST results available, 32 (6.3%) were Xpert MTB/XDR MTB NOT DETECTED. By the pDST reference standard, of these 32 specimens, two (6.3%) were resistant and 30 (93.8%) were susceptible.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 5%	Prevalence 10%		
True positives patients with INH resistance	9 (9 to 10)	47 (45 to 49)	94 (89 to 97)	994 (3)	⨁⨁⨁⨁ HIGH[a,b]
False negatives patients incorrectly classified as not having INH resistance	1 (0 to 1)	3 (1 to 5)	6 (3 to 11)		
True negatives patients without INH resistance	970 (942 to 982)	931 (904 to 942)	882 (857 to 893)	611 (3)	⨁⨁⨁⨁ HIGH[a]
False positives patients incorrectly classified as having INH resistance	20 (8 to 48)	19 (8 to 46)	18 (7 to 43)		

a. The median prevalence of isoniazid resistance in the included studies was 67.2% (range, 26.8% (DIAMA, Benin) to 93.9% (FIND, Moldova), higher than the three prevalences in the GRADE table. Applicability to settings with a lower prevalence of isoniazid resistance comes with some uncertainty. Although the population for this PICO question is 'irrespective of rifampicin resistance,' owing to enrollment criteria in the studies, we note that most participants were rifampicin resistant. We did not downgrade for indirectness.

b. Sensitivity estimates ranged from 81% (FIND, New Delhi) to 100% (DIAMA, Rwanda). Regarding the low sensitivity estimate in New Delhi, the study authors reported that sequencing did not show the presence of variants typically associated with resistance in many phenotypically isoniazid-resistant samples suggesting that

variants not analyzed by Xpert MTB/XDR might play a role. We did not downgrade for inconsistency. This was a judgement.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	The overall certainty of the evidence was high.	c.

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 1%	Effect per 1000 patients/year for pre-test probability of 5%	Effect per 1000 patients/year for pre-test probability of 10%	Importance
True positives	cross-sectional (cohort type accuracy study)	⨁⨁⨁◯ MODERATE[a,b]	9 (9 to 10)	47 (45 to 49)	94 (89 to 97)	
False negatives			1 (0 to 1)	3 (1 to 5)	6 (3 to 11)	
True negatives	cross-sectional (cohort type accuracy study)	⨁⨁⨁◯ MODERATE[a]	970 (942 to 982)	931 (904 to 942)	882 (857 to 893)	
False positives			20 (8 to 48)	19 (8 to 46)	18 (7 to 43)	

a. The median prevalence of isoniazid resistance in the included studies was 67.2% (range, 26.8% (DIAMA, Benin) to 93.9% (FIND, Moldova), higher than the three prevalences in the GRADE table. Applicability to settings with a lower prevalence of isoniazid resistance comes with some uncertainty. Although the population for this PICO question is 'irrespective of rifampicin resistance,' owing to enrollment criteria in the studies, we note that most participants were rifampicin resistant. We did not downgrade for indirectness.

b. Sensitivity estimates ranged from 81% (FIND, New Delhi) to 100% (DIAMA, Rwanda). Regarding the low sensitivity estimate in New Delhi, the study authors reported that sequencing did not show the presence of variants typically associated with resistance in many phenotypically isoniazid-resistant samples suggesting that variants not analyzed by Xpert MTB/XDR might play a role. We did not downgrade for inconsistency. This was a judgement.

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low	No adverse events or side effects were reported by any of the sites in the FIND study.	Although a diagnostic study may not capture

● Moderate ○ High ○ No included studies		adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	There are no randomized trials on the effect on patient-important outcomes of using the test.	A positive result for resistance would mean modification of the treatment regimen, and a negative result would mean preserving INH in the treatment regimen.

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	Observations from clinical practice suggest that clinicians will make decisions based on test results and individualise the regimen using them.	Clinicians and TB programmes are familiar with Xpert testing. The challenges with feasibility and the resources required mean that clinicians may not be able to order Xpert MTB/XDR testing in some settings. **WHO recommendation: In patients with confirmed rifampicin-susceptible, isoniazid-resistant tuberculosis, treatment with rifampicin, ethambutol, pyrazinamide and levofloxacin is recommended for a duration of 6 months.**

		WHO recommendation: In patients with confirmed rifampicin-susceptible, isoniazid-resistant tuberculosis, it is not recommended to add streptomycin or other injectable agents to the treatment regimen.
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies	This is the summary of the preceding points5-8	moderate certainty in accuracy
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me"), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence). Compared to existing tests/sputum microscopy, **healthcare professionals** appreciate the rapidity of CB-NAAT results, the accuracy of **CB-NAAT** results and the confidence that this generates to start treating and motivate patients, the diversity of sample types, the ability to detect drug resistance (earlier or at all, for as many drugs as possible and altering clinician's risk perception of drug resistance in children), as well as the consequence of avoiding costlier investigations or hospital stays when using CB-NAAT. (QES: high confidence). The cartridge has a quicker turnaround time for first and second line drug susceptibility testing, compared to other available diagnostic methods. People value faster TAT, the potential ability to reflex samples from the Xpert MTB/RIF to the Xpert MTB/XDR cartridge, and receiving information on multiple drugs as well as high or low level resistance simultaneously, as it could enable quicker diagnosis and optimized treatment for patients. (Interview study)**Laboratory technicians** appreciate the improvement of overall laboratory work that **CB-NAAT** brings compared to sputum microscopy in terms of ease of use, ergonomics, and biosafety (QES: high confidence). It requires minimal user steps and the GeneXpert platform is a familiar system which people feel comfortable running and interpreting (Interview study).	

	Laboratory managers appreciate that monitoring of laboratory work and training is easier than with sputum microscopy and that **CB-NAAT** eases staff retention, as it increases staff satisfaction and has a symbolic meaning of progress within the TB world (QES: low confidence)	

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ● Favors the intervention ○ Varies ○ Don't know		**The reference standard is phenotypic DST (the comparator)** Clinical benefit has not been evaluated here. Clinical benefit would be superior in terms of speed of treatment. in some settings the comparator Desirable outweight undesirable effects but there is uncertainty of the evidence which did not make all members of the panel confident that there is more benefit.

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderat e savings ○ Large savings ● Varies ○ Don't know	No direct evidence from published studies regarding total resources required. Resource requirements will include the purchase of cartridges ($19.80USD/cartridge), upgrading of existing platforms to 10-colour modules (an upgrade that will be required eventually for all Xpert platforms: $3860 to >$72,350) and operational and programmatic costs associated with implementing the novel diagnostic. Resource requirements for XDR treatment (drugs, hospital capacity, staff, etc.) likely will also increase with increasing numbers diagnosed. Total costs will vary depending on testing volume and prevalence of XDR in the population. Budget impact will depend on current standard of care and associated resource use.	

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations

○ Very low ○ Low ○ Moderate ○ High ● No included studies	Direct costs related to cartridge and machinery are provided from the manufacturer while several important items related to resource use including staff time, overhead and operational costs associated with implementing Xpert MTB/XDR have not been investigated. Differences in resource use between Xpert MTB/XDR and existing approaches will vary across settings using different phenotypic and genotypic DST. Important variability exists in costs of staff time and operational costs, such as testing volume across settings.	

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies ● No included	No cost-effectiveness studies were identified using XpertMTB/XDR. Extrapolation of cost-effectiveness data from Xpert MTB/RIF or other CBNATs is not advised due to differences in diagnostic accuracy, costs associated with XDR treatment and the testing and treatment cascade of care.	

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence). Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence). Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study). New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private	

	health system that is better equipped to quickly adopt new diagnostics and policies). (interview study) The identified challenges with **CB-NAAT** utilization and accumulated delays risk compromize the added value as identified by the users, ultimately leading to underutilization and hamper access to prompt and accurate testing and treatment particularly for vulnerable groups**.** (QES: High confidence)	
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Patients can be reluctant to test for TB/MDR-TB because of stigma related to MDR-TB or related to having interrupted treatment in the past, because of fears of side effects, the failure to recognize symptoms, the inability to produce sputum and the cost, distance and travel concerns related to (repeat) clinic visits. (QES: high confidence) Health workers can be reluctant to test for TB or MDR-TB because of TB associated stigma and consequences for their patients, fears of acquiring TB, fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified, fear of side effects of drugs in children, and community awareness of disease manifestations in children. (QES: high confidence) **CB-NAAT** appears widely acceptable by laboratory staff and clinicians based on its simple user steps, familiarity of the system, and due to the amount of important information it provides. (interview study)	
Feasibility Is the intervention feasible to implement?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	**CB-NAAT** seems to decrease workload in the laboratory in terms of freeing up time for laboratory staff, but in most settings the introduction of CB-NAAT increases workload of laboratory staff if added onto existing work without adjusting staffing arrangements, or if it does not replace existing diagnostic tests with the result that staff may be hesitant to accept testing with CB-NAAT**.** (QES: moderate confidence) **The CB-NAAT requires less user training** compared to other DST methods (such as LPA and culture), **making it more feasible to implement** compared to methods with more user steps and those methods which require significant additional training (interview study). However, implementation of new diagnostics must be accompanied with training for clinicians, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge leading to underutilization (QES: high confidence and interview study) or overreliance on **CB-NAAT** results at the expense of clinical acumen (QES: moderate confidence). **Furthermore**, introduction of new diagnostics must be accompanied by **guidelines and algorithms, which support clinicians and laboratories in communicating with each other**, such that they can discuss discordant results, and interpret laboratory results in the context of drug availability, patient history, and patient progress on a current drug regimen.(Interview study).	

	In addition, an **efficient sample transportation system, with sustainable funding mechanisms** is crucial for feasibility, especially if an algorithm requires multiple samples at different times, from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, the sample risks being contaminated and yielding inconclusive results on molecular diagnostics. Participants cited good personnel skill, standardized operating procedures, and significant laboratory infrastructure as essential in reducing sample contamination in their laboratory. (interview study) **Finally:** Implementation of new diagnostics must be accompanied with **training for clinicians**, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge (QES: high confidence and interview study). **BUT, Feasibility is challenged** by accumulation of diagnostic delays and/or underutilization at every step due to mainly health system factors: non-adherence to testing algorithms, testing for (MDR)-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor/delayed sample transport and sample quality, and result communication, delays in scheduling follow up visits and recalling patients, inconsistent result recording; lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners, and sputum containers; dusty environment, and delayed or absent local repair option); inefficient/unclear work- and patient flows (for instance inefficient organizational processes, poor links between providers, unclear follow up mechanisms or where patients need to go); and lack of data-driven and inclusive national implementation processes. These challenges lead to delays and/or underutilization. (QES: high confidence) **Feasibility for the CB-NAAT is also challenged by** the value of diagnosing MTB over DR TB at primary care, makes it less feasible as a baseline test, though it would fit at a district or intermediate level laboratory. The identified **feasibility challenges** with **CB-NAAT** utilization and accumulated delays at every step may compromize the added value/benefits as identified by the users (avoiding delays, keeping cost lost, accurate results, drug resistant information, easing laboratory work), ultimately leading to underutilization (QES: high confidence). We can assume that if these values are not met users are less likely to find CB-NAATs acceptable.	

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	Accurate	**Very accurate**		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of evidence of the test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of test's effects	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of management's effects	Very low	Low	Moderate	High			**No included studies**

	Judgement						
Certainty of the evidence of test result/management	**Very low**	Low	Moderate	High			No included studies
Certainty of effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	**Favors the intervention**	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In patients with bacteriologically-confirmed pulmonary TB, **automated nucleic acid amplification tests of low-complexity** should be used on sputum for detection of **resistance to isoniazid** (rather than culture based phenotypic DST) (Conditional recommendation; moderate certainty of evidence for diagnostic accuracy)

Remarks: considerations for isoniazid resistance testing include caring for patients with possible Hr-TB (isonizid (mono)resistance, rifampicin susceptible disease)

Need to be put in context LPA (both may be used) - this is not a direct comparison with LPA

Applies to population with confirmed and irrespective of rifampicin resistance (same as recommendation 26)

Justification

Cost was a key factor - see above

Implementation considerations

Large demand following this recommendation will require carefull planning

linking to treatment guidelines, including household contacts

Implementation processes have been challenged by lack of data on pragmatic effectiveness in operational conditions, lack of knowledge and awareness among providers beyond lab personnel, lack of guidelines and standardized training modules and instructions and a lack of national policy consensus and inclusive decision-making prior to roll out.

Performance may differ by geographic setting

Monitoring and evaluation

Ongoing surveillance of strains that are resistant but not detected by particular targets

Whether are appropriate regimens implemented.

Research priorities

More studies on performance of INH resistance associated with

Rs Cost and algorithm for overall diagnosis

Guidance on household contacts

Implementation may be on same sputum-SR mix and implications need investigation

Data on children

Use of NAAT on specimens other than sputum

Qualitative studies on acceptability across all groups

PICO 5. Should Moderate complexity automated NAATs on sputum be used to diagnose FQ resistance in patients with microbiologically confirmed pulmonary TB, irrespective of resistance to RIF, MRS?

Population:	patients with microbiologically confirmed pulmonary TB, irrespective of resistance to RIF, MRS
Intervention:	Moderate complexity automated NAAT on sputum

Assessment

Problem

Is the problem a priority?

<table>
<tr><th>Judgement</th><th>Research evidence</th><th>Additional considerations</th></tr>
<tr><td>○ No
○ Probably no
○ Probably yes
● Yes
○ Varies
○ Don't know</td><td>In case rifampicin resistance is detected, rapid molecular tests for resistance to at least fluoroquinolones should be performed promptly to inform the treatment of MDR-TB and XDR-TB. WHO currently recommends using commercially available molecular line probe assays (LPAs) as the initial test to detect resistance to fluoroquinolones for persons with a detected resistance to refampicin (WHO Consolidated Guidelines (Module 3) 2020). Nevertheless more automated, close to patient, accurate diagnostics for second-line anti-TB reagents are urgently needed.</td><td></td></tr>
<tr><td colspan="3">Test accuracy
How accurate is the test?</td></tr>
<tr><th>Judgement</th><th>Research evidence</th><th>Additional considerations</th></tr>
<tr><td>○ Very inaccurate
○ Inaccurate
○ Accurate
● Very accurate
○ Varies
○ Don't know</td><td>Test accuracy

MTB/XDR assay on sputum Sensitivity: 0.93 (95% CI: 0.88 to 0.96) Specificity: 0.98 (95% CI: 0.94 to 0.99)</td><td></td></tr>
<tr><td colspan="3">Desirable Effects
How substantial are the desirable anticipated effects?</td></tr>
<tr><th>Judgement</th><th>Research evidence</th><th>Additional considerations</th></tr>
<tr><td>○ Trivial
○ Small
○ Moderate
● Large
○ Varies
○ Don't know</td><td>Rapid detection of fluoroquinolone resistance is critical. FQs have an essential role in treating RR/MDR-TB and are also important for protecting second-line drugs like bedaquiline. The 2020 World Health Organization consolidated guidelines on drug resistant TB treatment recognize the importance of later generation fluoroquinolones in all-oral regimens of shorter duration.
<table>
<tr><th rowspan="2">Test result</th><th colspan="3">Number of results per 1000 patients tested (95% CI)</th><th rowspan="2">№ of participants (studies)</th><th rowspan="2">Certainty of the evidence (GRADE)</th></tr>
<tr><th>Prevalence 1%</th><th>Prevalence 5%</th><th>Prevalence 10%</th></tr>
<tr><td>True positives
patients with FQ resistance</td><td>9 (9 to 10)</td><td>47 (44 to 48)</td><td>93 (88 to 96)</td><td>384
(3)</td><td>⨁⨁⨁⨁
HIGH[a,b]</td></tr>
<tr><td>False negatives
patients incorrectly</td><td>1 (0 to 1)</td><td>3 (2 to 6)</td><td>7 (4 to 12)</td><td></td><td></td></tr>
</table></td><td></td></tr>
</table>

classified as not having FQ resistance						
True negatives patients without FQ resistance	973 (936 to 985)	934 (898 to 945)	885 (850 to 896)	953 (3)	⊕⊕⊕◯ MODERATE[a,c]	
False positives patients incorrectly classified as having FQ resistance	17 (5 to 54)	16 (5 to 52)	15 (4 to 50)			

a. The median prevalence of fluoroquinolone resistance in the included studies was 24.3% (range, 0.0% (DIAMA, Rwanda) to 58.4% (FIND, Mumbai), higher than the three prevalences listed in the GRADE table. Applicability to settings with lower prevalence of fluoroquinolone resistance comes with some uncertainty. Although the population for this PICO question is 'irrespective of rifampicin resistance,' owing to enrollment criteria in the studies, we note that most participants were rifampicin resistant. We did not downgrade for indirectness.

b. Sensitivity estimates ranged from 83% (FIND, New Delhi) to 100% (DIAMA, Benin and Cameroon). Except for New Delhi, sensitivity was > 90%. We did not downgrade for inconsistency.

c. Specificity estimates were inconsistent: 84% (FIND, Mumbai), 91% (FIND, New Delhi), and > 96% for other studies. We could not explain the heterogeneity in specificity estimates. We downgraded one level inconsistency.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	There is uncertainty about test performance in patients with paucibacillary disease.	The test's decreased ability to detect mutations causing low level fluoroquinolone resistance, especially in hetero-resistant strain populations, is a concern. Cepheid 2020 Of 530 specimens tested, 453 had pDST results available. Of these 453 specimens with pDST results available, 32 (7.1%), were Xpert

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 5%	Prevalence 10%		
True positives patients with FQ resistance	9 (9 to 10)	47 (44 to 48)	93 (88 to 96)	384 (3)	⨁⨁⨁⨁ HIGH[a,b]
False negatives patients incorrectly classified as not having FQ resistance	1 (0 to 1)	3 (2 to 6)	7 (4 to 12)		
True negatives patients without FQ resistance	973 (936 to 985)	934 (898 to 945)	885 (850 to 896)	953 (3)	⨁⨁⨁◯ MODERATE[a,c]
False positives patients incorrectly classified as having FQ resistance	17 (5 to 54)	16 (5 to 52)	15 (4 to 50)		

a. The median prevalence of fluoroquinolone resistance in the included studies was 24.3% (range, 0.0% (DIAMA, Rwanda) to 58.4% (FIND, Mumbai), higher than the three prevalences listed in the GRADE table. Applicability to settings with lower prevalence of fluoroquinolone resistance comes with some uncertainty. Although the population for this PICO question is 'irrespective of rifampicin resistance,' owing to enrollment criteria in the studies, we note that most participants were rifampicin resistant. We did not downgrade for indirectness.

b. Sensitivity estimates ranged from 83% (FIND, New Delhi) to 100% (DIAMA, Benin and Cameroon). Except for New Delhi, sensitivity was > 90%. We did not downgrade for inconsistency.

c. Specificity estimates were inconsistent: 84% (FIND, Mumbai), 91% (FIND, New Delhi), and > 96% for other studies. We could not explain the heterogeneity in specificity estimates. We downgraded one level inconsistency.

MTB/XDR MTB NOT DETECTED.

By the pDST reference standard, of these 32 specimens, one (3.1%) was resistant and 31 (96.9%) were susceptible.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate	The overall certainty of the evidence was moderate owing to serious inconsitency for specificity.	

○ High ○ No included studies	Detailed judgments are provided in the evidence profile.	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Although a diagnostic study may not capture adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.	will add the considerations from the FIND study

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	There are no randomized trials on the effect on patient-important outcomes of using the test. However, there is evidence that inclusion or exclusion of FQs from the regimen strongly impacts outcomes, whereas this is less clear for the other drugs ("…treatment outcomes were significantly better with use of linezolid, later generation fluoroquinolones, bedaquiline, clofazimine, and carbapenems for treatment of multidrug-resistant tuberculosis." Ahmad N, Ahuja SD, Akkerman OW, Alffenaar JC, Anderson LF, Baghaei P, and the Collaborative group for the meta-analysis of individual patient data in MDR-TB treatment. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet. 2018;392(10150):821–834.)	Not impacting all patient management (e.g. for susceptible TB). Judgment for resistant

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Observations from clinical practice suggest that clinicians will make decisions based on test results and individualise the regimen using them.	Clinicians and TB programmes are familiar with Xpert testing. The challenges with feasibility and the resources required mean that clinicians may not be able to order Xpert MTB/XDR testing in some settings.

Certainty of effects

What is the overall certainty of the evidence of effects of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies	This is the summary of the preceding points 5-8	moderate certainty in accuracy
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me"), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence). Compared to existing tests/sputum microscopy, **healthcare professionals** appreciate the rapidity of CB-NAAT results, the accuracy of **CB-NAAT** results and the confidence that this generates to start treating and motivate patients, the diversity of sample types, the ability to detect drug resistance (earlier or at all, for as many drugs as possible and altering clinician's risk perception of drug resistance in children), as well as the consequence of avoiding costlier investigations or hospital stays when using CB-NAAT. (QES: high confidence). The cartridge has a quicker turnaround time for first and second line drug susceptibility testing, compared to other available diagnostic methods. People value faster TAT, the potential ability to reflex samples from the Xpert MTB/RIF to the Xpert MTB/XDR cartridge, and receiving information on multiple drugs as well as high or low level resistance simultaneously, as it could enable quicker diagnosis and optimized treatment for patients. (Interview study) **Laboratory technicians** appreciate the improvement of overall laboratory work that **CB-NAAT** brings compared to sputum microscopy in terms of ease of use, ergonomics, and biosafety (QES: high confidence). It requires minimal user steps and the GeneXpert platform is a familiar system which people feel comfortable running and interpreting (Interview study). **Laboratory managers** appreciate that monitoring of laboratory work and training is easier than with sputum microscopy and that **CB-NAAT** eases staff retention, as it increases staff satisfaction and has a symbolic meaning of progress within the TB world (QES: low confidence)	
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations

○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ● Favors the intervention ○ Varies ○ Don't know		

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	No direct evidence from published studies regarding total resources required. Resource requirements will include the purchase of cartridges ($19.80USD/cartridge), upgrading of existing platforms to 10-colour modules (an upgrade that will be required eventually for all Xpert platforms: $3860 to >$72,350) and operational and programmatic costs associated with implementing the novel diagnostic. Resource requirements for XDR treatment (drugs, hospital capacity, staff, etc.) likely will also increase with increasing numbers diagnosed. Total costs will vary depending on testing volume and prevalence of XDR in the population. Budget impact will depend on current standard of care and associated resource use.	

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Direct costs related to cartridge and machinery are provided from the manufacturer while several important items related to resource use including staff time, overhead and operational costs associated with implementing Xpert MTB/XDR have not been investigated. Differences in resource use between Xpert MTB/XDR and existing approaches will vary across settings using different phenotypic and genotypic DST. Important variability exists in costs of staff time and operational costs, such as testing volume across settings.	

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations

○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies ● No included	No cost-effectiveness studies were identified using XpertMTB/XDR. Extrapolation of cost-effectiveness data from Xpert MTB/RIF or other CBNATs is not advised due to differences in diagnostic accuracy, costs associated with XDR treatment and the testing and treatment cascade of care.	

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence). Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence). Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study). New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies). (interview study) The identified challenges with **CB-NAAT** utilization and accumulated delays risk compromize the added value as identified by the users, ultimately leading to underutilization and hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence)	

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations

○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Patients can be reluctant to test for TB/MDR-TB because of stigma related to MDR-TB or related to having interrupted treatment in the past, because of fears of side effects, the failure to recognize symptoms, the inability to produce sputum and the cost, distance and travel concerns related to (repeat) clinic visits. (QES: high confidence) Health workers can be reluctant to test for TB or MDR-TB because of TB associated stigma and consequences for their patients, fears of acquiring TB, fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified, fear of side effects of drugs in children, and community awareness of disease manifestations in children. (QES: high confidence) **CB-NAAT** appears widely acceptable by laboratory staff and clinicians based on its simple user steps, familiarity of the system, and due to the amount of important information it provides. (interview study)	
Feasibility Is the intervention feasible to implement?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	**CB-NAAT** seems to decrease workload in the laboratory in terms of freeing up time for laboratory staff, but in most settings the introduction of CB-NAAT increases workload of laboratory staff if added onto existing work without adjusting staffing arrangements, or if it does not replace existing diagnostic tests with the result that staff may be hesitant to accept testing with CB-NAAT**.** (QES: moderate confidence) **The CB-NAAT requires less user training** compared to other DST methods (such as LPA and culture), **making it more feasible to implement** compared to methods with more user steps and those methods which require significant additional training (interview study). However, implementation of new diagnostics must be accompanied with training for clinicians, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge leading to underutilization (QES: high confidence and interview study) or overreliance on **CB-NAAT** results at the expense of clinical acumen (QES: moderate confidence). **Furthermore**, introduction of new diagnostics must be accompanied by **guidelines and algorithms, which support clinicians and laboratories in communicating with each other**, such that they can discuss discordant results, and interpret laboratory results in the context of drug availability, patient history, and patient progress on a current drug regimen.(Interview study). **In addition,** an **efficient sample transportation system, with sustainable funding mechanisms** is crucial for feasibility, especially if an algorithm requires multiple samples at different times, from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, the sample risks being contaminated and yielding inconclusive results on molecular diagnostics. Participants cited good personnel skill, standardized operating procedures, and significant laboratory infrastructure as essential in reducing sample contamination in their laboratory. (interview study) **Finally:** Implementation of new diagnostics must be accompanied with **training for clinicians**, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge (QES: high confidence and interview study). **BUT, Feasibility is challenged** by accumulation of diagnostic delays and/or underutilization at every step due to mainly health system factors: non-adherence to testing algorithms, testing for (MDR)-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor/delayed sample transport and sample quality, and result communication, delays in scheduling follow up	

	visits and recalling patients, inconsistent result recording; lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners, and sputum containers; dusty environment, and delayed or absent local repair option); inefficient/unclear work- and patient flows (for instance inefficient organizational processes, poor links between providers, unclear follow up mechanisms or where patients need to go); and lack of data-driven and inclusive national implementation processes. These challenges lead to delays and/or underutilization. (QES: high confidence) **Feasibility for the CB-NAAT is also challenged by** the value of diagnosing MTB over DR TB at primary care, makes it less feasible as a baseline test, though it would fit at a district or intermediate level laboratory. The identified **feasibility challenges** with **CB-NAAT** utilization and accumulated delays at every step may compromize the added value/benefits as identified by the users (avoiding delays, keeping cost lost, accurate results, drug resistant information, easing laboratory work), ultimately leading to underutilization (QES: high confidence). We can assume that if these values are not met users are less likely to find CB-NAATs acceptable.	

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	Accurate	**Very accurate**		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of evidence of the test accuracy	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of test's effects	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of management's effects	**Very low**	Low	Moderate	High			No included studies
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty of effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	**Favors the intervention**	Varies	Don't know

	Judgement						
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

Among patients with bacteriologically-confirmed pulmonary TB, **automated nucleic acid amplification tests of low-complexity** should be used on sputum for detection of **resistance to fluoroquinolones** rather based phenotypic DST (Conditional recommendation; moderate certainty of evidence for diagnostic accuracy).

Remarks: FLQ resistance testing is critical for all-oral 6-9 month DR-TB shorter regimen, and for 4-month Study 31 regimen for DS-TB.

Same judgments for RIF detected (question 30) - combine recommendations

PICO 6. Should Low complexity automated NAATs on sputum be used to diagnose ETO resistance in patients with microbiologically confirmed pulmonary TB, with detected resistance to RIF, gDST?

Population:	patients with microbiologically confirmed pulmonary TB, with detected resistance to RIF, gDST
Intervention:	Low complexity automated NAAT on sputum

Assessment

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Ethionamid resistance caused by *inhA mutations* is detected by the assay of interest. Ethionamid is an important second-line TB agent, which use is associated with high toxicity profile. That is why information on Ethionamid resistance is desirable. Currently information on Ethionamid resistance is inferred from LPA, however test of interest has potential offer more decentralized and automated solution.	

Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ○ Accurate ● Very accurate ○ Varies ○ Don't know	**Test accuracy** MTB/XDR assay on sputum Sensitivity: 0.98 (95% CI: 0.74 to 1.00) Specificity: 1.00 (95% CI: 0.83 to 1.00)	

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 20%	Prevalence 30%	Prevalence 50%		
True positives patients with ETO resistance	196 (148 to 200)	294 (223 to 300)	490 (371 to 500)	167 (1)	⨁◯◯◯ VERY LOW[a,b,c,d]
False negatives patients incorrectly classified as not having ETO resistance	4 (0 to 52)	6 (0 to 77)	10 (0 to 129)		
True negatives patients without ETO resistance	798 (668 to 800)	698 (584 to 700)	499 (418 to 500)	267 (1)	⨁◯◯◯ VERY LOW[a,b,e]
False positives patients incorrectly	2 (0 to 132)	2 (0 to 116)	1 (0 to 82)		

	classified as having ETO resistance	

a. We thought there was very serious risk of bias in the reference standard domain because the study did not include all of the loci (i.e. ethA, ethR, and inhA promoter) required for the reference standard to correctly classify the target condition. Of note, against a reference standard of pDST, the pooled sensitivity estimate was considerably lower at 51.7% (33.1 to 69.8). We downgraded two levels for risk of bias.
b. The median prevalence of ethionamide resistance in the included studies was 39.3%, range, 13.6% (FIND, New Delhi) to 61.5% (FIND, South Africa), higher than the three prevalences listed in the GRADE table. Applicability to settings with lower prevalence of ethionamide resistance comes with some uncertainty. We did not downgrade for indirectness.
c. Sensitivity estimates ranged from 78% (FIND, Moldova) to 100% (FIND, Moldova and Mumbai). The heterogeneity could in part explained by small numbers of resistant cases in Moldova and South Africa. We did not downgrade for inconsistency.
d. The 95% CI was wide. We thought the 95% CI around true positives and false negatives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.
e. We thought the 95% CI around true negatives and false positives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	Rapid extended drug resistance profiling allows for early initiation of optimised therapy and likely better patient outcomes. Amplification of drug resistance would be less likely. **For ethionamide, the drug is deprecated for use in WHO 2018 longer regimens containing bedaquiline because it does not appear to be effective in this context. When the drug is used as part of the standardised STREAM shorter regimen, resistance testing for ethionamide is not mandatory, though encouraged. Given this, there may be a smaller benefit for detecting ethionamide resistance.**	**True positive result** means rapid extended drug resistance profiling allows for early initiation of optimized therapy and likely better patient outcomes. Amplification of drug resistance would be less likely. Information on *inhA* promotor mutations could also guide high dose isoniazid therapy. **True negative result** will allow rapid exclusion of the TB diagnosis, decrease of stigma, better opportunities for diagnosis other diseases and likely better patient outcomes.Desirable effects less than for FQ - use of ETO (see comment under research evidence)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 20%	Prevalence 30%	Prevalence 50%		
True positives patients with ETO resistance	196 (148 to 200)	294 (223 to 300)	490 (371 to 500)	167 (1)	⨁◯◯◯ VERY LOW[a,b,c,d]

	False negatives patients incorrectly classified as not having ETO resistance	4 (0 to 52)	6 (0 to 77)	10 (0 to 129)		There would be 294 per 1000 true positive tests with Xpert MTB/XDR and those patients would benefit from being treated with an optimized regimen. There would be 698 per 1000 true negative tests with Xpert MTB/XDR and those patients could be maintained on the current treatment regimen and not suffer the consequences of unnecessary drug-resistant treatment.
	True negatives patients without ETO resistance	798 (668 to 800)	698 (584 to 700)	499 (418 to 500)	267 (1)	⨁◯◯◯ VERY LOW[a,b,e]
	False positives patients incorrectly classified as having ETO resistance	2 (0 to 132)	2 (0 to 116)	1 (0 to 82)		

a. We thought there was very serious risk of bias in the reference standard domain because the study did not include all of the loci (i.e. ethA, ethR, and inhA promoter) required for the reference standard to correctly classify the target condition. Of note, against a reference standard of pDST, the pooled sensitivity estimate was considerably lower at 51.7% (33.1 to 69.8). We downgraded two levels for risk of bias.
b. The median prevalence of ethionamide resistance in the included studies was 39.3%, range, 13.6% (FIND, New Delhi) to 61.5% (FIND, South Africa), higher than the three prevalences listed in the GRADE table. Applicability to settings with lower prevalence of ethionamide resistance comes with some uncertainty. We did not downgrade for indirectness.
c. Sensitivity estimates ranged from 78% (FIND, Moldova) to 100% (FIND, Moldova and Mumbai). The heterogeneity could in part explained by small numbers of resistant cases in Moldova and South Africa. We downgraded one level for inconsistency.
d. The 95% CI was wide. As we had already downgraded for inconsistency, we did not downgrade further for imprecision.
e. We thought the 95% CI around true negatives and false positives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations

○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know	There is uncertainty about test performance in patients with paucibacillary disease. The test may not detect all variants of ethionamide resistance. There is a discrepancy between genotypic and phenotypic DST. There is limited phenotypic DST availability in many settings. Hence utility in decision making is limited.	**False positive result** means unnecessary treatment, stigma, financial losses. **False negative result** would mean missed diagnosis, worse health outcomes, dissemination of TB infection. There would be 6 per 1000 false negative tests with Xpert MTB/XDR and those patients would suffer the consequences of not being treated with an optimized regimen. There would be 2 per 1000 false positive tests with Xpert MTB/XDR and those patients would suffer the consequences of unnecessary treatment for drug resistance.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 20%	Prevalence 30%	Prevalence 50%		
True positives patients with ETO resistance	196 (148 to 200)	294 (223 to 300)	490 (371 to 500)	167 (1)	⨁◯◯◯ VERY LOW[a,b,c,d]
False negatives patients incorrectly classified as not having ETO resistance	4 (0 to 52)	6 (0 to 77)	10 (0 to 129)		
True negatives patients without ETO resistance	798 (668 to 800)	698 (584 to 700)	499 (418 to 500)	267 (1)	⨁◯◯◯ VERY LOW[a,b,e]
False positives patients incorrectly classified as having ETO resistance	2 (0 to 132)	2 (0 to 116)	1 (0 to 82)		

a. We thought there was very serious risk of bias in the reference standard domain because the study did not include all of the loci (i.e. ethA, ethR, and inhA promoter) required for the reference standard to correctly classify the target condition. Of note, against a reference standard of pDST, the pooled sensitivity estimate was considerably lower at 51.7% (33.1 to 69.8). We downgraded two levels for risk of bias.

b. The median prevalence of ethionamide resistance in the included studies was 39.3%, range, 13.6% (FIND, New Delhi) to 61.5% (FIND, South Africa), higher than the three prevalences listed in the GRADE table. Applicability to settings with lower prevalence of ethionamide resistance comes with some uncertainty. We did not downgrade for indirectness.

	c. Sensitivity estimates ranged from 78% (FIND, Moldova) to 100% (FIND, Moldova and Mumbai). The heterogeneity could in part explained by small numbers of resistant cases in Moldova and South Africa. We downgraded one level for inconsistency. d. The 95% CI was wide. As we had already downgraded for inconsistency, we did not downgrade further for imprecision. e. We thought the 95% CI around true negatives and false positives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.	
Certainty of the evidence of test accuracy What is the overall certainty of the evidence of test accuracy?		
Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	The overall certainty of the evidence was very low owing to serious inconsistency for sensitivity and very serious risk of bias and serious imprecision for specificity. Detailed judgments are provided in the evidence profile.	
Certainty of the evidence of test's effects What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Although a diagnostic study may not capture adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.	to be reviewed
Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?		
Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	There are no randomized trials on the effect on patient-important outcomes of using the test. A positive result for resistance would mean modification of the treatment regimen, and a negative result would mean preserving ETO in the treatment regimen.	
Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
Judgement	Research evidence	Additional considerations

○ Very low ○ Low ○ Moderate ○ High ● No included studies	Observations from clinical practice suggest that clinicians will make decisions based on test results and individualise the regimen using them.	TB programmes and clinicians are familiar with Xpert testing. While we expect clinicians to have high confidence in Xpert MTB/XDR results, the challenges with feasibility and the resources required mean that clinicians may not be able to order Xpert MTB/XDR testing in some settings.
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ○ No included studies	This is the summary of the preceding points 5-8	very low certainty in accuracy
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me"), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence).	Compared to existing tests/sputum microscopy, **healthcare professionals** appreciate the rapidity of CB-NAAT results, the accuracy of **CB-NAAT** results and the confidence that this generates to start treating and motivate patients, the diversity of sample types, the ability to detect drug resistance (earlier or at all, for as many drugs as possible and altering clinician's risk perception of drug resistance in children), as well as the consequence of avoiding costlier investigations or hospital stays when using CB-NAAT. (QES: high confidence). The cartridge has a quicker turnaround time for first and second line drug susceptibility testing, compared to other available diagnostic methods. People value faster TAT, the potential ability to reflex

		samples from the Xpert MTB/RIF to the Xpert MTB/XDR cartridge, and receiving information on multiple drugs as well as high or low level resistance simultaneously, as it could enable quicker diagnosis and optimized treatment for patients. (Interview study)**Laboratory technicians** appreciate the improvement of overall laboratory work that **CB-NAAT** brings compared to sputum microscopy in terms of ease of use, ergonomics, and biosafety (QES: high confidence). It requires minimal user steps and the GeneXpert platform is a familiar system which people feel comfortable running and interpreting (Interview study). **Laboratory managers** appreciate that monitoring of laboratory work and training is easier than with sputum microscopy and that **CB-NAAT** eases staff retention, as it increases staff satisfaction and has a symbolic meaning of progress within the TB world (QES: low confidence)
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the interventio n ○ Varies ○ Don't know		**The reference standard is genotypic DST (the comparator)** Clinical benefit has not been evaluated here. Clinical benefit would be superior in terms of speed of treatment. in some settings the comparator Desirable outweight undesirable effects but there is uncertainty of the evidence which did not make all members of the panel

		confident that there is more benefit. The comparator is genotypic DST - same answer with the same speed The panel discussed that NAAT ETO is provided as together with other resistance data and leads to targetted regimens faster. genotypic DST not available for routine clinial use. This leads to diagnostic delays and the additional related concerns by patients (increased anxiety) The accuracy data for the comparison against genotypic DST come with the concern about the imperfect reference standard The toxicity of ETO and knowing about resistance to it helps to drop it from a regimen if it is included and resistance is present.
Resources required How large are the resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	No direct evidence from published studies regarding total resources required. Resource requirements will include the purchase of cartridges ($19.80USD/cartridge), upgrading of existing platforms to 10-colour modules (an upgrade that will be required eventually for all Xpert platforms: $3860 to >$72,350) and operational and programmatic costs associated with implementing the novel diagnostic. Resource requirements for XDR treatment (drugs, hospital capacity, staff, etc.) likely will also increase with increasing numbers diagnosed. Total costs will vary depending on testing volume and prevalence of XDR in the population. Budget impact will depend on current standard of care and associated resource use.	
Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations

○ Very low ○ Low ○ Moderate ○ High ● No included studies	Direct costs related to cartridge and machinery are provided from the manufacturer while several important items related to resource use including staff time, overhead and operational costs associated with implementing Xpert MTB/XDR have not been investigated. Differences in resource use between Xpert MTB/XDR and existing approaches will vary across settings using different phenotypic and genotypic DST. Important variability exists in costs of staff time and operational costs, such as testing volume across settings.	

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies	No cost-effectiveness studies were identified using XpertMTB/XDR. Extrapolation of cost-effectiveness data from Xpert MTB/RIF or other CBNATs is not advised due to differences in diagnostic accuracy, costs associated with XDR treatment and the testing and treatment cascade of care.	

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence). Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence). Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study). New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the	

	rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies). (interview study) The identified challenges with **CB-NAAT** utilization and accumulated delays risk compromize the added value as identified by the users, ultimately leading to underutilization and hamper access to prompt and accurate testing and treatment particularly for vulnerable groups**.** (QES: High confidence)	
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Patients can be reluctant to test for TB/MDR-TB because of stigma related to MDR-TB or related to having interrupted treatment in the past, because of fears of side effects, the failure to recognize symptoms, the inability to produce sputum and the cost, distance and travel concerns related to (repeat) clinic visits. (QES: high confidence) Health workers can be reluctant to test for TB or MDR-TB because of TB associated stigma and consequences for their patients, fears of acquiring TB, fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified, fear of side effects of drugs in children, and community awareness of disease manifestations in children. (QES: high confidence) **CB-NAAT** appears widely acceptable by laboratory staff and clinicians based on its simple user steps, familiarity of the system, and due to the amount of important information it provides. (interview study)	
Feasibility Is the intervention feasible to implement?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	**CB-NAAT** seems to decrease workload in the laboratory in terms of freeing up time for laboratory staff, but in most settings the introduction of CB-NAAT increases workload of laboratory staff if added onto existing work without adjusting staffing arrangements, or if it does not replace existing diagnostic tests with the result that staff may be hesitant to accept testing with CB-NAAT**.** (QES: moderate confidence) **The CB-NAAT requires less user training** compared to other DST methods (such as LPA and culture), **making it more feasible to implement** compared to methods with more user steps and those methods which require significant additional training (interview study). However, implementation of new diagnostics must be accompanied with training for clinicians, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge leading to underutilization (QES: high confidence and interview study) or overreliance on **CB-NAAT** results at the expense of clinical acumen (QES: moderate confidence). **Furthermore**, introduction of new diagnostics must be accompanied by **guidelines and algorithms, which support clinicians and laboratories in communicating with each other**, such that they can discuss discordant results, and interpret laboratory	

	results in the context of drug availability, patient history, and patient progress on a current drug regimen.(Interview study). **In addition,** an **efficient sample transportation system, with sustainable funding mechanisms** is crucial for feasibility, especially if an algorithm requires multiple samples at different times, from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, the sample risks being contaminated and yielding inconclusive results on molecular diagnostics. Participants cited good personnel skill, standardized operating procedures, and significant laboratory infrastructure as essential in reducing sample contamination in their laboratory. (interview study) **Finally:** Implementation of new diagnostics must be accompanied with **training for clinicians**, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge (QES: high confidence and interview study). **BUT, Feasibility is challenged** by accumulation of diagnostic delays and/or underutilization at every step due to mainly health system factors: non-adherence to testing algorithms, testing for (MDR)-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor/delayed sample transport and sample quality, and result communication, delays in scheduling follow up visits and recalling patients, inconsistent result recording; lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners, and sputum containers; dusty environment, and delayed or absent local repair option); inefficient/unclear work- and patient flows (for instance inefficient organizational processes, poor links between providers, unclear follow up mechanisms or where patients need to go); and lack of data-driven and inclusive national implementation processes. These challenges lead to delays and/or underutilization. (QES: high confidence) **Feasibility for the CB-NAAT is also challenged by** the value of diagnosing MTB over DR TB at primary care, makes it less feasible as a baseline test, though it would fit at a district or intermediate level laboratory. The identified **feasibility challenges** with **CB-NAAT** utilization and accumulated delays at every step may compromize the added value/benefits as identified by the users (avoiding delays, keeping cost lost, accurate results, drug resistant information, easing laboratory work), ultimately leading to underutilization (QES: high confidence). We can assume that if these values are not met users are less likely to find CB-NAATs acceptable.	

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	Accurate	**Very accurate**		Varies	Don't know
Desirable effects	Trivial	Small	**Moderate**	Large		Varies	Don't know
Undesirable effects	Large	**Moderate**	Small	Trivial		Varies	Don't know
Certainty of evidence of the test accuracy	**Very low**	Low	Moderate	High			No included studies

	Judgement						
Certainty of evidence of test's effects	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of management's effects	**Very low**	Low	Moderate	High			No included studies
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty of effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the interventio n	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In patients with bacteriologically-confirmed pulmonary TB and resistance to rifampicin, automated nucleic acid amplification tests of low-complexity may be used on sputum for detection of resistance to ethionamide (rather than genotypic sequencing for InhA) (Conditional recommendation; very low certainty of evidence for diagnostic accuracy)

Remarks: make remark about class based recommendations (here and elsewhere)

Implementation considerations

Comment about high specificity and use as rule in test.

PICO 7. Should Low complexity automated NAATs on sputum be used to diagnose AMK resistance in patients with microbiologically confirmed pulmonary TB, with detected resistance to RIF, MRS?

Population:	patients with microbiologically confirmed pulmonary TB, with detected resistance to RIF, MRS
Intervention:	Low complexity automated NAAT on sputum
Purpose of the test:	Anti-TB drugs resistance detection
Setting:	In/out-patient
Subgroups:	Children, PLHIV, patients with EP TB

Assessment

Problem

Is the problem a priority?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Amikacin is an injectable second-line TB agent, which use is associated with high toxicity profile and often patients discomfort. Yet its use is recommended for adults patients in specific situations when susceptibility has been demonstrated and adequate measures to monitor for adverse reactions can be ensured (WHO Module 4). That is why information on Amikacin resistance is desirable. Currently information on Amikacin resistance is provided by LPA, however test of interest has potential to offer more decentralized and automated solution.	

Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies	**Test accuracy** MTB/XDR assay on sputum Sensitivity: 0.86 (95% CI: 0.75 to 0.93) Specificity: 0.99 (95% CI: 0.93 to 1.00)	

○ Don't know

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 20%	Prevalence 30%	Prevalence 50%		
True positives patients with AMK resistance	172 (150 to 185)	258 (225 to 278)	431 (375 to 464)	65 (1)	⨁⨁◯◯ LOW[a,b,c]
False negatives patients incorrectly classified as not having AMK resistance	28 (15 to 50)	42 (22 to 75)	69 (36 to 125)		
True negatives patients without AMK resistance	791 (744 to 798)	692 (651 to 699)	495 (465 to 499)	425 (1)	⨁⨁⨁⨁ HIGH[a]
False positives patients incorrectly classified as having AMK resistance	9 (2 to 56)	8 (1 to 49)	5 (1 to 35)		

a. The median prevalence of amikacin resistance in the studies was 13.5%, range 5.7% (FIND, Moldova) to 36.1% (FIND, South Africa), lower than the prevalences listed in the table. Applicability to settings with higher prevalence of amikacin resistance comes with some uncertainty. We did not downgrade for indirectness.

b. Sensitivity estimates were somewhat inconsistent, ranging from 75% (FIND, New Delhi) to 95% (FIND, South Africa). Regarding the finding of low amikacin sensitivity estimates in the FIND study, the authors provided the following explanation. "This issue appears to be linked exclusively to samples with rrs c1402a and g1484t double mutations (12 in New Delhi, 3 in Moldova). The g1484t mutation was considered to be a marker of phenotypic amikacin resistance in the FIND analysis, but 14/15 of these mutated samples were pDST AMK-S (1 was pDST contaminated). Importantly, all of these pDST AMK-S/WGS AMK-R samples with the mutations noted above tested susceptible by Hain LPA as well as Xpert XDR, so we have more confidence in the Xpert (rather than WGS) result." We also note New Delhi had a small number of resistant cases. These explanations may in part explain the heterogeneity in sensitivity estimates. We did not downgrade for inconsistency. This was a judgement.

c. Although the 95% CI is wide, we thought that this was owing to heterogeneity (see explanation in Inconsistency domain). There was a very low number of participants with amikacin resistance contributing to this analysis for the observed sensitivity. We downgraded two levels for imprecision.

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement

○ Trivial
○ Small
○ Moderate
● Large
○ Varies
○ Don't know

Research evidence

Rapid extended drug resistance profiling allows for early initiation of optimised therapy and likely better patient outcomes. Amplification of drug resistance would be less likely.

It is helpful to know if amikacin can be used when newer all-oral RR/MDRTB regimens are not available, or the patient cannot be adequately treated by an all-oral regimen.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 6%	Prevalence 13%	Prevalence 20%		
True positives patients with AMK resistance	52 (45 to 56)	116 (101 to 125)	172 (150 to 185)	65 (1)	⨁⨁◯◯ LOW[a,b,c]
False negatives patients incorrectly classified as not having AMK resistance	8 (4 to 15)	19 (10 to 34)	28 (15 to 50)		
True negatives patients without AMK resistance	930 (874 to 938)	855 (804 to 863)	791 (744 to 798)	425 (1)	⨁⨁⨁⨁ HIGH[a]
False positives patients incorrectly classified as having	10 (2 to 66)	10 (2 to 61)	9 (2 to 56)		

Additional considerations

True positive result means rapid extended drug resistance profiling allows for early initiation of optimized therapy and likely better patient outcomes. Amplification of drug resistance would be less likely. Information on i*nhA* promotor mutations could also guide high dose isoniazid therapy.

True negative result will allow rapid exclusion of the TB diagnosis, decrease of stigma, better opportunities for diagnosis other diseases and likely better patient outcomes.Desirable effects less than for FQ - use of ETO (see comment under research evidence)

AMK resistance						

d. The median prevalence of amikacin resistance in the studies was 13.5%, range 5.7% (FIND, Moldova) to 36.1% (FIND, South Africa), lower than the prevalences listed in the table. Applicability to settings with higher prevalence of amikacin resistance comes with some uncertainty. We did not downgrade for indirectness.

e. Sensitivity estimates were somewhat inconsistent, ranging from 75% (FIND, New Delhi) to 95% (FIND, South Africa). Regarding the finding of low amikacin sensitivity estimates in the FIND study, the authors provided the following explanation. "This issue appears to be linked exclusively to samples with rrs c1402a and g1484t double mutations (12 in New Delhi, 3 in Moldova). The g1484t mutation was considered to be a marker of phenotypic amikacin resistance in the FIND analysis, but 14/15 of these mutated samples were pDST AMK-S (1 was pDST contaminated). Importantly, all of these pDST AMK-S/WGS AMK-R samples with the mutations noted above tested susceptible by Hain LPA as well as Xpert XDR, so we have more confidence in the Xpert (rather than WGS) result." We also note New Delhi had a small number of resistant cases. These explanations may in part explain the heterogeneity in sensitivity estimates. We did not downgrade for inconsistency. This was a judgement.

a. The 95% CI was wide. We thought the 95% CI around true positives and false negatives would likely lead to different decisions depending on which confidence limits are assumed. Also, there was a very low number of participants with amikacin resistance contributing to this analysis for the observed sensitivity. We downgraded two levels for imprecision.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	There is uncertainty about test performance in patients with paucibacillary disease. With the adoption of the new treatment regimens using all-oral medicines, the second-line injectable drugs are less relevant. Amikacin is identified as the preferred injectable and now regarded as a WHO category C (less important) drug for RR/MDR-TB treatment.	**False positive result** means unnecessary treatment, stigma, financial losses. **False negative result** would mean missed diagnosis, worse health outcomes, dissemination of TB infection. Toxicity of AMK even if a few false positives were treated may be substanital if the drug is used.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 6%	Prevalence 13%	Prevalence 20%		
True positives patients with AMK resistance	52 (45 to 56)	116 (101 to 125)	172 (150 to 185)	65 (1)	⨁⨁◯◯ LOW[a,b,c]

False negatives patients incorrectly classified as not having AMK resistance	8 (4 to 15)	19 (10 to 34)	28 (15 to 50)		
True negatives patients without AMK resistance	930 (874 to 938)	855 (804 to 863)	791 (744 to 798)	425 (1)	⨁⨁⨁⨁ HIGH[a]
False positives patients incorrectly classified as having AMK resistance	10 (2 to 66)	10 (2 to 61)	9 (2 to 56)		

a. The median prevalence of amikacin resistance in the studies was 13.5%, range 5.7% (FIND, Moldova) to 36.1% (FIND, South Africa), lower than the prevalences listed in the table. Applicability to settings with higher prevalence of amikacin resistance comes with some uncertainty. We did not downgrade for indirectness.

b. Sensitivity estimates were somewhat inconsistent, ranging from 75% (FIND, New Delhi) to 95% (FIND, South Africa). Regarding the finding of low amikacin sensitivity estimates in the FIND study, the authors provided the following explanation. "This issue appears to be linked exclusively to samples with rrs c1402a and g1484t double mutations (12 in New Delhi, 3 in Moldova). The g1484t mutation was considered to be a marker of phenotypic amikacin resistance in the FIND analysis, but 14/15 of these mutated samples were pDST AMK-S (1 was pDST contaminated). Importantly, all of these pDST AMK-S/WGS AMK-R samples with the mutations noted above tested susceptible by Hain LPA as well as Xpert XDR, so we have more confidence in the Xpert (rather than WGS) result." We also note New Delhi had a small number of resistant cases. These explanations may in part explain the heterogeneity in sensitivity estimates. We did not downgrade for inconsistency. This was a judgement.

c. The 95% CI was wide. We thought the 95% CI around true positives and false negatives would likely lead to different decisions depending on which confidence limits are assumed. Also, there was a very low number of participants with amikacin resistance contributing to this analysis for the observed sensitivity. We downgraded two levels for imprecision.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations

○ Very low ● Low ○ Moderate ○ High ○ No included studies	The certainty of the evidence was low owing to very serious imprecision for sensitivity. Detailed judgments are provided in the evidence profile.	
Certainty of the evidence of test's effects What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Although a diagnostic study may not capture adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.	
Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?		
Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	There are no randomized trials on the effect on patient-important outcomes of using the test. A positive result for resistance would mean modification of the treatment regimen, and a negative result would mean preserving AMK in the treatment regimen	
Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Observations from clinical practice suggest that clinicians will make decisions based on test results and individualise the regimen using them.	TB programmes and clinicians are familiar with Xpert testing. While we expect clinicians to have high confidence in Xpert MTB/XDR results, the challenges with feasibility and the resources required mean that clinicians may not be able to order Xpert MTB/XDR testing in some settings.
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
Judgement	Research evidence	Additional considerations

○ Very low ○ Low ○ Moderate ○ High ○ No included studies	This is the summary of the preceding points 5-8	Low certainty in accuracy

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing "what is wrong with me"), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence).	Compared to existing tests/sputum microscopy, **healthcare professionals** appreciate the rapidity of CB-NAAT results, the accuracy of **CB-NAAT** results and the confidence that this generates to start treating and motivate patients, the diversity of sample types, the ability to detect drug resistance (earlier or at all, for as many drugs as possible and altering clinician's risk perception of drug resistance in children), as well as the consequence of avoiding costlier investigations or hospital stays when using CB-NAAT. (QES: high confidence). The cartridge has a quicker turnaround time for first and second line drug susceptibility testing, compared to other available diagnostic methods. People value faster TAT, the potential ability to reflex samples from the Xpert MTB/RIF to the Xpert MTB/XDR cartridge, and receiving information on multiple drugs as well as high or low level resistance simultaneously, as it could enable quicker diagnosis and optimized treatment for patients. (Interview study)**Laboratory technicians** appreciate the improvement of overall laboratory work that **CB-NAAT** brings compared to sputum microscopy in terms of ease of use, ergonomics, and biosafety (QES: high confidence). It requires minimal user steps and the GeneXpert platform is a familiar system which people

		feel comfortable running and interpreting (Interview study). **Laboratory managers** appreciate that monitoring of laboratory work and training is easier than with sputum microscopy and that **CB-NAAT** eases staff retention, as it increases staff
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ● Favors the intervention ○ Varies ○ Don't know		**The reference standard is phenotypic DST (the comparator)** Clinical benefit has not been evaluated here. Clinical benefit would be superior in terms of speed of treatment. in some settings the comparator Desirable outweight undesirable effects but there is uncertainty of the evidence which did not make all members of the panel confident that there is more benefit.
Resources required How large are the resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	No direct evidence from published studies regarding total resources required. Resource requirements will include the purchase of cartridges ($19.80USD/cartridge), upgrading of existing platforms to 10-colour modules (an upgrade that will be required eventually for all Xpert platforms: $3860 to >$72,350) and operational and programmatic costs associated with implementing the novel diagnostic. Resource requirements for XDR treatment (drugs, hospital capacity, staff, etc.) likely will also increase with increasing numbers diagnosed. Total costs will vary depending on testing volume and prevalence of XDR in the population. Budget impact will depend on current standard of care and associated resource use.	

Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Direct costs related to cartridge and machinery are provided from the manufacturer while several important items related to resource use including staff time, overhead and operational costs associated with implementing Xpert MTB/XDR have not been investigated. Differences in resource use between Xpert MTB/XDR and existing approaches will vary across settings using different phenotypic and genotypic DST. Important variability exists in costs of staff time and operational costs, such as testing volume across settings.	
Cost effectiveness Does the cost-effectiveness of the intervention favor the intervention or the comparison?		
Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies	No cost-effectiveness studies were identified using XpertMTB/XDR. Extrapolation of cost-effectiveness data from Xpert MTB/RIF or other CBNATs is not advised due to differences in diagnostic accuracy, costs associated with XDR treatment and the testing and treatment cascade of care.	
Equity What would be the impact on health equity?		
Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence). Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence).	

	Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study). New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies). (interview study) The identified challenges with **CB-NAAT** utilization and accumulated delays risk compromize the added value as identified by the users, ultimately leading to underutilization and hamper access to prompt and accurate testing and treatment particularly for vulnerable groups**.** (QES: High confidence)	
Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Patients can be reluctant to test for TB/MDR-TB because of stigma related to MDR-TB or related to having interrupted treatment in the past, because of fears of side effects, the failure to recognize symptoms, the inability to produce sputum and the cost, distance and travel concerns related to (repeat) clinic visits. (QES: high confidence) Health workers can be reluctant to test for TB or MDR-TB because of TB associated stigma and consequences for their patients, fears of acquiring TB, fear from supervisors when reclassifying patients already on TB treatment who turn out to be misclassified, fear of side effects of drugs in children, and community awareness of disease manifestations in children. (QES: high confidence) **CB-NAAT** appears widely acceptable by laboratory staff and clinicians based on its simple user steps, familiarity of the system, and due to the amount of important information it provides. (interview study)	
Feasibility Is the intervention feasible to implement?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	**CB-NAAT** seems to decrease workload in the laboratory in terms of freeing up time for laboratory staff, but in most settings the introduction of CB-NAAT increases workload of laboratory staff if added onto existing work without adjusting staffing arrangements, or if it does not replace existing diagnostic tests with the result that staff may be hesitant to accept testing with CB-NAAT**.** (QES: moderate confidence) **The CB-NAAT requires less user training** compared to other DST methods (such as LPA and culture), **making it more feasible to implement** compared to methods with more user steps and those methods which require significant additional	

	training (interview study). However, implementation of new diagnostics must be accompanied with training for clinicians, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge leading to underutilization (QES: high confidence and interview study) or overreliance on **CB-NAAT** results at the expense of clinical acumen (QES: moderate confidence). **Furthermore**, introduction of new diagnostics must be accompanied by **guidelines and algorithms, which support clinicians and laboratories in communicating with each other**, such that they can discuss discordant results, and interpret laboratory results in the context of drug availability, patient history, and patient progress on a current drug regimen.(Interview study). **In addition,** an **efficient sample transportation system, with sustainable funding mechanisms** is crucial for feasibility, especially if an algorithm requires multiple samples at different times, from different collection points, as is the case when dealing with DR-TB. If mishandled during preparation, the sample risks being contaminated and yielding inconclusive results on molecular diagnostics. Participants cited good personnel skill, standardized operating procedures, and significant laboratory infrastructure as essential in reducing sample contamination in their laboratory. (interview study) **Finally:** Implementation of new diagnostics must be accompanied with **training for clinicians**, to help them interpret results from new molecular tests and understand how this relates to treatment of a patient. In the past, with introduction of CB-NAAT this has been a challenge (QES: high confidence and interview study). **BUT, Feasibility is challenged** by accumulation of diagnostic delays and/or underutilization at every step due to mainly health system factors: non-adherence to testing algorithms, testing for (MDR)-TB late in the process, empirical treatment, false negatives due to technology failure, large sample volumes and staff shortages, poor/delayed sample transport and sample quality, and result communication, delays in scheduling follow up visits and recalling patients, inconsistent result recording; lack of sufficient resources and maintenance (i.e. stock-outs; unreliable logistics; lack of funding, electricity, space, air conditioners, and sputum containers; dusty environment, and delayed or absent local repair option); inefficient/unclear work- and patient flows (for instance inefficient organizational processes, poor links between providers, unclear follow up mechanisms or where patients need to go); and lack of data-driven and inclusive national implementation processes. These challenges lead to delays and/or underutilization. (QES: high confidence) **Feasibility for the CB-NAAT is also challenged by** the value of diagnosing MTB over DR TB at primary care, makes it less feasible as a baseline test, though it would fit at a district or intermediate level laboratory. The identified **feasibility challenges** with **CB-NAAT** utilization and accumulated delays at every step may compromize the added value/benefits as identified by the users (avoiding delays, keeping cost lost, accurate results, drug resistant information, easing laboratory work), ultimately leading to underutilization (QES: high confidence). We can assume that if these values are not met users are less likely to find CB-NAATs acceptable.	

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know

	Judgement						
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of evidence of the test accuracy	Very low	**Low**	Moderate	High			No included studies
Certainty of evidence of test's effects	Very low	Low	**Moderate**	High			No included studies
Certainty of evidence of management's effects	**Very low**	Low	Moderate	High			No included studies
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty of effects	Very low	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	**Favors the intervention**	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention

○	○	○	●	○

Conclusions

Recommendation

In patients with bacteriologically-confirmed pulmonary TB **and resistance to rifampicin,** automated nucleic acid amplification tests of low-complexity may be used on sputum for detection of **resistance to amikacin,** rather than culture-based phenotypic DST (Conditional recommendation; low certainty of evidence for diagnostic accuracy);

Implementation considerations

drug toxicity monitoring and management

3.7 Evidence-to-decision tables: First-line line probe assay (FL-LPA)

PICO 1. Accuracy of LPAs for detecting rifampicin resistance by direct testing in sputum smear-positive TB patients compared with phenotypic culture-based DST

	Judgement	Research evidence	Additional considerations
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Currently, only 26% of an estimated 480 000 cases of MDR-TB are diagnosed, and often a diagnosis of MDR-TB comes too late. This is in large part due to a lack of access to accurate and rapid diagnostics. LPAs are able to detect *Mycobacterium tuberculosis* and resistance to rifampicin and isoniazid. LPAs normally take at least 1 working day to perform and require a controlled laboratory infrastructure.	
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* LPA for direct testing compared with phenotypic DST Sensitivity: 0.96 (95% CI: 0.95–0.97); specificity: 0.98 (95% CI: 0.97–0.99)	
Desirable	**How substantial are the desirable anticipated effects?**		The decrease in the time to results is a critical reason for the large benefits.

	○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	**Test result**	**Number of results per 1 000 patients tested (95% CI)**		**Number of participants (number of studies)**	**Quality of the evidence (GRADE)**	LPA results are more likely to be interpretable compared with results from culture-based DST. Benefits are greater when direct LPA is compared with indirect.
			5% prevalence	15% prevalence			
Undesirable effects	**How substantial are the undesirable anticipated effects?** ○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	**True positives** (patients with rifampicin resistance)	48 (47–49)	144 (142–146)	2 876 (48)	⊕⊕⊕◯ MODERATE	The toxic effects of anti-TB agents on patients who are false positive by LPA are of concern. When a composite reference standard is used, some of the false positives may become true positives, thus improving sensitivity.
		False negatives (patients incorrectly classified as not having rifampicin resistance)	2 (1–3)	6 (4–8)			
		True negatives (patients without rifampicin resistance)	933 (923–939)	835 (826–840)	7 684 (48)	⊕⊕⊕◯ MODERATE	
		False positives (patients incorrectly classified as having rifampicin resistance)	17 (11–27)	15 (10–24)			
Certainty of the evidence of tests accuracy	**What is the overall certainty of the evidence of the test's accuracy?** ○ Very low ○ Low ● Moderate ○ High ○ No included studies	The risk of bias was considered to be serious for all studies. Indirectness was considered not to be serious. Inconsistency was considered not to be serious. Imprecision was considered not to be serious. Publication bias: none.					
Certainty of the evidence	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or**	No studies were included.					

	burdens of the test? ○ Very low ○ Low ○ Moderate ○ High ● No included studies		
Certainty of the evidence of management effects	**What is the overall certainty of the evidence of effects of the management that is guided by the test results?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	In theory, test results should guide management decisions, provided that the use of the test is adopted as national policy. Given the high accuracy of LPAs, a positive test result should be sufficient to start treating a patient. There are insufficient data about how the test performs in smear-negative samples.	
Certainty of the evidence of test result/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Although this systematic review was not designed to evaluate the clinical impact of LPAs, it was noted that 12 studies attempted to measure the impact of LPAs on clinical impacts, such as turnaround time and cost. For turnaround time, most studies reported the time from a positive culture result to LPA results, with results varying from 8 hours to 5 days and most reporting 1 to 2 days. This was faster than phenotypic DST with liquid cultures, which typically took 9 to 25 days, and solid cultures, which took more than 30 days. One systematic review focused on reductions in diagnostic and treatment delays. The analysis showed that using LPAs reduced diagnostic delays by an average of 47 days (95% CI: 29–64) compared with culture.	
Certainty of effects	**What is the overall certainty of the evidence of the effects of the test?** ○ Very low ○ Low ● Moderate ○ High ○ No included studies	This question is intended to summarize information from the previous four questions about the certainty of the evidence.	
Values	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or	There is no important uncertainty or variability in how people value the main outcomes. **For detecting rifampicin resistance:** LPAs have demonstrated good diagnostic accuracy when compared with both the phenotypic as well as the composite reference standard.	

	variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability ○ No known undesirable outcomes		
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ● Favours the intervention ○ Varies ○ Don't know	LPAs' good performance in sensitivity and specificity for detecting rifampicin resistance indicates that they are accurate tests, with small numbers of false-negative and false-positive results. Reductions in diagnostic and treatment delays have been documented.	
Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	Cost and cost–effectiveness studies were not assessed. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance. Based on a cost–effectiveness study done in 2011, LPAs are cost–effective compared with conventional DST	

Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Cost and cost–effectiveness studies were not assessed for this guideline. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance.	
Cost-effectiveness	**Does the cost–effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	Cost and cost–effectiveness studies were not assessed for this guideline. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Because more patients would have access to the test, health equity may be positively affected.	
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes	The test may be acceptable for implementation in settings with a high prevalence of MDR-TB. Implementing the test requires additional human resources, as it is labour intensive, as well as additional infrastructure (three separate rooms) and increased biosafety standards. For patients, the burdens and adverse effects are potentially insignificant.	

	○ Varies ○ Don't know		
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	In 2008, WHO recommended using this test to diagnose rifampicin-resistant TB in AFB-positive smears and cultures. During the Guideline Development Group meeting there was some disagreement about how feasible it would be to implement LPAs. A sophisticated laboratory infrastructure and skilled staff are required to perform the test, which are usually available at the intermediate- and reference-levels of laboratory networks. Hence, implementing the test would require additional funding and technical support to train staff and procure equipment. Quality assurance strategies will be needed as well.	

AFB: acid-fast bacilli; CI: confidence interval; DST: drug-susceptibility testing; GRADE: Grading of Recommendations Assessment, Development and Evaluation; LPA: line probe assay; MDR-TB: multidrug-resistant TB.

Summary of judgements

	Judgement						**Implications**
Problem	No	Probably no	Probably yes	**Yes**	Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate	Varies	Don't know	
Desirable effects	Trivial	Small	Moderate	**Large**	Varies	Don't know	
Undesirable effects	Large	Moderate	**Small**	Trivial	Varies	Don't know	
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High	No included studies		

	Judgement							Implications
Certainty of the evidence of test effects	Very low	Low	Moderate	High		**No included studies**		
Certainty of the evidence of management's effects	Very low	Low	Moderate	High		**No included studies**		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High		**No included studies**		
Certainty of effects	Very low	Low	Moderate	High		**No included studies**		
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**		No known undesirable outcomes		
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	**Favours the intervention**	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High		**No included studies**		

	Judgement							Implications
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know	

Conclusions

Should LPA by direct testing (compared with phenotypic DST) be used to detect rifampicin resistance in pulmonary TB?

Type of recommendation	Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ●	Strong recommendation for the intervention ○
Recommendation	**For patients with smear-positive TB, the WHO guideline panel suggests using direct LPA for the detection of rifampicin resistance instead of phenotypic DST (conditional recommendation, moderate certainty in the evidence for test accuracy).**				
Justification	There is uncertainty about the impact on cost. Feasibility concerns are moderated by international roll-out of LPA but cannot be ignored; patients who have rifampicin monoresistance by LPA should still have specimens cultured.				
Implementation considerations	Positive results should be interpreted with caution in settings with a very low prevalence of rifampicin resistance; such results possibly require confirmation and repeat testing, but therapy should not be delayed. Implementation should be phased-in gradually along with biosafety upgrades, starting at reference-level laboratories. Facilities requirements must be met (three separate rooms); there must be adequate supplies; and quality assurance strategies must be implemented, as well as reporting mechanisms. Staff training and internal laboratory procedures may need to be revised and changes should be implemented as necessary. Clinicians will need aids for interpreting results.				

Research priorities	Priorities for research include direct clinical trials to assess the impact on patient outcomes of knowing isoniazid-resistance status.

PICO 2. Accuracy of LPAs for detecting rifampicin resistance by indirect testing of *Mycobacterium tuberculosis* complex culture isolates compared with phenotypic culture- based DST

<table>
<tr><th></th><th>Judgement</th><th>Research evidence</th><th>Additional considerations</th></tr>
<tr><td>Problem</td><td>Is the problem a priority?

○ No
○ Probably no
○ Probably yes
● Yes
○ Varies
○ Don't know</td><td>Currently, only 26% of an estimated 480 000 cases of MDR-TB are diagnosed, and often a diagnosis of MDR-TB comes too late. This is in large part due to a lack of access to accurate and rapid diagnostics. LPAs are able to detect Mycobacterium tuberculosis and resistance to rifampicin and isoniazid . LPAs normally take at least 1 working day to perform and require a controlled laboratory infrastructure.</td><td></td></tr>
<tr><td>Test accuracy</td><td>How accurate is the test?

○ Very inaccurate
○ Inaccurate
● Accurate
○ Very accurate
○ Varies
○ Don't know</td><td>Test accuracy
LPA for indirect testing compared with phenotypic DST

Sensitivity: 0.97 (95% CI: 0.95–0.98); specificity: 0.99 (95% CI: 0.99–1.00)</td><td></td></tr>
<tr><td>Desirable effects</td><td>How substantial are the desirable anticipated effects?

○ Trivial
○ Small
○ Moderate
● Large
○ Varies
○ Don't know</td><td>
<table>
<tr><th rowspan="2">Test result</th><th colspan="2">Number of results per 1 000 patients tested (95% CI)</th><th rowspan="2">Number of participants (number of studies)</th><th rowspan="2">Quality of the evidence (GRADE)</th></tr>
<tr><th>5% prevalence</th><th>15% prevalence</th></tr>
<tr><td>True positives (patients with rifampicin resistance)</td><td>48 (48–49)</td><td>145 (143–147)</td><td rowspan="2">3 913 (43)</td><td rowspan="2">⊕⊕⊕◯
MODERATE</td></tr>
<tr><td>False negatives</td><td>2 (1–2)</td><td>5 (3–7)</td></tr>
</table>
</td><td>The decrease in the time to results is a critical reason for the large benefits.

The time gained depends on the medium used: LPA takes at least 3 weeks less than solid culture and 1 week less than liquid culture.

LPA results are more likely to be interpretable compared with results from culture-based DST. Benefits are greater when direct LPA is</td></tr>
</table>

<table>
<tr><td></td><td></td><td>
<table>
<tr><td>(patients incorrectly classified as not having rifampicin resistance)</td><td></td><td></td><td></td><td></td></tr>
<tr><td>True negatives (patients without rifampicin resistance)</td><td>943 (937–946)</td><td>844 (838–847)</td><td>6 783 (483)</td><td rowspan="2">⊕⊕⊕◯
MODERATE</td></tr>
<tr><td>False positives (patients incorrectly classified as having rifampicin resistance)</td><td>7 (4–13)</td><td>6 (3–12)</td></tr>
</table>
</td><td>compared with indirect.</td></tr>
<tr><td>Undesirable effects</td><td>How substantial are the undesirable anticipated effects?

○ Large
○ Moderate
● Small
○ Trivial
○ Varies
○ Don't know</td><td></td><td>The toxic effects of anti-TB agents on patients who are false positive by LPA are of concern. When a composite reference standard is used, some of the false positives may become true positives, thus improving sensitivity.</td></tr>
<tr><td>Certainty of the evidence of test accuracy</td><td>What is the overall certainty of the evidence of the test’s accuracy?

○ Very low
○ Low
● Moderate
○ High
○ No included studies</td><td>The risk of bias was considered to be serious for all studies. Indirectness was considered not to be serious. Inconsistency was considered not to be serious. Imprecision was considered not to be serious.

Publication bias: none.</td><td></td></tr>
<tr><td>Certainty of the evidence of test effects</td><td>What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burdens of the test?

○ Very low
○ Low
○ Moderate
○ High
● No included studies</td><td>No studies were included.</td><td></td></tr>
<tr><td>Certainty of the evidence</td><td>What is the overall certainty of the evidence of effects of the management that is guided by the test results?</td><td>In theory, test results should guide management decisions, provided that the use of the test is adopted as national policy. Given the high accuracy of LPAs, a positive test result should be sufficient to start treating a patient. There are insufficient data about how the test performs in smear-negative samples.</td><td></td></tr>
</table>

	○ Very low ○ Low ○ Moderate ○ High ● No included studies		
Certainty of the evidence of test result/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Although this systematic review was not designed to evaluate the clinical impact of LPAs, it was noted that 12 studies attempted to measure the impact of LPAs on clinical impacts, such as turnaround time and cost. For turnaround time, most studies reported the time from a positive culture result to LPA results, with results varying from 8 hours to 5 days and most reporting 1 to 2 days. This was faster than phenotypic DST with liquid cultures, which typically took 9 to 25 days, and solid cultures, which took more than 30 days. One systematic review focused on reductions in diagnostic and treatment delays. The analysis showed that using LPAs reduced diagnostic delays by an average of 47 days (95% CI: 29–64) compared with culture.	
Certainty of effects	**What is the overall certainty of the evidence of the effects of the test?** ○ Very low ○ Low ● Moderate ○ High ○ No included studies	This question is intended to summarize information from the previous four questions about the certainty of the evidence.	
Values	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability ○ No known undesirable outcomes	There is no important uncertainty or variability in how people value the main outcomes. **For detecting rifampicin resistance:** LPAs have demonstrated good diagnostic accuracy when compared with both the phenotypic as well as the composite reference standard.	

Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ● Favours the intervention ○ Varies ○ Don't know	LPAs' good performance in sensitivity and specificity for detecting rifampicin resistance indicates that they are accurate tests, with small numbers of false-negative and false-positive results. Reductions in diagnostic and treatment delays have been documented.	
Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	Cost and cost–effectiveness studies were not assessed. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance.	
Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Cost and cost–effectiveness studies were not assessed for this guideline. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance.	

Cost-effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	Cost and cost–effectiveness studies were not assessed. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Because more patients would have access to the test, health equity may be positively affected. However, the test may introduce barriers to health equity in self-payment environments.	
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable for implementation in settings with a high prevalence of MDR-TB. Implementing the test requires additional human resources, as it is labour intensive, as well as additional infrastructure (three separate rooms) and increased biosafety standards. For patients, the burdens and adverse effects are potentially insignificant.	
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes	In 2008, WHO recommended using this test to diagnose rifampicin-resistant TB in AFB-positive smears and cultures. During the Guideline Development Group meeting there was some disagreement about how feasible it would be to implement LPAs.	

	○ Varies ○ Don't know	A sophisticated laboratory infrastructure and skilled staff are required to perform the test, which are usually available at the intermediate- and reference-levels of laboratory networks. Hence, implementing the test would require additional funding and technical support to train staff and procure equipment. Quality assurance strategies will be needed as well.	

AFB: acid-fast bacilli; CI: confidence interval; DST: drug-susceptibility testing; GRADE: Grading of Recommendations Assessment, Development and Evaluation; LPA: line probe assay; MDR-TB: multidrug-resistant TB.

Summary of judgements

	Judgement						Implications
Problem	No	Probably no	Probably yes	**Yes**	Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate	Varies	Don't know	
Desirable effects	Trivial	Small	Moderate	**Large**	Varies	Don't know	
Undesirable effects	Large	Moderate	**Small**	Trivial	Varies	Don't know	
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High	No included studies		
Certainty of the evidence of test effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of management's effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High	**No included studies**		

	Judgement							Implications
Certainty of effects	Very low	Low	Moderate	High		**No included studies**		
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**		No known undesirable outcomes		
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	**Favours the intervention**	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High		**No included studies**		
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes	Varies		Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes	Varies		Don't know	

Conclusions

Should LPA by indirect testing (compared with phenotypic DST) be used to detect rifampicin resistance in *Mycobacterium tuberculosis* complex culture isolates?

Type of recommendation	Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ●	Strong recommendation for the intervention ○
Recommendation	**For patients with culture-positive TB, the WHO guideline panel suggests using indirect LPA for the detection of rifampicin resistance instead of phenotypic DST on *Mycobacterium tuberculosis* complex culture isolates (conditional recommendation, moderate certainty in the evidence for test accuracy).**				
Justification	There is uncertainty about the impact on cost. Feasibility concerns are moderated by international roll-out of LPA but cannot be ignored; patients who have rifampicin monoresistance by LPA should still have specimens cultured.				
Implementation considerations	Positive results should be interpreted with caution in settings with a very low prevalence of rifampicin resistance; such results possibly require confirmation and repeat testing, but therapy should not be delayed. Implementation should be phased-in gradually along with biosafety upgrades, starting at reference-level laboratories. Facilities requirements must be met (three separate rooms); there must be adequate supplies; and quality assurance strategies must be implemented, as well as reporting mechanisms. Staff training and internal laboratory procedures may need to be revised and changes should be implemented as necessary. Clinicians will need aids for interpreting results.				
Research priorities	Priorities for research include direct clinical trials to assess the impact on patient outcomes of knowing rifampicin-resistance status.				

PICO 3. Accuracy of LPAs for detecting isoniazid resistance by direct testing in sputum smear- positive TB patients compared with phenotypic culture-based DST

	Judgement	Research evidence	Additional considerations
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	*Mycobacterium tuberculosis* causes 9 million cases of TB and 1.5 million deaths annually, and it is estimated that 3.6 million cases of TB go undiagnosed each year. The emergence of MDR-TB and XDR-TB is a major threat to global TB control. Culture and conventional DST using solid and liquid media take from 8 days to 2 months. Hence, the development of rapid molecular diagnostic tests for identifying *M. tuberculosis* and drug resistance have become research and implementation priorities. LPAs detect isoniazid resistance by identifying mutations in *kat*G and *inh*A genes. However, the mutations that cause isoniazid resistance are located in several genes and regions. On average, 80–85% of isoniazid-resistant strains have been found to contain mutations in codon 315 of the *kat*G gene in the *inh*A regulatory region.	
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* LPA for direct testing compared with phenotypic DST Sensitivity: 0.89 (95% CI: 0.86–0.92); specificity: 0.98 (95% CI: 0.97–0.99)	
Desirable effects	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	The anticipated desirable effect is the correct diagnosis of isoniazid-resistant cases (true positives) as well as isoniazid-susceptible cases (true negatives). LPA would correctly identify the majority of isoniazid-resistant cases at pre-test probabilities of 5%, 15% and 90% (see nested table below). Correctly identifying isoniazid-resistant cases (true positives) should lead to higher cure rates, fewer sequelae for the patient, and less transmission in the community. Correctly identifying isoniazid-susceptible cases (true negatives) should allow patients to avoid unnecessary treatment with additional anti-TB agents and the increased risk of severe adverse events; it should also avoid higher costs. The anticipated undesirable effect is the incorrect identification of individuals as isoniazid-sensitive cases when their TB is resistant to isoniazid (false negative). In the pooled data, LPAs misclassified 5–97 cases at pre-test probabilities of 5%, 15% and 90% (see nested	

Undesirable effects

How substantial are the undesirable anticipated effects?

○ Large
○ Moderate
● Small
○ Trivial
○ Varies
○ Don't know

table below). Incorrectly identifying an individual as an isoniazid-resistant case may lead to increased suffering for the patient and TB transmission in the community due to the use of a suboptimal regimen.

Among the undesirable effects, false-negative cases are harmed the most. False-positive diagnoses may result in unnecessary additional treatment with the potential for serious adverse effects.

Test result	Number of results per 1 000 patients tested (95% CI)			Number of participants (number of studies)	Quality of the evidence (GRADE)
	5% prevalence	15% prevalence	90% prevalence		
True positives (patients with isoniazid resistance)	45 (43–46)	134 (129–138)	803 (772–827)	3 576 (46)	⊕⊕⊕◯ 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)	935 (926–940)	836 (829–841)	98 (97–99)	6 896 (46)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly classified as having isoniazid resistance)	15 (10–24)	14 (9–21)	2 (1–3)		

A 90% prevalence of isoniazid resistance is likely to occur in a population of MDR-TB patients when a patient is diagnosed by the Xpert MTB/RIF assay.

Certainty of evidence

What is the overall certainty of the evidence of the test's accuracy?

The risk of bias was considered to be serious for all studies. Indirectness was considered not to be serious. Inconsistency was considered not to be serious. Imprecision was considered not to be serious.

	○ Very low ○ Low ● Moderate ○ High ○ No included studies	Publication bias: none.	
Certainty of the evidence of test effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burdens of the test?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The test is labour intensive and adds to the burdens of the health-worker. Sophisticated laboratory infrastructure and skilled staff are required to perform the test, which is usually available only at intermediate- and central-level laboratories. There may be a diagnostic delay due to the need to transport samples from lower levels of the network to the intermediate or central levels.	
Certainty of the evidence of management' s effects	**What is the overall certainty of the evidence of effects of the management that is guided by the test results?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	In theory, test results should guide management decisions, provided that the use of the test is adopted as national policy. Given the high accuracy of LPAs, a positive test result should be sufficient to start treating a patient. There are insufficient data about how the test performs in smear-negative samples.	
Certainty of the evidence of test	**How certain is the link between test results and management decisions?** ○ Very low ○ Low	Although this systematic review was not designed to evaluate the clinical impact of LPAs, it was noted that 12 studies attempted to measure the impact of LPAs on clinical impacts, such as turnaround time and cost. For turnaround time, most studies reported the time from a positive culture result to LPA results, with results varying from 8 hours to 5 days and most reporting 1 to 2 days. This was faster than phenotypic DST with liquid cultures, which typically took 9 to 25 days, and solid cultures, which took more than 30 days.	

	○ Moderate ○ High ● No included studies	One systematic review focused on reductions in diagnostic and treatment delays. The analysis showed that using LPAs reduced diagnostic delays by an average of 47 days (95% CI: 29–64) compared with culture.	
Certainty of effects	**What is the overall certainty of the evidence of the effects of the test?** ○ Very low ○ Low ● Moderate ○ High ○ No included studies	This question is intended to summarize information from the previous four questions about the certainty of the evidence.	
Values	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ● No important uncertainty or variability ○ No known undesirable outcomes	There is no important uncertainty or variability.	

Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know	At high prevalences there will be large numbers of false-negative results.	The turnaround time for LPAs is faster than that for conventional DST.
Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	Cost and cost–effectiveness studies were not assessed. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance.	

Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	There are no data about resource requirements.	
Cost–effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	There are no data about the cost–effectiveness of the intervention.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased	Because more patients would have access to the test, health equity may be positively affected.	

	○ Increased ○ Varies ○ Don't know		
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable for implementation in settings with a high prevalence of MDR-TB. Implementing the test requires additional human resources, as it is labour intensive, as well as additional infrastructure (three separate rooms) and increased biosafety standards. For patients, the burdens and adverse effects are potentially insignificant.	
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	In 2008, WHO recommended using this test to diagnose rifampicin-resistant TB in AFB-positive smears and cultures. During the Guideline Development Group meeting there was some disagreement about how feasible it would be to implement LPAs. A sophisticated laboratory infrastructure and skilled staff are required to perform the test, which are usually available at the intermediate- and reference-levels of laboratory networks. Hence, implementing the test would require additional funding and technical support to train staff and procure equipment. Quality assurance strategies will be needed as well.	

AFB: acid-fast bacilli; CI: confidence interval; DST: drug-susceptibility testing; GRADE: Grading of Recommendations Assessment, Development and Evaluation; LPA: line probe assay; MDR-TB: multidrug-resistant TB; XDR-TB: extensively drug-resistant TB.

Summary of judgements

	Judgement						**Implications**
Problem	No	Probably no	Probably yes	**Yes**	Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate	Varies	Don't know	
Desirable effects	Trivial	Small	**Moderate**	Large	Varies	Don't know	

	JUDGEMENT						IMPLICATIONS
Undesirable effects	Large	Moderate	**Small**	Trivial	Varies	Don't know	
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High	No included studies		
Certainty of the evidence of test effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of management's effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High	**No included studies**		
Certainty of effects	Very low	Low	Moderate	High	**No included studies**		
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**	No known undesirable outcomes		
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the intervention	Varies	Don't know

	Judgement							Implications
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High		**No included studies**		
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes	Varies		Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes	Varies		Don't know	

Conclusions

Should LPA by direct testing (compared with phenotypic DST) be used to diagnose isoniazid resistance in patients with pulmonary TB?

Type of recommendation	Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ●	Strong recommendation for the intervention ○
Recommendation	**For patients with smear-positive TB, the WHO guideline panel suggests using direct LPA for the detection of isoniazid resistance instead of phenotypic DST (conditional recommendation, moderate certainty in the evidence for test accuracy).**				
Justification	There is uncertainty about the impact on cost. Feasibility concerns are moderated by international roll-out of LPA but cannot be ignored; patients who have isoniazid monoresistance by LPA should still have specimens cultured.				
Implementation considerations	Positive results should be interpreted with caution in settings with a very low prevalence of rifampicin resistance; such results possibly require confirmation and repeat testing, but therapy should not be delayed. Implementation should be phased-in gradually along with				

	biosafety upgrades, starting at reference-level laboratories. Facilities requirements must be met (three separate rooms); there must be adequate supplies; and quality assurance strategies must be implemented, as well as reporting mechanisms. Staff training and internal laboratory procedures may need to be revised and changes should be implemented as necessary. Clinicians will need aids for interpreting results.
Research priorities	Priorities for research include direct clinical trials to assess the impact on patient outcomes of knowing isoniazid-resistance status.

PICO 4. Accuracy of LPA for detecting isoniazid resistance by indirect testing of *Mycobacterium tuberculosis* complex culture isolates compared with phenotypic culture-based DST

	Judgement	Research evidence	Additional considerations
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	*Mycobacterium tuberculosis* causes 9 million cases of TB and 1.5 million deaths annually, and it is estimated that 3.6 million cases of TB go undiagnosed each year. The emergence of MDR-TB and XDR-TB is a major threat to global TB control. Culture and conventional DST using solid and liquid media take from 8 days to 2 months. Hence, the development of rapid molecular diagnostic tests for identifying *M. tuberculosis* and drug resistance have become research and implementation priorities. LPAs detect isoniazid resistance by identifying mutations in *kat*G and *inh*A genes. However, the mutations that cause isoniazid resistance are located in several genes and regions. On average, 80–85% of isoniazid-resistant strains have been found to contain mutations in codon 315 of the *kat*G gene in the *inh*A regulatory region.	
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	*Test accuracy* LPA for indirect testing compared with phenotypic DST Sensitivity: 0.91 (95% CI: 0.89–0.93); specificity: 1.00 (95% CI: 0.99–1.00)	
Desirable effects	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small	The anticipated desirable effect is the correct diagnosis of isoniazid-resistant cases (true positives) as well as isoniazid-susceptible cases (true negatives). LPA would correctly identify the majority of isoniazid-resistant cases at pre-test probabilities of 5%, 15% and 90% (see nested table below). Correctly identifying isoniazid-resistant cases (true positives) should lead to higher cure rates, fewer sequelae for the patient, and less transmission in the community. Correctly identifying isoniazid-susceptible cases (true negatives) should allow patients to avoid unnecessary treatment	

	● Moderate ○ Large ○ Varies ○ Don't know	with additional anti-TB agents and the increased risk of severe adverse events; it should also avoid higher costs. The anticipated undesirable effect is the incorrect identification of individuals as isoniazid-sensitive cases when their TB is resistant to isoniazid (false negative). In the pooled data, LPAs misclassified 4 cases at the pre-test probability of 5%, 13 cases at the pre-test probability of 15%, and 81 cases at the pre-test probability 90% (see nested table below). Incorrectly identifying an individual as an isoniazid-resistant case may lead to increased suffering for the patient and TB transmission in the community due to the use of a suboptimal regimen. Among the undesirable effects, false-negative cases are harmed the most. False-positive diagnoses may result in unnecessary additional treatment with the potential for serious adverse effects.	
Undesirable effects	**How substantial are the undesirable anticipated effects?** ○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know		A 90% prevalence of isoniazid resistance is likely to occur in a population of MDR-TB patients when a patient is diagnosed by the the Xpert MTB/RIF assay.

Test result	Number of results per 1 000 patients tested (95% CI)			Number of participants (number of studies)	Quality of the evidence (GRADE)
	5% prevalence	15% prevalence	90% prevalence		
True positives (patients with isoniazid resistance)	46 (44–47)	137 (133–140)	819 (797–837)	4 559 (43)	⊕⊕⊕◯ 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)	947 (943–950)	847 (844–850)	100 (99–100)	5 903 (43)	⊕⊕⊕◯ MODERATE
False positives (patients incorrectly classified as having isoniazid	3 (0–7)	3 (0–6)	0 (0–1)		

		resistance)					
Certainty of the evidence of the test accuracy	**What is the overall certainty of the evidence of the test's accuracy?** ○ Very low ○ Low ● Moderate ○ High ○ No included studies	The risk of bias was considered to be serious for all studies. Indirectness was considered not to be serious. Inconsistency was considered not to be serious. Imprecision was considered not to be serious. Publication bias: none.					
Certainty of the evidence of the test effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burdens of the test?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	No studies were included.					
Certainty of the evidence of management' s effects	**What is the overall certainty of the evidence of effects of the management that is guided by the test results?**	In theory, test results should guide management decisions, provided that the use of the test is adopted as national policy. Given the high accuracy of LPAs, a positive test result should be sufficient to start treating a patient. There are insufficient data about how the test performs in smear-negative samples.					

	○ Very low ○ Low ○ Moderate ○ High ● No included studies		
Certainty of the evidence of test result/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Although this systematic review was not designed to evaluate the clinical impact of LPAs, it was noted that 12 studies attempted to measure the impact of LPAs on clinical impacts, such as turnaround time and cost. For turnaround time, most studies reported the time from a positive culture result to LPA results. with results varying from 8 hours to 5 days and most reporting 1 to 2 days. This was faster than phenotypic DST with liquid cultures, which typically took 9 to 25 days, and solid cultures, which took more than 30 days. One systematic review focused on reductions in diagnostic and treatment delays. The analysis showed that using LPAs reduced diagnostic delays by an average of 47 days (95% CI: 29–64) compared with culture.	
Certainty of effects	**What is the overall certainty of the evidence of the effects of the test?** ○ Very low ○ Low ● Moderate ○ High ○ No included studies	This question is intended to summarize information from the previous four questions about the certainty of the evidence.	
Values	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or	There is no important uncertainty or variability.	

	variability ● No important uncertainty or variability ○ No known undesirable outcomes		
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know	At high prevalences there will be large numbers of false-negative results.	The turnaround time for LPAs is faster than that for conventional DST.
Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings	Cost and cost–effectiveness studies were not assessed. Potential areas needing investment include infrastructure, sample referral procedures, equipment and maintenance.	

	○ Varies ● Don't know		
Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	There are no data about resource requirements.	
Cost–effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	There are no data about the cost–effectiveness of the intervention.	
Equity	**What would be the**	Because more patients would have access to the test, health equity may be positively affected.	

	impact on health equity? ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know		
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable for implementation in settings with a high prevalence of MDR-TB. Implementing the test requires additional human resources, as it is labour intensive, as well as additional infrastructure (three separate rooms) and increased biosafety standards. For patients, the burdens and adverse effects are potentially insignificant.	
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	In 2008, WHO recommended using this test to diagnose rifampicin-resistant TB in AFB-positive smears and cultures. During the Guideline Development Group meeting there was some disagreement about how feasible it would be to implement LPAs. A sophisticated laboratory infrastructure and skilled staff are required to perform the test, which are usually available at the intermediate- and reference-levels of laboratory networks. Hence, implementing the test would require additional funding and technical support to train staff and procure equipment. Quality assurance strategies will be needed as well.	

AFB: acid-fast bacilli; CI: confidence interval; DST: drug-susceptibility testing; LPA: line probe assay; MDR-TB: multidrug-resistant TB; XDR-TB: extensively drug-resistant TB.

Summary of judgements

	Jusdgement						**Implication**
Problem	No	Probably no	Probably yes	**Yes**	Varies	Don't know	

	Judgement						Implication
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate	Varies	Don't know	
Desirable effects	Trivial	Small	**Moderate**	Large	Varies	Don't know	
Undesirable effects	Large	Moderate	**Small**	Trivial	Varies	Don't know	
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High	No included studies		
Certainty of the evidence of test effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of management's effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High	**No included studies**		
Certainty of effects	Very low	Low	Moderate	High	**No included studies**		
Values	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**	No known undesirable outcomes		

	Judgement							Implications
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the intervention	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High		**No included studies**		
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know	

Conclusions

Should LPA by indirect testing (compared with phenotypic DST) be used to diagnose isoniazid resistance in *Mycobacterium tuberculosis* complex culture isolates?

Type of recommendation	Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
	○	○	○	●	○
Recommendation	**For patients with culture-positive TB, the WHO guideline panel suggests using indirect LPA for detection of isoniazid resistance instead of phenotypic DST in *Mycobacterium tuberculosis* complex culture isolates (conditional recommendation, moderate certainty in the evidence for test accuracy).**				

Justification	There is uncertainty about the impact on cost. Feasibility concerns are moderated by international roll-out of LPA but cannot be ignored; patients who have isoniazid mono-resistance by LPA should still have specimens cultured.
Implementation consideration	Positive results should be interpreted with caution in settings with a very low prevalence of rifampicin resistance; such results possibly require confirmation and repeat testing, but therapy should not be delayed. Implementation should be phased-in gradually along with biosafety upgrades, starting at reference-level laboratories. Facilities requirements must be met (three separate rooms); there must be adequate supplies; and quality assurance strategies must be implemented, as well as reporting mechanisms. Staff training and internal laboratory procedures may need to be revised and changes should be implemented as necessary. Clinicians will need aids for interpreting results.
Research priorities	Priorities for research include direct clinical trials to assess the impact on patient outcomes of knowing isoniazid-resistance status.

3.8 Evidence-to-decision tables: second-line line probe assay (SL-LPA)

PICO 1: Evidence to recommendation: Accuracy of MTBDRsl by direct testing for detection of fluoroquinolone resistance in patients with rifampicin-resistant or MDR-TB

	Judgement	Research evidence	Additional considerations
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	In 2014 WHO has estimated that 9.7% of the 480,000 cases of MDR-TB, were actually XDR TB, i.e. MDR TB with added resistance to at least one FQ and one SLID. Genotypic (molecular) methods have considerable advantages for scaling up programmatic management and surveillance of drug-resistant TB, offering speed of diagnosis, standardized testing, potential for high through-put, and fewer requirements for laboratory biosafety. Molecular tests for detecting drug resistance such as the MTBDR*sl* assay have shown promise for the diagnosis of drug-resistant tuberculosis (TB). The MTBDR*sl* assay incorporates probes to detect mutations within genes (*gyr*A and *rrs* for version 1.0 and, in addition, *gyr*B and the *eis* promoter for version 2.0), which are associated with resistance to the class of fluoroquinolones or the class of second-line injectable drugs (SLID).	Additional regions associated with resistance to FQ and SLIDs are included in the version 2.0 assay. Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ○ Accurate	In this review – data from the 9 studies, 1771 patients, reference standard: culture based DST *Test accuracy* MTBDR*sl* by direct testing for fluoroquinolones: Sensitivity: 86% (95% CI: 75% to 93%) Specificity: 99% (95% CI: 97% to 99%)	The presence of mutations in these regions does not necessarily imply resistance to all the drugs within that class. Although

	○ Very accurate ● Varies ○ Don't know	More data is needed to better understand the correlation of the presence of certain fluoroquinolone resistance conferring mutations with phenotypic DST resistance for moxifloxacin and patient outcomes.	specific mutations within these regions may be associated with different levels of resistance to each drug within these classes, the extent of this is not completely understood.
Desirable effects	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	The anticipated **desirable** effect is the correct diagnosis of fluoroquinolone (FQ) resistant cases (TP) as well as FQ susceptible cases (TN). MTBDR*s*/ would correctly identify 43 cases out of 50 per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%. For 10-15% there would be 86 and 129 patients respectively (see table below). **Correct identification of FQ resistant cases should lead to higher cure rates, less sequelae to the individual patient, and less transmission in the community.** Similarly MTBDR*s*/ would correctly identify 937 FQ-susceptible (TN) out of 950 per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%. For 10-15% prevalence's there would be 887 and 838 patients respectively (see table below). **Correct identification of FQ susceptible cases should lead to avoiding unnecessary treatment with additional drugs with increased risk of severe adverse events and greater costs.**	**Desirable anticipated effects** MTBDR*s*/ can be performed in a single day to allow the initiation of an appropriate treatment regimen. Phenotypic DST more difficult to perform.
Undesirable effects	**How substantial are the undesirable anticipated effects?** ○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know	The anticipated **undesirable** effect is the incorrect identification of an individual as a FQ susceptible or FQ resistant case (FN or FP). MTBDR*s*/ would misclassify 7 cases as FN per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%, and 14 to 21 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as FQ susceptible may have a potential increased risk of patient morbidity and mortality, continued risk of community transmission of drug-resistant TB.** However, the harm may be lessened as patients without resistance detected to fluoroquinolones may be eligible for an MDR-TB regimen which would include either moxifloxacin or gatifloxacin. MTBDR*s*/ had misclassified 13 cases as FP per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%, and 13 to 12 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as FQ resistant may lead to patient anxiety, possible delays in further diagnostic evaluation, prolonged and unnecessary treatment with drugs that may have additional serious adverse effects.** **Should MTBDRsl by direct testing be used to diagnose FQ resistance in patients with RR or MDR TB?**	**Undesirable anticipated effects per drug:** FN results are of main concern as patients may not be given an effective treatment regimen. Less concern for FP results. Conventional phenotypic DST should be used in the follow-up evaluation of patients with a negative result especially in settings with a high pre-test

Test result	Number of results per 1000 patients tested (95% CI)			Number of participants (studies)
	Prevalence 5%	Prevalence 10%	Prevalence 15%	
True positives (patients with FQ resistance)	43 (37 to 47)	86 (75 to 93)	129 (112 to 140)	519 (9)

		False negatives (patients incorrectly classified as not having FQ resistance)	7 (3 to 13)	14 (7 to 25)	21 (10 to 38)		probability for resistance ⊕◯◯◯ to fluoroquinolones. MODERATE
		True negatives (patients without FQ resistance)	937 (921 to 944)	887 (872 to 895)	838 (824 to 845)	1252 (9)	⊕⊕⊕⊕ HIGH Two GDG members thought that the undesirable effects were large.
		False positives (patients incorrectly classified as having FQ resistance)	13 (6 to 29)	13 (5 to 28)	12 (5 to 26)		
		Implications for the detection of FQ conferring mutations among RR-TB persons TP: Test result suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen. FP: Test result suggests modification of a WHO recommended MDR-TB regimen. Increased risk of serious adverse effects. Patient receive optimal regimen. FN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. Patient receive suboptimal regimen. No benefits. TN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen.					Physicians should be guided by the MTBDRs/ assay in their initial choice of an MDR-TB treatment regimen.
Certainty of the evidence of test accuracy	What is the overall certainty of the evidence of test accuracy? ○ Very low ○ Low ● Moderate ○ High ○ No included studies	In this review the **risk of bias** was not serious **Indirectness w**as considered not serious **Inconsistency** was considered serious for test sensitivity and not serious for test specificity **Imprecision** was considered not serious for both sensitivity and specificity **Publication bias** – none for all studies.					**Quality of evidence for test accuracy** is: Sensitivity –moderate quality of evidence Specificity – high quality of evidence
Certainty of the evidence of the test effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?** ○ Very low ○ Low ○ Moderate ○ High ● No	The test is labour-intensive and presents certain burden for the health worker. There is a need for appropriate infrastructure with separate rooms and biosafety requirements, which assumes a considerable investment. The burden and adverse effects are potentially insignificant for the patient.					

	included studies		
Certainty of the evidence of the management effects	**What is the overall certainty if the evidence of effects of the management that is guided by the test results?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Ideally test results should guide management decisions, provided use of test is adopted by national policy. A positive test result should be sufficient for a patient to start treatment.	
Certainty of the evidence of the test result/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The link between test results and management decisions may be uncertain in various settings. In some occasions clinicians use empirical treatment for TB. In others capacity of health system may be insufficient to provide the patient with necessary treatment.	Turnaround time would be faster than for conventional DST The need for sample referral may cause delays
Certainty of effects	**What is the overall certainty of the evidence of effects of the test?** ○ Very low ○ Low ● Moderate ○ High	This question is intended to summarize previous four questions on the certainty of the evidence.	

	○ No included studies		
Values	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability ○ No known undesirable outcomes	There is no important uncertainty about or variability in how much people value the main outcomes.	
Balance of evffects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison	FN results increase with increasing pre-test probability for FQ resistance. Conventional phenotypic DST should be used in the follow-up evaluation of patients with a negative result especially in settings with a high pre-test probability for resistance to fluoroquinolones.	

	○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know		
Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	No research evidence was identified.	
Certainty of the evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High	No research evidence was identified.	

	● No included studies		
Cost-effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	No research evidence was identified.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably	System incorporating molecular methods provides more equity.	

	reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know		
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable to be implemented in reference settings, where infrastructure and qualified staff to perform MTBDR*sl* exist. If MTBDR*sl* is implemented for first-line DST the MTBDR*sl* assay could be performed on the same specimen for rifampicin-resistant TB or MDR-TB cases.	
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Implementation of the test would require additional funding and technical support for the infrastructure upgrade, training of staff and procuring the equipment.	

Summary of judgments

	Judgement						Implications
Problem	No	Probably no	Probably yes	**Yes**	Varies	Don't know	

	Judgement						Implications
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate	Varies	Don't know	
Desirable effects	Trivial	Small	Moderate	**Large**	Varies	Don't know	
Undesirable effects	Large	**Moderate**	Small	Trivial	Varies	Don't know	
Certainty of the evidence of test accuracy	Very low	Low	**Moderate**	High	**No included studies**		
Certainty of the evidence of test effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of management's effects	Very low	Low	Moderate	High	**No included studies**		
Certainty of the evidence of test result/management	Very low	Low	Moderate	High	**No included studies**		
Certainty of effects	Very low	Low	**Moderate**	High	**No included studies**		
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability	**No known undesirable outcomes**		
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the intervention	Varies	Don't know

	Judgement							Implications
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High		**No included studies**		
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know	

Conclusions

Should MTBDR*s*/ by direct testing be used to diagnose FQ resistance in patients with RR or MDR TB?					
Type of recommendation	Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ●	Strong recommendation for the intervention ○
Recommendation	For patients with confirmed rifampicin-resistant TB or MDR-TB, the WHO guideline development group suggests using direct testing of patient specimens with the MTBDR*s*/ assay as the initial test, over culture and phenotypic DST, to detect resistance to FQ (Conditional recommendation, Moderate certainty in the evidence for test accuracy).				
Subgroup considerations	Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.				

Implementation considerations	Adoption of the MTBDR*sl* assay does not eliminate the need for conventional culture and DST capability. Despite good specificity of the MTBDR*sl* for the detection of resistance to FQs, culture and phenotypic DST is required to completely exclude resistance to this drug class. However, the demand for conventional culture and DST capacity may change, based on the prevalence of resistance to second-line anti-TB drugs in patients with confirmed RR-TB or MDR-TB. The following implementation considerations apply: • MTBDR*sl* cannot determine resistance to individual drugs in the class of fluoroquinolones. Phenotypic resistance to ofloxacin and levofloxacin is highly correlated with resistance conferring mutations detected by the MTBDR*sl* assay. Uncertainty remains about the susceptibility to moxifloxacin and gatifloxacin for such strains with mutations; • MTBDR*sl* assay should be used in the direct testing of sputum samples irrespective of whether samples are smear-negative or smear-positive from patients with confirmed rifampicin-resistant TB or MDR-TB; • MTBDR*sl* assay is designed to TB and resistance to second-line injectable drugs from processed sputum samples. Other respiratory samples (e.g. bronchoalveolar lavage and gastric aspirates) or extrapulmonary samples (tissue samples, CSF or other body fluids) have not been adequately evaluated; • Culture and phenotypic DST plays a critical role in the monitoring of patients' response to treatment and for detecting additional resistance to second-line drugs during treatment. Patients with false negative resistance results using the MTBDR*sl* can be identified and captured through treatment monitoring. Patients with false positive results might benefit from the addition of other drugs; • The availability of additional second-line drugs is critical.
Monitoring and evaluations	System of quality assurance is necessary.
Research priorities	Current recommendations on the MTBDR*sl* assay should not prevent or restrict further research on new rapid molecular DST tests, especially for assays that can be used as close as possible to where patients are initially diagnosed with RR-TB and MDR-TB and where treatment can be initiated. Further operational research on the MTBDR*sl* test should focus on the following priorities: • Develop and improved understanding of the correlation between the detection of resistance conferring mutations with phenotypic DST results and patient outcomes; • Develop improved knowledge of the presence of specific mutations detected with the MTBDR*sl* assay correlated with MICs for individual drugs within the class of fluoroquinolones; • Review evidence to confirm or revise different critical concentrations used in phenotypic DST methods; • Determine the limit of detection of MTBDR*sl* for the detection of heteroresistance; • Determine training, competency, and quality assurance needs; • Gather more evidence on the impact on appropriate MDR-TB treatment initiation and mortality; • Meet "Standards for Reporting Diagnostic accuracy studies" (STARD) for future studies; • Perform country-specific cost-effectiveness and cost-benefit analyses of MTDDR*sl* assay use in different programmatic settings.

PICO 2: Evidence to Decisions tables: **Accuracy of MTBDR*sl* by direct testing for detection of SLID resistance in patients with rifampicin-resistant or MDR-TB**

Judgement	Research evidence	Additional considerations

Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	In 2014 WHO has estimated that 9.7% of the 480,000 cases of MDR-TB , were actually XDR TB, i.e. MDR TB with added resistance to at least one FQ and one SLID. Genotypic (molecular) methods have considerable advantages for scaling up programmatic management and surveillance of drug-resistant TB, offering speed of diagnosis, standardized testing, potential for high through-put, and fewer requirements for laboratory biosafety. Molecular tests for detecting drug resistance such as the MTBDR*sl* assay have shown promise for the diagnosis of drug-resistant tuberculosis (TB). The MTBDR*sl* assay incorporates probes to detect mutations within genes (*gyr*A and *rrs* for version 1.0 and, in addition, *gyr*B and the *eis* promoter for version 2.0), which are associated with resistance to the class of fluoroquinolones or the class of second-line injectable drugs (SLID).	Additional regions associated with resistance to FQ and SLIDs are included in the version 2.0 assay. Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ○ Accurate ○ Very accurate ● Varies ○ Don't know	In this review – data from the 8 studies, 1639 patients, reference standard: culture based DST *Test accuracy* MTBDR*sl* by direct testing for SLID: Sensitivity: 87% (95% CI: 38% to 99%) Specificity: 99% (95% CI: 94% to 100%) MTBDR*sl* by direct testing for Amikacin: Sensitivity: 92% (95% CI: 71% to 98%) Specificity: 100% (95% CI: 95% to 100%) MTBDR*sl* by direct testing for Kanamycin: Sensitivity: 79% (95% CI: 12% to 99%) Specificity: 100% (95% CI: 94% to 100%) MTBDR*sl* by direct testing for Capreomycin: Sensitivity: 77% (95% CI: 61% to 87%7) Specificity: 98% (95% CI: 93% to 100%)	The accuracy varies with the different SLID. The variability is explained in part by the use of different drugs, critical concentrations , types of culture media in the reference standard and likely presence of *eis* resistance-conferring mutations in patients in Eastern European countries.
Desirable effects	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	The anticipated **desirable** effect is the correct diagnosis of SLID resistant cases (TP) as well as SLID susceptible cases (TN). MTBDR*sl* would correctly identify 44 cases out of 50 per 1000 individuals tested if the pre-test probability of TB is 5%. For 10-15% there would be 87 and 131 patients respectively (see table below). **Correct identification of SLID resistant cases should lead to higher cure rates, less sequelae to the individual patient, and less transmission in the community.** Similarly MTBDR*sl* would correctly identify 945 TB-free cases (TN) out of 950 per 1000 individuals tested if the pre-test probability of TB is 5%. For 10-15% prevalence's there would be 896 and 846 patients respectively (see table below). **Correct identification of SLID susceptible cases should lead to avoiding unnecessary treatment with additional drugs with increased risk of severe adverse events and greater costs.**	**Desirable anticipated effects per drug:** Amikacin – Large desirable effects Capreomycin – Large desirable effects Kanamycin – Large desirable effects

Undesirable effects

How substantial are the undesirable anticipated effects?

○ Large
○ Moderate
○ Small
○ Trivial

● Varies
○ Don't know

The anticipated **undesirable** effect is the incorrect identification of an individual as a SLID susceptible or resistant case (FN or FP).

MTBDR*s*l would misclassify 6 cases as FN per 1000 individuals tested if the pre-test probability of TB with SLID resistance is 5%, and 13 to 19 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as SLID susceptible may have a potential increased risk of patient morbidity and mortality, and continued risk of community transmission of drug-resistant TB as well initiation of an MDR-TB regimen which includes a SLID with doubtful efficacy.**

MTBDR*s*l had misclassified 5 cases as FP per 1000 individuals tested if the pre-test probability of TB is 5%, and 4 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as SLID resistant may lead to patient anxiety, possible delays in further diagnostic evaluation, prolonged and unnecessary treatment with drugs that may have additional serious adverse effects.**

Should MTBDRsl by direct testing be used to diagnose SLID resistance in patients with RR or MDR TB?

Test result	Number of results per 1000 patients tested (95% CI)			Number of participants (studies)
	Prevalence 5%	Prevalence 10%	Prevalence 15%	
True positives (patients with SLID resistance)	44 (19 to 49)	87 (38 to 99)	131 (57 to 148)	348 (8)
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)	945 (889 to 950)	896 (842 to 900)	846 (796 to 850)	1291 (8)
False positives (patients incorrectly classified as having SLID resistance)	5 (0 to 61)	4 (0 to 58)	4 (0 to 54)	

Implications for the detection of SLID conferring mutations among RR-TB persons

TP: Test result suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen.
FP: Test result suggests modification of a WHO recommended MDR-TB regimen. Increased risk of serious adverse effects. Patient receive optimal regimen.
FN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. Patient receive suboptimal regimen. No benefits.
TN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen.

Undesirable anticipated effects per drug:

Amikacin – Small undesirable effects

Capreomycin and kanamycin – moderate undesirable effects

Two GDG members thought that the undesirable effects were large.

Physicians should be guided by the MTBDR*s*l assay in their initial choice of an MDR-TB treatment regimen.

Conventional phenotypic DST should be used in the follow-up evaluation of patients with a negative result especially in settings with a high pre-test probability for resistance to SLIDs.

Certainty of the evidence of test accuracy	What is the overall certainty of the evidence of test accuracy? ○ Very low ● Low ○ Moderate ○ High ○ No included studies	In this review the **risk of bias** was serious **Indirectness w**as considered not serious **Inconsistency** was considered not serious **Imprecision** was considered serious for sensitivity and not serious for specificity **Publication bias** – none for all studies	**Quality of evidence for test accuracy** is: Sensitivity –low quality of evidence Specificity – moderate quality of evidence Kanamycin- low certainty Capreomycin- low certainty Amikacin – moderate certainty
Certainty of the evidence of tests' effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The test is labour-intensive and presents certain burden for the health worker. There is a need for appropriate infrastructure with separate rooms and biosafety requirements, which assumes a considerable investment. The burden and adverse effects are potentially insignificant for the patient.	
Certainty of the evidence of the management' effects	**What is the overall certainty if the evidence of effects of the managem ent that is guided by the test results?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Ideally test results should guide management decisions, provided use of test is adopted by national policy. A positive test result should be sufficient for a patient to start treatment.	

Certainty of the evidence of test result/management	**How certain is the link between test results and management decisions ?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The link between test results and management decisions may be uncertain in various settings. In some occasions clinicians use empirical treatment for TB. In others capacity of health system may be insufficient to provide the patient with necessary treatment.	Turnaround time would be faster than for conventional DST The need for sample referral may cause delays
Certainty of effects	**What is the overall certainty of the evidence of effects of the test?** ○ Very low ● Low ○ Moderate ○ High ○ No included studies	This question is intended to summarize previous four questions on the certainty of the evidence.	Kanamycin and Capreomycin – low certainty Amikacin – moderate certainty
Values	**Is there important uncertainty about or variability in how much people value the main outcomes ?** ○ Important uncertainty or variability ○ Possibly important uncertainty	There is no important uncertainty about or variability in how much people value the main outcomes.	

	or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability ○ No known undesirable outcomes		
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know	Desirable (Amikacin, Kanamycin, Capreomycin) - Large, Large, Large Undesirable (Amikacin, Kanamycin, Capreomycin) - Small, Moderate, Moderate	Concern - FN Accuracy for detecting amikacin resistance is better than for capreomycin or kanamycin.

Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	No research evidence was identified.	
Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	No research evidence was identified.	
Cost effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison	No research evidence was identified.	

	○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies		
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	System incorporating molecular methods provides more equity.	
Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable to be implemented in reference settings, where infrastructure and qualified staff to perform MTBDR*sl* exist. If MTBDR*sl* is implemented for first-line DST the MTBDR*sl* assay could be performed on the same sample	

Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Implementation of the test would require additional funding and technical support for the infrastructure upgrade, training of staff and procuring the equipment.	

Summary of judgments

	Judgement							Implications
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	Accurate	Very accurate		**Varies**	Don't know	
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know	
Undesirable effects	Large	Moderate	Small	Trivial		**Varies**	Don't know	
Certainty of the evidence of test accuracy	Very low	**Low**	Moderate	High			No included studies	
Certainty of the evidence of test effects	Very low	Low	Moderate	High			**No included studies**	
Certainty of the evidence of management's effects	Very low	Low	Moderate	High			**No included studies**	
Certainty of the evidence of test	Very low	Low	Moderate	High			**No include**	

	JUDGEMENT							IMPLICATIONS
							d studies	
Certainty of effects	Very low	**Low**	Moderate	High			No included studies	
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			No known undesirable outcomes	
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the intervention	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**	
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know	

Conclusions

Should MTBDRsl by direct testing be used to diagnose SLID resistance in patients with RR or MDR TB?					
Type of recommendation	Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ●	Strong recommendation for the intervention ○
Recommendation	For patients with confirmed rifampicin-resistant TB or MDR-TB, the WHO guideline development group suggests using direct testing of patient specimens with the MTBDR*s*l assay as the initial test, over culture and phenotypic DST, to detect resistance to SLID (Conditional recommendation, Low certainty in the evidence for test accuracy).				
Subgroup considerations	Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.				
Implementation considerations	Adoption of the MTBDR*s*l assay does not eliminate the need for conventional culture and DST capability. Despite good specificity of the MTBDRsl for the detection of resistance to SLIDs, culture and phenotypic DST is required to completely exclude resistance to these drug classes. However, the demand for conventional culture and DST capacity may change, based on the prevalence of resistance to second-line anti-TB drugs in patients with confirmed RR-TB or MDR-TB. The following implementation considerations apply: • Mutations in some regions (e.g., the eis promoter region) may be responsible for causing resistance to one drug in a class more than other drugs within that class. The *eis* C14T mutation is associated with kanamycin resistance in strains from Eastern Europe; • MTBDR*s*l assay should be used in the direct testing of sputum samples irrespective of whether samples are smear-negative or smear-positive from patients with confirmed rifampicin-resistant TB or MDR-TB; • MTBDR*s*l assay is designed to TB and resistance to second-line injectable drugs from processed sputum samples. Other respiratory samples (e.g. bronchoalveolar lavage and gastric aspirates) or extrapulmonary samples (tissue samples, CSF or other body fluids) have not been adequately evaluated; • Culture and phenotypic DST plays a critical role in the monitoring of patients' response to treatment and for detecting additional resistance to second-line drugs during treatment. Patients with false negative resistance results using the MTBDR*s*l can be identified and captured through treatment monitoring. Patients with false positive results might benefit from the addition of other drugs; • The availability of additional second-line drugs is critical.				
Monitoring and evaluations	System of quality assurance is necessary.				
Research priorities	Current recommendations on the MTBDR*s*l assay should not prevent or restrict further research on new rapid molecular DST tests, especially for assays that can be used as close as possible to where patients are initially diagnosed with RR-TB and MDR-TB and where treatment can be initiated. Further operational research on the MTBDR*s*l test should focus on the following priorities:				

	• Develop and improved understanding of the correlation between the detection of resistance conferring mutations with phenotypic DST results and patient outcomes; • Develop improved knowledge of the presence of specific mutations detected with the MTBDR*sl* assay correlated with MICs for individual drugs within the class of SLIDs; • Review evidence to confirm or revise different critical concentrations used in phenotypic DST methods; • Determine the limit of detection of MTBDR*sl* for the detection of heteroresistance; • Determine training, competency, and quality assurance needs; • Gather more evidence on the impact on appropriate MDR-TB treatment initiation and mortality; • Meet "Standards for Reporting Diagnostic accuracy studies" (STARD) for future studies; • Perform country-specific cost-effectiveness and cost-benefit analyses of MTDDR*sl* assay use in different programmatic settings.

PICO 3: Evidence to recommendations: Accuracy of MTBDRsl by indirect testing for detection of fluoroquinolone resistance in patients with rifampicin-resistant or MDR-TB

	Judgement	Research evidence	Additional considerations
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	In 2014 WHO has estimated that 9.7% of the 480,000 cases of MDR-TB , were actually XDR TB, i.e. MDR TB with added resistance to at least one FQ and one SLID. Genotypic (molecular) methods have considerable advantages for scaling up programmatic management and surveillance of drug-resistant TB, offering speed of diagnosis, standardized testing, potential for high through-put, and fewer requirements for laboratory biosafety. Molecular tests for detecting drug resistance such as the MTBDR*sl* assay have shown promise for the diagnosis of drug-resistant tuberculosis (TB). The MTBDR*sl* assay incorporates probes to detect mutations within genes (*gyr*A and *rrs* for version 1.0 and, in addition, *gyr*B and the *eis* promoter for version 2.0), which are associated with resistance to the class of fluoroquinolones or the class of second-line injectable drugs (SLID).	Additional regions associated with resistance to FQ and SLIDs are included in the version 2.0 assay. Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies	In this review – data from the 19 studies, 2223 patients, reference standard: culture based DST *Test accuracy* MTBDR*sl* by indirect testing for fluoroquinolones: Sensitivity: 86%(95% CI: 79% to 90%) Specificity: 99% (95% CI: 97% to 99%) More data is needed to better understand the correlation of the presence of certain fluoroquinolone resistance conferring mutations with phenotypic DST resistance for moxifloxacin and patient outcomes.	The presence of mutations in these regions does not necessarily imply resistance to all the drugs within that class. Although specific mutations within these regions may be associated with different

	○ Don't know	The diagnostic accuracy of MTBDRsl is similar when performed using either direct or indirect testing.	levels of resistance to each drug within these classes, the extent of this is not completely understood.
Desirable effects	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	The anticipated **desirable** effect is the correct diagnosis of fluoroquinolone (FQ) resistant cases (TP) as well as FQ susceptible cases (TN). MTBDR*sl* would correctly identify 43 cases out of 50 per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%. For 10-15% there would be 86 and 128 patients respectively (see table below). **Correct identification of FQ resistant cases should lead to higher cure rates, less sequelae to the individual patient, and less transmission in the community.** Similarly MTBDR*sl* would correctly identify 937 FQ-susceptible (TN) out of 950 per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%. For 10-15% prevalence's there would be 887 and 838 patients respectively (see table below). **Correct identification of FQ susceptible cases should lead to avoiding unnecessary treatment with additional drugs with increased risk of severe adverse events and greater costs.** The anticipated **undesirable** effect is the incorrect identification of an individual as a FQ susceptible or FQ resistant case (FN or FP).	**Desirable anticipated effects** Indirect testing with MTBDR*sl* can be performed in a single day once the culture is grown. The method is faster and easier to perform than phenotypic DST.
Undesirable effects	**How substantial are the undesirable anticipated effects?** ○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know	MTBDR*sl* would misclassify 7 cases as FN per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%, and 14 to 22 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as FQ susceptible may have a potential increased risk of patient morbidity and mortality, continued risk of community transmission of drug-resistant TB.** However, the harm may be lessened as patients without resistance detected to fluoroquinolones may be eligible for an MDR-TB regimen which would include either moxifloxacin or gatifloxacin. MTBDR*sl* had misclassified 13 cases as FP per 1000 individuals tested if the pre-test probability of TB with FQ resistance is 5%, and 13 to 12 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as FQ resistant may lead to patient anxiety, possible delays in further diagnostic evaluation, prolonged and unnecessary treatment with drugs that may have additional serious adverse effects.** **Should MTBDRsl by indirect testing be used to diagnose FQ resistance in patients with RR or MDR TB?**	**Undesirable anticipated effects per drug:** FN results are of main concern as patients may not be given an effective treatment regimen. Less concern for FP results. Conventional phenotypic DST should be used in the follow-up evaluation of patients with a negative result especially in settings with a high pre-test probability for resistance to

Test result	Number of results per 1000 patients tested (95% CI)			Number of participants (studies)
	Prevalence 5%	Prevalence 10%	Prevalence 15%	
True positives (patients with FQ resistance)	43 (40 to 45)	86 (79 to 90)	128 (119 to 133)	869 (19)
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)	937 (921 to 944)	887 (872 to 895)	838 (824 to 845)	1354 (19)	fluoroquinolon es. ⊕⊕◯◯ LOW
		False positives (patients incorrectly classified as having FQ resistance)	13 (6 to 29)	13 (5 to 28)	12 (5 to 26)		
		Implications for the detection of FQ conferring mutations among RR-TB persons TP: Test result suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen. FP: Test result suggests modification of a WHO recommended MDR-TB regimen. Increased risk of serious adverse effects. Patient receive optimal regimen. FN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. Patient receive suboptimal regimen. No benefits. TN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen.					
Certainty of the evidence of test accuracy	What is the overall certainty of the evidence of test accuracy? ● Very low ○ Low ○ Moderate ○ High ○ No included studies	In this review the **risk of bias** was serious **Indirectness** was considered serious **Inconsistency** was considered serious for test sensitivity and not serious for test specificity **Imprecision** was considered not serious for sensitivity and specificity **Publication bias** – none for all studies.					**Quality of evidence for test accuracy** is: Sensitivity – very low quality of evidence Specificity – low quality of evidence
Certainty of the evidence of the test effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The test is labour-intensive and presents certain burden for the health worker. There is a need for appropriate infrastructure with separate rooms and biosafety requirements, which assumes a considerable investment. The burden and adverse effects are potentially insignificant for the patient.					

Certainty of the evidence of managements effects	**What is the overall certainty if the evidence of effects of the management that is guided by the test results?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	Ideally test results should guide management decisions, provided use of test is adopted by national policy. A positive test result should be sufficient for a patient to start treatment.	
Certainty of the evidence of the test result/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The link between test results and management decisions may be uncertain in various settings. In some occasions clinicians use empirical treatment for TB. In others capacity of health system may be insufficient to provide the patient with necessary treatment.	Turnaround time would be faster than for conventional DST The need for sample referral may cause delays
Certainty of effects	**What is the overall certainty of the evidence of effects of the test?** ○ Very low ○ Low ● Moderate ○ High ○ No included studies	This question is intended to summarize previous four questions on the certainty of the evidence.	

Values	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability ○ No known undesirable outcomes	There is no important uncertainty about or variability in how much people value the main outcomes.	
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour	FN results increase with increasing pre-test probability for FQ resistance. Conventional phenotypic DST should be used in the follow-up evaluation of patients with a negative result especially in settings with a high pre-test probability for resistance to fluoroquinolones.	

	either the intervention or the comparison ● Probably favours the intervention ○ Favours the intervention ○ Varies ○ Don't know		
Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	No research evidence was identified.	
Certainty of the evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	No research evidence was identified.	

Cost effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	No research evidence was identified.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	System incorporating molecular methods provides more equity.	

Acceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable to be implemented in reference settings, where infrastructure and qualified staff to perform MTBDR*sl* exist. If MTBDR*sl* is implemented for first-line DST the MTBDR*sl* assay could be performed on the same specimen for rifampicin-resistant TB or MDR-TB cases.	
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Implementation of the test would require additional funding and technical support for the infrastructure upgrade, training of staff and procuring the equipment.	

Summary of judgments

	Judgement							Implications
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know	
Desirable effects	Trivial	Small	**Moderate**	Large		Varies	Don't know	
Undesirable effects	Large	**Moderate**	Small	Trivial		Varies	Don't know	

	Judgement							Implications
Certainty of the evidence of test accuracy	**Very low**	Low	Moderate	High			No included studies	
Certainty of the evidence of test effects	Very low	Low	Moderate	High			**No included studies**	
Certainty of the evidence of management's effects	Very low	Low	Moderate	High			**No included studies**	
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**	
Certainty of effects	**Very low**	Low	Moderate	High			No included studies	
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			No known undesirable outcomes	
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the intervention	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**	

	Judgement							Implications
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know	

Conclusions

Should MTBDR*sl* by direct testing be used to diagnose FQ resistance in patients with RR or MDR TB?					
Type of recommendation	Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ●	Strong recommendation for the intervention ○
Recommendation	For patients with confirmed rifampicin-resistant TB or MDR-TB, the WHO guideline development group suggests using indirect testing of cultured isolates of *M.tuberculosis* with the MTBDR*sl* assay as the initial test, over culture and phenotypic DST, to detect resistance to FQ (Conditional recommendation, Very low certainty in the evidence for test accuracy).				
Subgroup considerations	Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.				
Implementation considerations	Adoption of the MTBDR*sl* assay does not eliminate the need for conventional culture and DST capability. Despite good specificity of the MTBDR*sl* for the detection of resistance to FQs, culture and phenotypic DST is required to completely exclude resistance to this drug class. However, the demand for conventional culture and DST capacity may change, based on the prevalence of resistance to second-line anti-TB drugs in patients with confirmed RR-TB or MDR-TB. The following implementation considerations apply: • MTBDR*sl* cannot determine resistance to individual drugs in the class of fluoroquinolones. Phenotypic resistance to ofloxacin and levofloxacin is highly correlated with resistance conferring mutations detected by the MTBDRsl assay. Uncertainty remains about the susceptibility to moxifloxacin and gatifloxacin for such strains with mutations;				

	• Culture and phenotypic DST plays a critical role in the monitoring of patients' response to treatment and for detecting additional resistance to second-line drugs during treatment. Patients with false negative resistance results using the MTBDR*sl* can be identified and captured through treatment monitoring. Patients with false positive results might benefit from the addition of other drugs; • The availability of additional second-line drugs is critical.
Monitoring and evaluation	System of quality assurance is necessary.
Research priorities	Current recommendations on the MTBDR*sl* assay should not prevent or restrict further research on new rapid molecular DST tests, especially for assays that can be used as close as possible to where patients are initially diagnosed with RR-TB and MDR-TB and where treatment can be initiated. Further operational research on the MTBDR*sl* test should focus on the following priorities: • Develop and improved understanding of the correlation between the detection of resistance conferring mutations with phenotypic DST results and patient outcomes; • Develop improved knowledge of the presence of specific mutations detected with the MTBDR*sl* assay correlated with MICs for individual drugs within the class of fluoroquinolones; • Review evidence to confirm or revise different critical concentrations used in phenotypic DST methods; • Determine the limit of detection of MTBDR*sl* for the detection of heteroresistance; • Determine training, competency, and quality assurance needs; • Gather more evidence on the impact on appropriate MDR-TB treatment initiation and mortality; • Meet "Standards for Reporting Diagnostic accuracy studies" (STARD) for future studies; • Perform country-specific cost-effectiveness and cost-benefit analyses of MTDDR*sl* assay use in different programmatic settings.

PICO 4: Accuracy of MTBDR*sl* by indirect testing for detection of SLID resistance in patients with rifampicin-resistant or MDR-TB

	Judgement	Research evidence	Additional considerations
Problem	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	In 2014 WHO has estimated that 9.7% of the 480,000 cases of MDR-TB , were actually XDR TB, i.e. MDR TB with added resistance to at least one FQ and one SLID. Genotypic (molecular) methods have considerable advantages for scaling up programmatic management and surveillance of drug-resistant TB, offering speed of diagnosis, standardized testing, potential for high through-put, and fewer requirements for laboratory biosafety. Molecular tests for detecting drug resistance such as the MTBDR*sl* assay have shown promise for the diagnosis of drug-resistant tuberculosis (TB). The MTBDR*sl* assay incorporates probes to detect mutations within genes (*gyr*A and *rrs* for version 1.0 and, in addition, *gyr*B and the *eis* promoter for version 2.0), which are associated with resistance to the class of fluoroquinolones or the class of second-line injectable drugs (SLID).	Additional regions associated with resistance to FQ and SLIDs are included in the version 2.0 assay. Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.
Test accuracy	**How accurate is the test?** ○ Very inaccurate ○ Inaccurate ○ Accurate ○ Very accurate ● Varies ○ Don't know	In this review – data from the 16 studies, 1921 patients, reference standard: culture based DST *Test accuracy* MTBDR*sl* by direct testing for SLID: Sensitivity: 76.5% (95% CI: 63.3% to 86.0%) Specificity: 99.1% (95% CI: 97.3% to 99.7%) MTBDR*sl* by direct testing for Amikacin: Sensitivity: 84.9% (95% CI: 79.2% to 89.1%) Specificity: 99.1% (95% CI: 97.6% to 99.6%) MTBDR*sl* by direct testing for Kanamycin: Sensitivity: 66.9% (95% CI: 44.1% to 83.8%) Specificity: 98.6% (95% CI: 96.1% to 99.5%) MTBDR*sl* by direct testing for Capreomycin: Sensitivity: 79.5% (95% CI: 58.3% to 91.4%) Specificity: 95.8% (95% CI: 93.4% to 97.3%)	The accuracy varies with the different SLID. The variability is explained in part by the use of different drugs, critical concentrations , types of culture media in the reference standard and likely presence of *eis* resistance-conferring mutations in patients in Eastern European countries.
Desirable effects	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small ○ Moderate ● Large	The anticipated **desirable** effect is the correct diagnosis of SLID resistant cases (TP) as well as SLID susceptible cases (TN). MTBDR*sl* would correctly identify 32 cases out of 50 per 1000 individuals tested if the pre-test probability of TB is 5%. For 10-15% there would be 77 and 115 patients respectively (see table below). **Correct identification of SLID resistant cases should lead to higher cure rates, less sequelae to the individual patient, and less transmission in the community.** Similarly MTBDR*sl* would correctly identify 941 TB cases susceptible to SLID (TN) out of 950 per 1000 individuals tested if the pre-test probability of TB is 5%. For 10-15% prevalence's there would be 896 and 846 patients	**Desirable anticipated effects per drug:** Amikacin – Large desirable effects Capreomycin – Large

	○ Varies ○ Don't know	respectively (see table below). **Correct identification of SLID susceptible cases should lead to avoiding unnecessary treatment with additional drugs with increased risk of severe adverse events and greater costs.** The anticipated **undesirable** effect is the incorrect identification of an individual as a SLID susceptible or resistant case (FN or FP).	desirable effects Kanamycin – Large desirable effects
Undesirable effects	**How substantial are the undesirable anticipated effects?** ○ Large ○ Moderate ○ Small ○ Trivial ● Varies ○ Don't know	MTBDR*s*l would misclassify 12 cases as FN per 1000 individuals tested if the pre-test probability of TB with SLID resistance is 5%, and 23 to 35 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as SLID susceptible may have a potential increased risk of patient morbidity and mortality, and continued risk of community transmission of drug-resistant TB as well initiation of an MDR-TB regimen which includes a SLID with doubtful efficacy.** MTBDR*s*l had misclassified 9 cases as FP per 1000 individuals tested if the pre-test probability of TB with re sistance to SLID is 5%, and 8 cases under pre-test probabilities of 10-15%. **Incorrect identification of an individual as SLID resistant may lead to patient anxiety, possible delays in further diagnostic evaluation, prolonged and unnecessary treatment with drugs that may have additional serious adverse effects.** **Should MTBDRsl by indirect testing be used to diagnose SLID resistance in patients with RR or MDR TB?**	**Undesirable anticipated effects per drug:** Amikacin – Small undesirable effects Capreomycin and kanamycin – moderate undesirable effects Physicians should be guided by the MTBDR*s*l assay in their initial choice of an MDR-TB treatment regimen. Conventional phenotypic DST should be used in the follow-up evaluation of patients with a negative result especially in settings with a high pre-test probability for resistance to SLIDs.

Test result	Number of results per 1000 patients tested (95% CI)			Number of participants (studies)
	Prevalence 5%	Prevalence 10%	Prevalence 15%	
True positives (patients with SLID resistance)	38 (32 to 43)	77 (63 to 86)	115 (95 to 129)	575 (16)
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)	941 (924 to 947)	892 (876 to 897)	842 (827 to 847)	1346 (16)
False positives (patients incorrectly classified as having SLID resistance)	9 (3 to 26)	8 (3 to 24)	8 (3 to 23)	

Implications for the detection of SLID conferring mutations among RR-TB persons

TP: Test result suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen.
FP: Test result suggests modification of a WHO recommended MDR-TB regimen. Increased risk of serious adverse effects. Patient receive optimal regimen.

		FN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. Patient receive suboptimal regimen. No benefits. TN: Test result do not suggests modification of a WHO recommended MDR-TB regimen. No additional harms. Patient receive optimal regimen.	
Certainty of the evidence of the test accuracy	What is the overall certainty of the evidence of test accuracy? • Very low ○ Low ○ Moderate ○ High ○ No included studies	In this review the **risk of bias** was serious **Indirectness** was considered serious **Inconsistency** was considered serious for sensitivity and not serious for specificity **Imprecision** was considered not serious for sensitivity and specificity **Publication bias** – none for all studies (both direct and indirect testing).	**Quality of evidence for test accuracy** is: Sensitivity – very low quality of evidence Specificity – low quality of evidence
Certainty of the evidence of the test effects	**What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?** ○ Very low ○ Low ○ Moderate ○ High • No included studies	The test is labour-intensive and presents certain burden for the health worker. There is a need for appropriate infrastructure with separate rooms and biosafety requirements, which assumes a considerable investment. The burden and adverse effects are potentially insignificant for the patient.	
Certainty of the evidence of the management effects	**What is the overall certainty if the evidence of effects of the management that is guided by**	Ideally test results should guide management decisions, provided use of test is adopted by national policy. A positive test result should be sufficient for a patient to start treatment.	

	the test results? ○ Very low ○ Low ○ Moderate ○ High ● No included studies		
Certainty of the evidence of the test result/management	**How certain is the link between test results and management decisions?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	The link between test results and management decisions may be uncertain in various settings. In some occasions clinicians use empirical treatment for TB. In others capacity of health system may be insufficient to provide the patient with necessary treatment.	Turnaround time would be faster than for conventional DST The need for sample referral may cause delays
Certainty of effects	**What is the overall certainty of the evidence of effects of the test?** ○ Very low ● Low ○ Moderate ○ High ○ No included studies	This question is intended to summarize previous four questions on the certainty of the evidence.	Kanamycin and Capreomycin – low certainty Amikacin – moderate certainty
Values	**Is there important uncertainty about or variability in how much people value the**	There is no important uncertainty about or variability in how much people value the main outcomes.	

	main outcomes ? ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability ○ No known undesirable outcomes		
Balance of effects	**Does the balance between desirable and undesirable effects favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ● Probably favours the intervention ○ Favours the	Desirable (Amikacin, Kanamycin, Capreomycin) - Large, Large, Large Undesirable (Amikacin, Kanamycin, Capreomycin) - Small, Moderate, Moderate	Concern - FN Accuracy for detecting amikacin resistance is better than for capreomycin or kanamycin.

	intervention ○ Varies ○ Don't know		
Resources required	**How large are the resource requirements (costs)?** ○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ● Don't know	No research evidence was identified.	
Certainty of evidence of required resources	**What is the certainty of the evidence of resource requirements (costs)?** ○ Very low ○ Low ○ Moderate ○ High ● No included studies	No research evidence was identified.	

Cost effectiveness	**Does the cost-effectiveness of the intervention favour the intervention or the comparison?** ○ Favours the comparison ○ Probably favours the comparison ○ Does not favour either the intervention or the comparison ○ Probably favours the intervention ○ Favours the intervention ○ Varies ● No included studies	No research evidence was identified.	
Equity	**What would be the impact on health equity?** ○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	System incorporating molecular methods provides more equity.	

Aceptability	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	The test may be acceptable to be implemented in reference settings, where infrastructure and qualified staff to perform MTBDR*s*l exist. If MTBDR*s*l is implemented for first-line DST the MTBDR*s*l assay could be performed on the same culture isolate.	
Feasibility	**Is the intervention feasible to implement?** ○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Implementation of the test would require additional funding and technical support for the infrastructure upgrade, training of staff and procuring the equipment.	

Summary of judgments

	Judgement							**Implications**
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know	
Test accuracy	Very inaccurate	Inaccurate	Accurate	Very accurate		**Varies**	Don't know	
Desirable effects	Trivial	Small	Moderate	**Large**		Varies	Don't know	
Undesirable effects	Large	Moderate	Small	Trivial		**Varies**	Don't know	

	Judgement							Implications
Certainty of the evidence of test accuracy	**Very low**	Low	Moderate	High			No included studies	
Certainty of the evidence of test effects	Very low	Low	Moderate	High			**No included studies**	
Certainty of the evidence of management's effects	Very low	Low	Moderate	High			**No included studies**	
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**	
Certainty of effects	**Very low**	Low	Moderate	High			No included studies	
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			No known undesirable outcomes	
Balance of effects	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	**Probably favours the intervention**	Favours the intervention	Varies	Don't know	
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	**Don't know**	
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**	

	Judgement							Implications
Cost effectiveness	Favours the comparison	Probably favours the comparison	Does not favour either the intervention or the comparison	Probably favours the intervention	Favours the intervention	Varies	**No included studies**	
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know	
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know	

Conclusions

Should MTBDRsl by direct testing be used to diagnose SLID resistance in patients with RR or MDR TB?					
Type of recommendation	Strong recommendation against the intervention ○	Conditional recommendation against the intervention ○	Conditional recommendation for either the intervention or the comparison ○	Conditional recommendation for the intervention ●	Strong recommendation for the intervention ○
Recommendation	For patients with confirmed rifampicin-resistant TB or MDR-TB, the WHO guideline development group suggests using indirect testing of a culture of *M.tuberculosis* with the MTBDR*sl* assay as the initial test, over culture and phenotypic DST, to detect resistance to SLID (Conditional recommendation, Very low certainty in the evidence for test accuracy).				
Justification					
Subgroup consideration	Accuracy of version 2.0 assay is expected to be no worse than version 1.0 and should have improved sensitivity for detection of resistance for these drug classes.				
Implementation considerations	Adoption of the MTBDR*sl* assay does not eliminate the need for conventional culture and DST capability. Despite good specificity of the MTBDR*s*l for the detection of resistance to SLIDs, culture and phenotypic DST is required to completely exclude resistance to these drug classes. However, the demand for conventional culture and DST capacity may change, based on the prevalence of resistance to second-line anti-TB drugs in patients with confirmed RR-TB or MDR-TB. The following implementation considerations apply: • Mutations in some regions (e.g., the eis promoter region) may be responsible for causing resistance to one drug in a class more than other drugs within that class. The *eis* C14T mutation is associated with kanamycin resistance in strains from Eastern Europe; • Culture and phenotypic DST plays a critical role in the monitoring of patients' response to treatment and for detecting additional resistance to second-line drugs				

	during treatment. Patients with false negative resistance results using the MTBDR*s*/ can be identified and captured through treatment monitoring. Patients with false positive results might benefit from the addition of other drugs; • The availability of additional second-line drugs is critical.
Monitoring and evaluation	System of quality assurance is necessary.
Research priorities	Current recommendations on the MTBDR*s*/ assay should not prevent or restrict further research on new rapid molecular DST tests, especially for assays that can be used as close as possible to where patients are initially diagnosed with RR-TB and MDR-TB and where treatment can be initiated. Further operational research on the MTBDR*s*/ test should focus on the following priorities: • Develop and improved understanding of the correlation between the detection of resistance conferring mutations with phenotypic DST results and patient outcomes; • Develop improved knowledge of the presence of specific mutations detected with the MTBDR*s*/ assay correlated with MICs for individual drugs within the class of SLIDs; • Review evidence to confirm or revise different critical concentrations used in phenotypic DST methods; • Determine the limit of detection of MTBDR*s*/ for the detection of heteroresistance; • Determine training, competency, and quality assurance needs; • Gather more evidence on the impact on appropriate MDR-TB treatment initiation and mortality; • Meet "Standards for Reporting Diagnostic accuracy studies" (STARD) for future studies; • Perform country-specific cost-effectiveness and cost-benefit analyses of MTDDR*s*/ assay use in different programmatic settings.

3.9 Evidence-to-decision tables: High complexity reverse hybridization- based NAATs

PICO 8. Should high complexity hybridization based NAAT on isolates be used to diagnose PZA resistance in patients with microbiologically confirmed PTB, irrespective of resistance to RIF, pDST?	
POPULATION:	patients with microbiologically confirmed PTB, irrespective of resistance to RIF, pDST
INTERVENTION:	high complexity hybridization based NAAT on isolates

Assessment

Problem Is the problem a priority?		
Judgement	Research evidence	Additional considerations
○ No ○ Probably no ○ Probably yes ● Yes	Pyrazinamide (PZA) remains an important antibiotic for the treatment of both drug susceptible and drug resistant TB due to its unique ability to eradicate persisting bacilli and its synergistic properties with other antibiotics. While mono-resistance to PZA is rare, PZA resistance is strongly associated with MDR/RR-TB, with an	

○ Varies ○ Don't know	estimated 30-60% of MDR/RR-TB also resistant to PZA (Whitfield 2015, 2016). For people diagnosed with RR-TB, it is thus important to detect the presence of PZA resistance so that clinicians can make an informed decision on whether to include or exclude PZA in the treatment regimen.	

Test accuracy

How accurate is the test?

Judgement	Research evidence	Additional considerations
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	**Test accuracy** PZA LPA assay on isolates Sensitivity: 0.81 (95% CI: 0.75 to 0.86) Specificity: 0.98 (95% CI: 0.96 to 0.99)	

Desirable Effects

How substantial are the desirable anticipated effects?

Judgement	Research evidence	Additional considerations
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	(see table below)	**True positive** (PZA resistant): stop PZA and avid toxicity of ineffective drug. **True negative** (PZA susceptible): HCW increased confidence in regimen, patient inched likelihood of knowing they receive an effective regimen

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 8%	Prevalence 50%	Prevalence 90%		
True positives patients with PZA resistance	65 (60 to 69)	406 (377 to 429)	731 (679 to 772)	214 (7)	⨁◯◯◯ VERY LOW[a,b,c]
False negatives patients incorrectly classified as not having PZA resistance	15 (11 to 20)	94 (71 to 123)	169 (128 to 221)		
True negatives patients without PZA resistance	900 (888 to 907)	489 (483 to 493)	98 (96 to 99)	750 (7)	⨁⨁◯◯ LOW[a,b]

False positives patients incorrectly classified as having PZA resistance	20 (13 to 32)	11 (7 to 17)	2 (1 to 4)			

a. Studies suffered from selection bias, as they selected isolates with a wide range of different pncA mutations instead of a representative sample from a population. We downgraded one level for risk of bias.
b. Studies included do not directly address the review question. We downgraded one level for indirectness.
c. Burhan trial and Rienthong study are outliers for their sensitivities compared to the other studies. We downgraded one level for inconsistency.

Undesirable Effects

How substantial are the undesirable anticipated effects?

Judgement	Research evidence					Additional considerations
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	**Test result**	**Number of results per 1000 patients tested (95% CI)**			**№ of participants (studies)**	**Certainty of the evidence (GRADE)**	**False positive** (PZA resistance in case of susceptible): elimination of an effective drug from the regimen **False negative** (PZA resistance missed): HCW and patient believe regimen is highly effective even though an ineffective drug is included
		Prevalence 8%	**Prevalence 50%**	**Prevalence 90%**			
	True positives patients with PZA resistance	65 (60 to 69)	406 (377 to 429)	731 (679 to 772)	214 (7)	⨁◯◯◯ VERY LOW[a,b,c]	
	False negatives patients incorrectly classified as not having PZA resistance	15 (11 to 20)	94 (71 to 123)	169 (128 to 221)			
	True negatives patients without PZA resistance	900 (888 to 907)	489 (483 to 493)	98 (96 to 99)	750 (7)	⨁⨁◯◯ LOW[a,b]	

	False positives patients incorrectly classified as having PZA resistance	20 (13 to 32)	11 (7 to 17)	2 (1 to 4)		

a. Studies suffered from selection bias, as they selected isolates with a wide range of different pncA mutations instead of a representative sample from a population. We downgraded one level for risk of bias.
b. Studies included do not directly address the review question. We downgraded one level for indirectness.
c. Burhan trial and Rienthong study are outliers for their sensitivities compared to the other studies. We downgraded one level for inconsistency.

Certainty of the evidence of test accuracy

What is the overall certainty of the evidence of test accuracy?

Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	Certainty of the evidence of test accuracy: VERY LOW	

Certainty of the evidence of test's effects

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No direct evidence was considered here. Although a diagnostic study may not capture adverse effects as effectively as a treatment trial, if major adverse effects had occurred, it is likely that these would be reported.	

Certainty of the evidence of management's effects

What is the overall certainty of the evidence of effects of the management that is guided by the test results?

Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	There are no current observational or randomized controlled studies on patient-important outcomes of using the test. Testing for resistance to pyrazinamide is important ahead of starting treatment for Hr-TB (p.8)	very low - copy recommendations about PZA from the recmap

	For longer MDR-TB regimen pyrazinamide is counted as an effective agent only when DST results confirm susceptibility (p.29) Serious adverse events associated with PZA on long regimens occured a median of 8.8% (Table 3.3, p.31)	

Certainty of the evidence of test result/management

How certain is the link between test results and management decisions?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies		The management decisions may differ for rifampicin sensitive and rifampicin resistant patients, with more pronounced effects for rifampicin resistant population.

Certainty of effects

What is the overall certainty of the evidence of effects of the test?

Judgement	Research evidence	Additional considerations
● Very low ○ Low ○ Moderate ○ High ○ No included studies	This is the summary of the preceding judgements 5-8	very low for accuracy very low for treatment

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

Judgement	Research evidence	Additional considerations
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	**Patients** in high-burden TB settings value 1) getting an accurate diagnosis and reaching diagnostic closure (finally knowing what is wrong with me), 2) avoiding diagnostic delays as they exacerbate existing financial hardships and emotional and physical suffering and make patients feel guilty for infecting others (especially children), 3) having accessible facilities and 4) reducing diagnosis-associated costs (travel, missing work) as important outcomes of the diagnostic. (QES: moderate confidence) **The PZA LPA** addresses some preferences/values of laboratory staff and clinicians. It provides quicker results regarding PZA resistance, compared to other available methods (e.g. culture DST), can provide information on different concentration levels, and targets a drug that is widely used in first-line TB treatment. (Interview study)	

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the interventio n ○ Varies		The reference standard is pDST (the comparator) Clinical benefit has not been evaluated here. Clinical benefit would be superior in terms of speed of treatment.

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	No direct evidence from published studies regarding total resources required. Resource requirements will include the purchase of test kits (Genoscholar PZA: $16 USD/test kit consumables only), and the equipment which is available for $14,000USD. Operational costs are frequently several fold greater than test kit costs and are not accounted for, and will vary across settings. Unit test costs for Genotype MTBDRsl and MTBDRplus ranged from $23.46 to $108.70, with higher unit test costs coming from settings and countries such as South Africa and China and largely driven by higher staff wages and operational costs. Extrapolations from unit test costs using different LPAs should be done with caution and are not intended to be directly transferrable estimates. These indirect data do suggest that total unit test cost of the Genoscholar PZA LPA is likely several fold higher than unit test kit consumable cost of only $16USD. Total costs will vary depending on testing volume, numbers eligible for testing and prevalence of PZA resistance in the population. Budget impact will depend on current standard of care, diagnostic and care pathways and associated resource use.	

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ○ Moderate ○ High ● No included studies	Direct costs related to test kits and machinery are available while several important items related to resource use including staff time, overhead and operational costs associated with implementing Genoscholar PZA LPA have not been investigated. Differences in resource use between Genoscholar PZA LPA and existing approaches will vary across settings using different phenotypic and genotypic DST. Important variability exists in costs of staff time and operational costs, such as testing volume across settings.	

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the interventio n ○ Varies	No cost-effectiveness studies were identified using the Genoscholar PZA-TB II. Extrapolation of cost-effectiveness data from other line probe assays is not advised due to differences in diagnostic accuracy, resistance prevalence, and the testing and treatment cascade of care.	

Equity

What would be the impact on health equity?

Judgement	Research evidence	Additional considerations
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Lengthy diagnostic delays, underutilization of diagnostics, lack of TB diagnostic facilities at lower levels and too many eligibility restrictions, hamper access to prompt and accurate testing and treatment particularly for vulnerable groups. (QES: High confidence for CB NAAT, applicability to 3 index tests also confirmed in interview study) Staff and managers voiced concerns regarding sustainability of funding and maintenance, complex conflicts of interest between donors and implementers and concerns related to the strategic and equitable use of resources, which negatively affects creating equitable access to cartridge-based diagnostics. (QES: High confidence) Access to clear, comprehensible, and dependable information on what TB diagnostics are available to them and how to interpret results is a vital component to equity and represents an important barrier for patients (interview study). New treatment options need to be matched with new diagnostics: it is important to improve access to treatment based on new diagnostics, it is equally important to improve access to diagnostics for new treatment options (Interview study). The speed at which WHO guidelines are changing does not match the speed at which many country programmes are able to implement the guidelines. This translates into differential access to new TB diagnostics and treatment at an inter-country level (i.e. between countries that can and cannot quickly keep up with the rapidly changing TB diagnostic environment) as well at an intra-country level (i.e. between patients who can and cannot afford the private health system that is better equipped to quickly adopt new diagnostics and policies). (interview study)	

Acceptability

Is the intervention acceptable to key stakeholders?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Acceptability of **Hybridization-based Technology (PZA LPA)** is dependent on how well it performs on different samples, as laboratory staff question how well LPA methods work on smear-negative samples. If samples first need to be cultured in order to run PZA LPA this may undermine the benefits of this method's quicker TAT compared to phenotypic DST for PZA. Acceptability also depends on how well it actually detects mutations specific to PZA resistance and clincians and laboratory staff may require further clarification/justification in some settings as to why this specific DST drug test is being prioritized, as it is not currently part of routine DST. Specific feasibility challenges (training and infrastructure requirements, sample quality result interpretation system) and general feasibility challenges (as identified in interview study and QES respectively), and accumulated delays risk undoing the added value/benefits as identified by the users (avoiding delays, drug resistant information). (combination QES and interview study)	

Feasibility

Is the intervention feasible to implement?

Judgement	Research evidence	Additional considerations
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	**Feasibility of PZA LPA is challenged by the significant training and laboratory infrastructure required to implement this method, including proper sample handling** and quality sample. **Feasibility for this test also hinges on the availability of an automated interpretation system**, as it is difficult to interpret.(interview study).	

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable effects	Trivial	Small	**Moderate**	Large		Varies	Don't know
Undesirable effects	Large	Moderate	**Small**	Trivial		Varies	Don't know
Certainty of evidence of the test accuracy	**Very low**	Low	Moderate	High			No included studies
Certainty of evidence of test's effects	Very low	Low	Moderate	High			**No included studies**
Certainty of evidence of management's effects	**Very low**	Low	Moderate	High			No included studies

	Judgement						
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty of effects	**Very low**	Low	Moderate	High			No included studies
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the interventio n	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	Very low	Low	Moderate	High			**No included studies**
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the interventio n	Favors the interventio n	Varies	**No included studies**
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In people with microbiologically confirmed TB, hybridization-based technology may be used on isolates for detection of pyrazinamide resistance (rather than culture based phenotypic DST) (conditional recommendation; very low certainty of evidence for diagnostic accuracy)

Remarks: Recommendation includes people with RR/MDRTB and INH mono Resistant TB

Subgroup considerations

no special considerations given that both tests depend on the availability of isolates (no subgroup considerations for PLHIV or children)

Implementation considerations

Infrastructure, lab and clinical training (expereienced lab for LPA) on interpretation of

results Can only be implemented where culture facilities are available

Quality control and assurance required

Equipment maintenance

Sample transport conditions

Monitoring and evaluation

Quality control and assurance required

Results feedback

Research priorities

Research: accuracy on direct specimen testing; further research on genotype/phenotype/clinical outcome relationship

-Impact of test result on treatment decisions- would also be a research priority

research on testing of sputum and EPTB and specimens in general (which should include PLHIV and children) - smear positive and negative populations

Direct evidence of testing on people important outcomes (which should include PLHIV and children)

Values of outcomes, feasibility, acceptability, equity and economic

evaluation research on how to interpret the index test when compared to

sequencing

3.10 Targeted Next-Generation Sequencing (NGS)

Question

Should targeted Next Generation Sequencing as an initial test be used to diagnose drug resistance to rifampicin (RIF), isoniazid (INH), flouroquinolones (FQs), pyrazinamide (PZA), and ethambutol (EMB) in patients with bacteriologically confirmed pulmonary TB disease?	
Population:	Patients with bacteriologically confirmed pulmonary TB disease
Intervention:	Targeted next-generation sequencing (tNGS) technologies **for** use on respiratory samples to detect resistance to rifampicin (RIF), isoniazid (INH), flouroquinolones (FQ), pyrazinamide (PZA), ethambutol (EMB) compared to culture-based phenotypic drug sensitivity testing (DST)
Purpose of the test:	Used as a test for TB drug resistance
Role of the test:	An initial test, following bacteriological diagnosis
Linked treatments:	Correct treatment for drug sensitive or drug resistant TB
Anticipated outcomes:	Improved treatment outcomes based on drug resistance detected
Setting:	TB programmes worldwide
Perspective:	Public health perspective
Subgroups:	N/A
Conflict of interests:	All guideline panel members completed declaration of interest forms

Assessment

Judgement	**Research evidence**	**Additional considerations**
Problem **Is the problem a priority?**		
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Drug-resistant tuberculosis (DR-TB) is a major threat to global TB control. There are estimated to be over 10 million cases and 1.5 million deaths of TB annually, including 450,000 cases of DR-TB, only a third of whom are diagnosed and treated appropriately (WHO Global TB Report 2022). WHO recommendations include initial tests for resistance to RIF and INH among all TB patients and tests for resistance to fluoroquinolones (FQs) among those with RIF-resistant or RIF-susceptible INH-resistant TB. Phenotypic drug susceptibility testing (pDST), remains the reference standard for drug resistance detection to most drugs. It is a culture-based approach that requires several weeks for results to be available and performed at specialized sites with limited access. In recent years, nucleic-acid amplification tests (NAATs) have offered options for molecular detection of drug resistance, including line-probe assays and rapid molecular tests that can detect drug resistance in a fraction of the time required for culture-based methods. However, they have limitations in the number of drugs they can test for resistance, the ability to distinguish mutations with differing resistance potential and how quickly they can incorporate new data on genetic information about drug resistance.	The group noted that this question is slightly less of a priority than PICO 2 for certain drugs, given that in this population there are currently available rapid molecular tests that give resistance information for some drugs (RIF, INH). However it is also noted that,

	The recent introduction of new drugs and repurposing of existing antimicrobial agents for the treatment of TB have generated new TB treatment regimens at a comparatively rapid rate, providing improvements to treatment options, outcomes, and quality of life during treatment among DR-TB patients. However, as resistance to these new and repurposed drugs in the community gradually increases, there is concern about the lack of options for rapid detection of resistance to these drugs by currently approved methods. Gene sequencing technologies provide an option for rapid, accurate genetic analysis and detection of mutations indicating resistance in a fraction of the time required for culture-based methods for detecting resistance. Recently several commercial “End-to-End Solutions” for targeted next-generation sequencing (tNGS) for detection of drug resistance have become available that promise a higher throughput, a significantly faster time to result, and greater accuracy across more TB drugs than current WHO-recommended molecular methods for DST, and offer the potential to rapidly assimilate new information on genetic markers for resistance as they become known. The question to the GDG is to evaluate the available evidence on tNGS technologies and to generate guidance on their use in programmatic management of DR-TB globally.	given there are other TB diagnostics products in development that do not always include resistance testing, this will become more important in the future.
Test accuracy How accurate is the test?		
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	**Test accuracy** RIF (comp): Sensitivity: 0.93 (95% CI: 0.87 to 0.99) Specificity: 0.96 (95% CI: 0.89 to 1.00) INH (pDST): Sensitivity: 0.96 (95% CI: 0.93 to 0.99) Specificity: 0.97 (95% CI: 0.95 to 0.99) LFX (pDST): Sensitivity: 0.94 (95% CI: 0.88 to 1.00) Specificity: 0.96 (95% CI: 0.93 to 0.99) MFX (pDST): Sensitivity: 0.96 (95% CI: 0.92 to 0.99) Specificity: 0.96 (95% CI: 0.93 to 1.00) PZA (comp): Sensitivity: 0.85 (95% CI: 0.80 to 0.90) Specificity: 0.94 (95% CI: 0.92 to 0.96) EMB (comp): Sensitivity: 0.88 (95% CI: 0.82 to 0.94) Specificity: 0.94 (95% CI: 0.91 to 0.97)	Judgement for all drugs is "accurate". The group notes that, based on the test accuracies seen here, all drugs should be classified as "accurate" or "very accurate" compared to other available tests. However, it is noted that the relatively high indeterminate rates for the test across all drugs impact the group's considerations of how accurate the test is (see "Undesirable effects"). The group also notes that this

		data is limited to processed sputum samples; it is not yet known how well the test performs on raw sputum or other specimen types.

Desirable Effects
How substantial are the desirable anticipated effects?

○ Trivial
○ Small
○ Moderate
● Large
○ Varies
○ Don't know

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 15%		
True positives patients with drug resistance to rifampin (RIF) (composite)	19 (17 to 20)	93 (87 to 99)	140 (131 to 149)	1436 (9)	⨁⨁⨁◯ Moderate [a]
False negatives patients incorrectly classified as not having drug resistance to rifampin (RIF) (composite)	1 (0 to 3)	7 (1 to 13)	10 (1 to 19)		
True negatives patients without drug resistance to rifampin (RIF) (composite)	941 (872 to 980)	864 (801 to 900)	816 (757 to 850)	271 (7)[b]	⨁⨁◯◯ Low[a,c]
False positives patients incorrectly	39 (0 to 108)	36 (0 to 99)	34 (0 to 93)		

The group agrees that the magnitude of desirable effects varies across country settings with differing prevalences of drug resistance, but the group agreed that across all drugs considered the desirable effects could be considered "large".

The group notes, however, that desirable effects will be smaller for patients with lower bacillary loads as it will lead to higher indeterminate rates (see "Undesirable effects"). Also the level of desirable versus undesirable effects will vary based on the prevalence of resistance to all the drugs under consideration.

classified as having drug resistance to rifampin (RIF) (composite)					

NOTE: A True Positive test indicates that the patient is correctly treated with appropriately modified regimen for resistance pattern; risk of treatment failure or developing further resistance are minimized. A True Negative test indicates that the patient is correctly treated with appropriate regimen; treatment burden minimized.

a. All studies enriched for samples that were rifampicin resistant. Prevalence of resistance to rifampicin (composite) across data used in the model was 83% (CI 81% to 85%). However, prevalence should not significantly impact sensitivity or specificity, therefore not downgraded for bias, just for indirectness.
b. 115 observations from ONT dropped by model as variable 'duplicate=2' (i.e. ONT) predicts the outcome perfectly (115 TN results)
c. 95% confidence interval for specificity spans >10%, therefore the result was downgraded for imprecision.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 2%	Prevalence 10%	Prevalence 15%		
True positives patients with drug resistance to isoniazid (INH) (pDST)	19 (19 to 20)	96 (93 to 99)	144 (140 to 149)	1440 (12)	⨁⨁⨁◯ Moderate[a]
False negatives patients incorrectly classified as not having drug resistance to	1 (0 to 1)	4 (1 to 7)	6 (1 to 10)		

isoniazid (INH) (pDST)					
True negatives patients without drug resistance to isoniazid (INH) (pDST)	951 (931 to 970)	873 (855 to 891)	825 (808 to 842)	517 (12)	⊕⊕⊕◯ Moderate[a]
False positives patients incorrectly classified as having drug resistance to isoniazid (INH) (pDST)	29 (10 to 49)	27 (9 to 45)	25 (8 to 42)		

d. All studies enriched for samples that were rifampicin resistant. Prevalence of resistance to isoniazid across data used in the model was 74% (CI 72% to 76%). However, prevalence should not significantly impact sensitivity or specificity, therefore not downgraded for bias, just for indirectness.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 5%	Prevalence 10%		
True positives patients with drug resistance to levofloxacin (LFX) (pDST)	9 (9 to 10)	47 (44 to 50)	94 (88 to 100)	654 (6)	⊕⊕◯◯ Low[a,b]
False negatives patients	1 (0 to 1)	3 (0 to 6)	6 (0 to 12)		

incorrectly classified as not having drug resistance to levofloxacin (LFX) (pDST)					
True negatives patients without drug resistance to levofloxacin (LFX) (pDST)	950 (921 to 980)	912 (884 to 941)	864 (837 to 891)	913 (7)	⨁⨁⨁◯ Moderate [a]
False positives patients incorrectly classified as having drug resistance to levofloxacin (LFX) (pDST)	40 (10 to 69)	38 (9 to 66)	36 (9 to 63)		

e. All studies enriched for samples that were rifampicin resistant. Prevalence of resistance to Levofloxacin across data used in the model was 42% (CI 39% to 44%). However, prevalence should not significantly impact sensitivity or specificity, therefore not downgraded for bias, just for indirectness.
f. One of the larger studies performed much worse for sensitivity and therefore the result was downgraded for inconsistency.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 5%	Prevalence 10%		
True positives patients with drug	10 (9 to 10)	48 (46 to 50)	96 (92 to 99)	652 (6)	⨁⨁⨁◯ Moderate [a]

resistance to moxifloxacin (MFX) (pDST)					
False negatives patients incorrectly classified as not having drug resistance to moxifloxacin (MFX) (pDST)	0 (0 to 1)	2 (0 to 4)	4 (1 to 8)		
True negatives patients without drug resistance to moxifloxacin (MFX) (pDST)	950 (921 to 990)	912 (884 to 950)	864 (837 to 900)	921 (8)	⨁⨁⨁◯ Moderate [a]
False positives patients incorrectly classified as having drug resistance to moxifloxacin (MFX) (pDST)	40 (0 to 69)	38 (0 to 66)	36 (0 to 63)		

g. All studies enriched for samples that were rifampicin resistant. Prevalence of resistance to Moxifloxacin across data used in the model was 41% (CI 39% to 44%). However, prevalence should not significantly impact sensitivity or specificity, therefore not downgraded for bias, just for indirectness.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 3%	Prevalence 10%		

True positives patients with drug resistance to pyrazinamide (PZA) (composite)	9 (9 to 9)	26 (26 to 28)	88 (85 to 92)	346 (3)	⊕⊕⊕◯ Moderate [a]
False negatives patients incorrectly classified as not having drug resistance to pyrazinamide (PZA) (composite)	1 (1 to 1)	4 (2 to 4)	12 (8 to 15)		
True negatives patients without drug resistance to pyrazinamide (PZA) (composite)	980 (960 to 990)	960 (941 to 970)	891 (873 to 900)	269 (3)	⊕⊕⊕◯ Moderate [a]
False positives patients incorrectly classified as having drug resistance to pyrazinamide (PZA) (composite)	10 (0 to 30)	10 (0 to 29)	9 (0 to 27)		

h. All studies enriched for samples that were rifampicin resistant. Prevalence of resistance to Pyrazinamide (composite) across data used in the model was 56% (CI 52% to 60%). However, prevalence should not significantly impact sensitivity or specificity, therefore not downgraded for bias, just for indirectness.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 3%	Prevalence 10%		

	True positives patients with drug resistance to ethambutol (EMB) (composite)	10 (9 to 10)	29 (28 to 29)	96 (94 to 98)	432 (4)	⊕⊕◯◯ Low[a,b]
	False negatives patients incorrectly classified as not having drug resistance to ethambutol (EMB) (composite)	0 (0 to 1)	1 (1 to 2)	4 (2 to 6)		
	True negatives patients without drug resistance to ethambutol (EMB) (composite)	980 (970 to 990)	960 (951 to 970)	891 (882 to 900)	268 (4)[c]	⊕⊕◯◯ Low[a,b]
	False positives patients incorrectly classified as having drug resistance to ethambutol (EMB) (composite)	10 (0 to 20)	10 (0 to 19)	9 (0 to 18)		

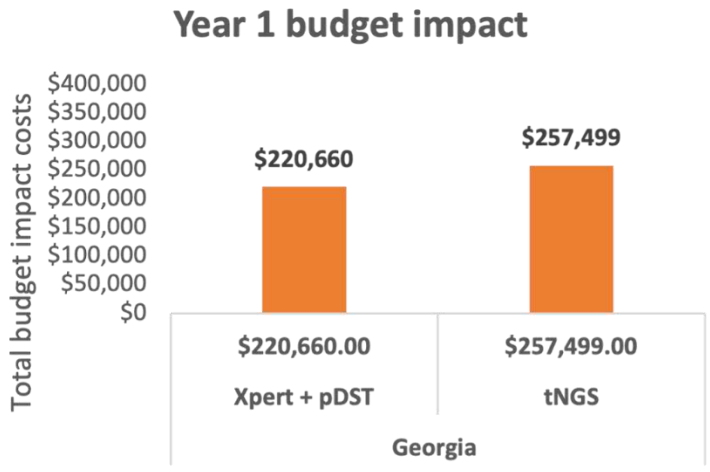
i. All studies enriched for samples that were rifampicin resistant. Prevalence of resistance to ethambutol (composite) across data used in the model was 62% (CI 58% to 65%). However, prevalence should not significantly impact sensitivity or specificity, therefore not downgraded for bias, just for indirectness.
j. Different samples used for tNGS and reference test
k. The model does not control for rifampicin resistance as this variable was collinear in the original model.

	I.	
Undesirable Effects How substantial are the undesirable anticipated effects?		
○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know	See tables above for numbers of false positive and false negative test results expected for each drug (see "Desirable Effects"). NOTE: • A False Positive test may result in incorrect and inappropriate treatment regimens, and put people with TB at risk of unnecessary adverse effects. • A False Negative test may result in people with TB receiving incorrect treatment and causing delays in receiving appropriate treatment, putting them at risk of treatment failure, mortality, developing further resistance, and transmission of DR-TB to others. Indeterminate rates: RIF (comp) = 12.0% (10.5-13.6) INH (pDST) = 14.6% (13.0-16.2%) LFX (pDST) = 9.2% (7.8-10.7%) MFX (pDST) = 9.3% (7.9-10.9%) PZA (comp) = 17.6% (14.6-20.8%) EMB (comp) = 16.3% (13.5-19.2%)	The group notes that the high indeterminate rates affect this decision significantly, as the lack of clinically actionable results from the test from 10 to 20% of the time reduces its clinical utility and increases the effective per-patient test cost. The group also notes that, in some settings with lower prevalences of drug resistance, the false-positive results will outnumber the true-positive ones, with significant clinical implications.
Certainty of the evidence of test accuracy What is the overall certainty of the evidence of test accuracy?		
○ Very low ● Low ○ Moderate ○ High ○ No included studies	Certainty of test accuracy: - MODERATE for INH, PZA, MFX - LOW for RIF, EMB, LFX	Given the range of data included in this combined PICO, the composite measure of certainty of evidence for test accuracy is LOW.
Certainty of the evidence of test's effects What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?		

○ Very low ○ Low ○ Moderate ○ High ● No included studies	No included studies. tNGS is an "in vitro" test, and therefore no adverse effects are expected for patients beyond discomfort from producing sputum. tNGS is likely to have a faster turn-around time than culture-based tests as it can be completed in a few days' time versus several weeks to months required for culture growth. However, placement of the test in the health system and batching/multiplexing of the technology will impact the turn-around time experienced in a particular setting.	The group notes that for some populations like children, sputum is not easy to produce for testing, requiring more burdensome sample types like gastric aspirates. There is also a note that for patients with low bacillary load there may be a need to collect more than one sample.
Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?		
○ Very low ○ Low ○ Moderate ● High ○ No included studies	Treatment regimen depends on the results of drug susceptibility testing. **RIF-susceptible:** · New patients with pulmonary TB should receive a regimen containing 6 months of rifampicin: 2HRZE/4HR **(strong recommendation, high certainty of evidence).** · People aged 12 years or older with drug-susceptible pulmonary TB, may receive a 4-month regimen of isoniazid, rifapentine, moxifloxacin and pyrazinamide **(conditional recommendation, moderate certainty of evidence).** · In patients with confirmed rifampicin-susceptible, isoniazid-resistant tuberculosis, treatment with rifampicin, ethambutol, pyrazinamide and levofloxacin is recommended for a duration of 6 months. (**Conditional recommendation, very low certainty evidence).** **RIF-resistant:** · WHO suggests the use of a 6-month treatment regimen, composed of Bedaquiline, Pretomanid, Linezolid (600 mg) and Moxifloxacin (BPaLM), rather than the 9-month or longer (18-month) regimens in MDR/RR-TB patients (**Conditional recommendation, very low certainty evidence**) · WHO suggests the use of the 9-month all-oral regimen rather than longer (18-month) regimens in patients with MDR/RR-TB and in whom resistance to fluoroquinolones has been excluded. (**Conditional recommendation, very low certainty evidence**) · In multidrug- or rifampicin-resistant tuberculosis (MDR/RR-TB) patients on longer regimens, all three Group A agents and at least one Group B agent should be included to ensure that treatment starts with at least four TB agents likely to be effective, and that at least three agents are included for the rest of the treatment if bedaquiline is stopped. If only one or two Group A agents are used, both Group B agents are to be included. If the regimen cannot be composed with agents from Groups A and B alone, Group C agents are added to complete it. (**Conditional recommendation, very low certainty evidence**)	The evidence base underpinning all TB treatment recommendations is known and captured here. The group notes that there is uncertainty in how much confidence clinicians will have in the results of this new test.

Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No included studies for this question about whether those tested access management (linkage to care). If clinicians receive the test results, there is a high likelihood that the test results will be used, and treatment decisions will be based on the test results for resistance detection. TB programmes have processes in place to link laboratory test results with clinicians treating patients. Linkage of laboratory results to patients in a timely manner impacts on loss to follow-up of patients and retention in care. tNGS is likely to have a faster turn-around time than culture-based tests as it can be completed in a shorter time space compared to several weeks to months required for culture growth. However, placement of the test in the health system and batching/multiplexing of the technology will impact the turn-around time experienced in any particular setting. In most contexts, TB medicines are available. The availability of TB medicines will impact the ability to treat patients according to the test results. <u>Indirect evidence from a systematic review:</u> "Use of line probe assays (LPAs) compared to pDST reduced diagnostic delay by 40.09 days (95% CI 26.82–53.37) and treatment initiation delay by 45.32 days (95% CI 30.27–60.37) in comparison to any culture DST methods. " https://doi.org/10.1186/s12879-022-07855-9	The group notes that there are many unknowns here - it is not yet known what the uptake would be for this technology if approved. In addition, while tNGS is expected to have a faster turnaround time to results compared to culture-based DST methods, there are many health system factors that will affect how it will impact on patient care. Experiences with implementation of other rapid DST options (mWRDs, LPA) have illustrated how health system issues affect turnaround time for supposedly rapid tests.
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
○ Very low ○ Low ○ Moderate ○ High ● No included studies	There are no available comparative intervention studies on the effectiveness of tNGS compared to the current standard of care on patient-important outcomes. <u>Indirect evidence from other studies with other tests that result in faster turnaround time to results:</u> A comparative cohort study from China found that patients with early available molecular DST results had a more rapid culture conversion (aHR1.94 95% CI: 1.37-2.73; median,12 vs 24 months, respectively; $P < 0.001$) and a higher rate of treatment success (68% vs 47%, $P < 0.01$) (1). Similarly, a pragmatic trial from Brazil showed that compared to the MGIT	It is noted that there are no specific comparative effectiveness data for this question.

	group, culture conversion after 6 months was higher for Xpert in arm 1 (90.9% vs 79.3%, p=0.39) and LPA in arm 2 (80.0% vs 83.0%, p=0.81) (2). In contrast, a study from Ethiopia did not show any difference in treatment outcomes between Xpert, LPA and MGIT used for detection drug-resistant TB (3). References 1. Shi W, Davies Forsman L, Hu Y, Zheng X, Gao Y, Li X, et al. Improved treatment outcome of multidrug-resistant tuberculosis with the use of a rapid molecular test to detect drug resistance in China. International Journal of Infectious Diseases. 2020;96:390-7.doi.org/10.1016/j.ijid.2020.04.049 2. Kritski A, Oliveira MM, Almeida IN de, Ramalho D, Andrade MK de N, Carvalho M, et al.. Clinical Impact of the Line Probe Assay and Xpert® MTB/RIF Assay in the Presumptive Diagnosis of Drug-Resistant Tuberculosis in Brazil: A Pragmatic Clinical Trial. Rev Soc Bras Med Trop [Internet]. 2022;55:e0191–2021. Available from: https://doi.org/10.1590/0037-8682-0191-2021 doi.org/10.1590%2F0037-8682-0191-2021 3. Kassa GM, Merid MW, Muluneh AG, Wolde HF. Comparing the impact of genotypic based diagnostic algorithm on time to treatment initiation and treatment outcomes among drug-resistant tuberculosis patients in Amhara region, Ethiopia. PLOS ONE. 2021;16(2):e0246938. doi.org/10.1371/journal.pone.0246938	
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	WHO commissioned a qualitative evidence synthesis (QES) which did not find any included studies. A primary qualitative study of user experiences (technicians and implementers in the FIND studies) was conducted but it did not explicitly explore how much people value the outcomes or their preferences with respect to the intervention compared to the comparison. Indirect evidence: A qualitative evidence synthesis of recipient and provider perspectives on rapid molecular tests for TB and drug resistance found that people with tuberculosis valued reaching diagnostic closure with an accurate diagnosis, avoiding diagnostic delays, and keeping diagnostic-associated costs low. Similarly, healthcare providers valued aspects of accuracy and the resulting confidence in low-complexity NAAT results, rapid turnaround times, and low costs to people seeking a diagnosis. https://doi.org/10.1002/14651858.CD014877.pub2	The group agrees that there is probably no important uncertainty and that patients likely value an accurate test with rapid results.
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the interventio	The balance of desirable and undesirable effects **probably favors** the intervention versus the comparison.	The group notes that areas with higher prevalence of drug resistances (>5%) will have a greater benefit from TNGS compared to phenotypic DST.

n or the comparison ● Probably favors the interventio n ○ Favors the interventio n ○ Varies ○ Don't know		The group also notes that modeling suggests that the prevalence of INH resistance drives the balance, with a higher INH resistance prevalence favoring TNGS more compared to phenotypic DST.
Resources required How large are the resource requirements (costs)?		
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know	Literature reviews suggest unit test costs for tNGS are consistently higher than phenotypic drug susceptibility testing (pDST). The largest cost components were test kit reagents, ancillary consumables, and costs associated with sequencing. Key cost drivers include specific sequencer used, depth and breadth of coverage, inefficiencies in initial sample runs, the economics of scale via batching or cross-batching, operational efficiency, availability of trained personnel, sequencers being used to full capacity, bulk purchases, and complexity of the infectious pathogen. Empirical costing estimates for tNGS unit cost per sample for Deeplex Myc-TB Genoscreen tNGS ranged from: (unit costs includes consumables, equipment, staffing and overgead where available, costs assume tNGS testing for all drugs) **- $134 to $257 in South Africa,** **- $120 to $198 in Georgia and** **- $121 to $175 in India,** depending on patient volume, batching and negotiated tNGS kit cost. Budget Impact assessment suggests tNGS will be marginally more costly in the Georgian setting (PICO1) compared with Xpert + pDST  	The group notes that the question of required costs depends on the drug under consideration and what alternatives exist for DST, as well as if costs for setting up the systems for tNGS or DST are considered or not. For example, for RIF , there are many tests available that are likely less expensive. Likewise for INH there are other rapid DST options. For other drugs with fewer other rapid DST options, the resources required for tNGS and pDST are similar. However the group also notes that the costs of the test for tNGS produce results for all drugs at once, as compared to drug-specific

		tests. The group notes that initial costs for the test (eg during the initial set-up period) are likely to be quite large, but those are time-limited investments. The group feels that this is a research gap and there is a need for Health Technology Assessment studies for tNGS in specific country contexts.
Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
● Very low ○ Low ○ Moderate ○ High ○ No included studies	The systematic review identified 10 manuscripts with very limited economic data and no cost-effectiveness analyses. Data on total implementation costs are also limited, with only one budget impact assessment retrieved in the systematic review and one commissioned for the GDG meeting. Several key scenarios were assessed in the empirical costing to derive unit cost ranges and account for underlying uncertainty.	The group notes that the certainty of evidence in this space is very low, given that the knowledge of the costs and budget impact of these tests is very limited.
Cost effectiveness Does the cost-effectiveness of the intervention favor the intervention or the comparison?		
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors	Cost-effectiveness modelling was commissioned and found that using tNGS as the initial test for drug resistance in patients with bacteriologically confirmed pulmonary TB in the Georgian setting, where the standard of care was mWRD (GeneXpert) followed by phenotypic DST among rifampicin resistant individuals, to be cost-effective (ICER= $9261; 95% UR: of $5,258-$32,040) with 80% of simulated iterations below the willingness to pay threshold of $15,069 for Georgia. In sensitivity analysis, prevalence of INH mono-resistance and indeterminate rate of tNGS were important drivers of cost-effectiveness. tNGS was not cost-effective when INH mono-resistance was less than 9% and when indeterminate rate was greater than 26%. Implementing tNGS as the initial test for DST may be beneficial in settings with high INH resistance and where DST of Group A second line drugs are not being performed universally. This model explored the cost-effectiveness of tNGS when drugs susceptibility testing (DST) is being done for only rifampicin, isoniazid and fluoroquinolone.	The group notes again that the specific cost effectiveness of the tNGS technology depends heavily on the country context. The modeling results that the group has to consider are for the Georgia context only.

the intervention ● Varies ○ No included studies		
Equity What would be the impact on health equity?		
○ Reduced ○ Probably reduced ○ Probably no impact ○ Probably increased ○ Increased ● Varies ○ Don't know	From a qualitative study of relevant stakeholders, the following considerations on the impact of tNGS on equity were found: **- Centralized vs decentralized placement may have equity implications for access.** Given high-level specialised laboratory infrastructure, specialized human resources and technical complexity, tNGS technology is only suitable for placement at centralized laboratories. This may have equity access considerations as it may mean less access for some regions of a country without specialized central laboratories. **- Affordability and cost-effectiveness are major concerns:** There was a major concern about financial costs of the tNGS technology and the affordability for LMICs. Participants were worried about not only the cost of the equipment, but also the costs of ongoing specialist supplies, especially for reagents, as well as the cost of maintaining equipment. They noted that costing calculations should be comprehensive and include the cost of specialist consumables, extra general laboratory consumables, and additional infrastructure needs (such as extra space, temperature control, and internet connectivity). There were concerns that cost-effectiveness calculations should also include an assessment of the impact of the use of tNGS testing on improving TB outcomes in comparative studies. **- The MDR-TB case burden of a country could influence equitable access at centralized level**s. In some settings with high caseloads, the tNGS technology in central laboratories may not be sufficient for processing large caseloads in good time, and in settings with low caseloads, waiting for sufficient samples to batch will also cause delays.	The group notes that tNGS has the potential to increase equity compared to phenotypic DST in many settings, especially considering the shorter turnaround time compared to phenotypic DST, the capacity for tNGS to conduct comprehensive DST to more fully guide treatment, and the potential scalability of tNGS. However, compared to other rapid tests available for DST for certain drugs, tNGS probably does not improve equity. The group notes that tNGS is a centralized test, which may result in decreasing access for some patients. Also the group notes that the high indeterminate

		rates, especially among paucibaccilary samples, will disproportionately affect people living with HIV, children, and other patients with low bacterial load or patients being diagnosed with non sputum-based sample types. Some countries may quickly gain access to this technology and some will not, so this may increase global inquity. To truly undertand the impact of TNGS on equity in each specific country context, each country will need to explore.
Acceptability Is the intervention acceptable to key stakeholders?		
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	From a qualitative study of relevant stakeholders (laboratory staff and management who were involved with testing tNGS platforms in the three FIND trial sites, India, Georgia, and South Africa) **acceptability of tNGS technology was high.** There was an overwhelmingly positive sentiment for the potential utility of tNGS, and it was seen as a 'major advancement' in molecular MDR TB diagnostics. - The main reasons for the **high level of acceptability were the comprehensiveness** (resistance diagnosis for more drugs and for new and repurposed drugs), **the convenience of using sputum sample** (as compared to culture samples), **and the rapidness** (quick results compared to phenotypic testing times; 3-5 days as compared to 4-6 weeks). - There was also the sense that there is **a good window of opportunity for the utility of tNGS technology;** that the technology is arriving at the right time given that resistance to newer TB drugs is likely to increase as use if these drugs become routine.	The group notes that the data suggests the test is acceptable to laboratory personnel, including manageres and staff conducting the assay.
Feasibility Is the intervention feasible to implement?		

○ No ○ Probably no ○ Probably yes ○ Yes ● Varies ○ Don't know	From a qualitative study of stakeholders from the FIND studies, the following considerations regarding the feasibility of TNGS were reported: **- Start-up and setting up challenges:** There were multiple starting and setting up problems. Some related to the newness of the technology and the trial setting, problems with importing technology and specialist supplies, problems related to the absence of in-country technical assistance for problem-solving, and need for more hands-on training practice. **- High technical complexity of the test is a challenge:** tNGS technology was viewed as a high-complexity molecular test that was technically challenging. For example, preparing the sample for sequencing involves multiple steps that require attention to detail, precision, and little room for error. The complexity of the library preparation phase was more particular for the Deeplex platform. However, both the Deeplex and the Nanopore platforms were thought to have different pros and cons in terms of complexity. Both platforms were thought to have insufficient opportunities for early error recognition and correction, increasing the risk of failed runs. **- Specialized laboratory infrastructure and human resources are required which are potentially challenging**: As tNGS is a molecular-based testing platform, the platform requires highly specialised laboratory infrastructure that includes multiple rooms to prevent contamination and specialized cold storage facilities. Highly specialized molecular/medical scientists are needed to perform the tests. In these LMIC settings, such specialized laboratory infrastructure and staff may only be available at centralized laboratories and not necessarily at regional laboratories. **- Specialist requirements for operating the test are potentially challenging:** In addition to highly specialized laboratory infrastructure and staff, the testing technology also requires an uninterrupted supply of electricity, high internet connectivity, high computing capacity, clean water, and temperature controls - requirements that may pose challenges in some LMIC settings. **- Supply chain challenges were an obstacle:** A major concern was the supply chain challenges - procurement bottlenecks and delays jeopardized continuous access to specialist supplies. **- Data management and storage requirements presented challenges:** There were concerns that data analysis and data storage requirements were not fully developed, including systems for backing up data, data ownership and data security considerations. Consideration is needed for how tNGS and routine laboratory information systems would be interlinked. **- Continuous updating of the WHO mutations reference library would be required:** There is the sense that the usefulness of the tNGS technology is dependent on the informational support provided by the WHO mutations reference library, which allows for meaningful interpretation of resistance data; and thus, there is a need for the WHO reference library to be continuously updated. **- There are different feasibility concerns for the different tNGS platforms**: The overall sentiment that is that all three tNGS platforms needed to be further developed before being fully ready for operational use, some more than others. The high level of technical complexity of the sample preparation stages (mainly the library preparation stage) was considered a key challenge for the Deeplex platform. The need for improved computer analysis and storage capacity was a challenge for the Oxford Nanopore (ONP) platform. However, both required a high level of precision and attention to detail and more steps for early error recognition. The third platform was not ready for testing in two sites. Participants did not want to express an explicit preference for one tNGS platform over the other, noting that both Deeplex and ONT had their pros and cons and that both needed further development to be fit for purpose.	The group notes that there may be many challenges in scaling up tNGS as the initial test for DST for all TB patients, as illustrated by the data shared. The group also notes, however, that despite the implementation challenges, most all countries have succeeded in implementing sequencing-based tests for COVID. While many in the group support a judgement of "probably yes" for this space, the group voted to move forward with "varies" given the great amount of debate and variation to be considered in this question. However, the group also notes that a potential WHO recommendation made in favor of the technology would help make it more feasible by supporting funding, implementation aid, etc.

Summary of judgements

	Judgement						
Problem	No	Probably no	Probably yes	**Yes**		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable Effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable Effects	Large	**Moderate**	Small	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	**Low**	Moderate	High			No included studies
Certainty of the evidence of test's effects	Very low	Low	Moderate	High			**No included studies**
Certainty of the evidence of management's effects	Very low	Low	Moderate	**High**			No included studies
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**

	Judgement						
Certainty of effects	Very low	Low	Moderate	High			**No included studies**
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	**Very low**	Low	Moderate	High			No included studies
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	**Varies**	No included studies
Equity	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	**Varies**	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	Probably yes	Yes		**Varies**	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In people with bacteriologically confirmed pulmonary TB, targeted next-generation sequencing technologies may be used on respiratory samples to diagnose resistance to rifampicin, isoniazid, fluoroquinolones, pyrazinamide and ethambutol rather than culture-based phenotypic drug susceptibility testing. *(conditional recommendation, certainty of evidence moderate [isoniazid and pyrazinamide],low [rifampicin, fluoroquinolones and ethambutol]).*

Subgroup considerations

In individuals with confirmed bacteriological pulmonary tuberculosis (TB) disease, priority should be assigned to those at higher risk of resistance to first-line treatment medications, including individuals who:

- continue to be smear or culture positive after two months or more of treatment or experience treatment failure, particularly those with initial results indicating rifampicin-susceptibility.
- have a history of previous TB treatment
- are in contact with a person with known drug resistance
- reside in settings or belonging to sub-groups where the probability for resistance to either rifampicin, isoniazid or fluoroquinolone (used in new shorter regimens) is high, or where there is a high prevalence of *Mycobacterium tuberculosis* strains harbouring mutations not detected by other rapid molecular tests

Priority should be given to samples with a high bacillary load as determined by initial bacteriological tests (e.g., semi-quantitative high/medium or smear-positive grading). In situations where the bacillary load is low (e.g., semi-quantitative grading of low/very low/trace or smear-negative), the recommendations still hold while acknowledging the higher rates of indeterminate results and the potential need for repeat testing.

Similarly, the recommendations apply to children, adolescents, and people living with HIV (PLHIV), acknowledging the higher risk of indeterminate results in these populations due to higher frequency of samples with low bacterial load.

The recommendation is based on data obtained from sputum and bronchoalveolar lavage specimens and can be extrapolated to other lower respiratory tract samples (e.g., endotracheal aspirates). However, further research is needed to evaluate the use of these tests on alternative sample types for diagnosing pulmonary TB in children (such as nasopharyngeal and stool samples) and diagnosing extra-pulmonary TB.

Implementation considerations

- Targeted next-generation sequencing is a high-complexity test in its current format and is most suitable for centralized laboratories equipped with specialized skills and infrastructure.
- These tests do not replace rapid tests that are more accessible and easier to perform for detecting resistance to rifampicin, isoniazid, and fluoroquinolones. However, they can be considered as an alternative initial option for prioritized populations. Individuals requiring rapid and comprehensive DST but with limited access to phenotypic DST will benefit most from these tests.

Monitoring and evaluation

- Standardize the nomenclature for result reporting across different targeted NGS technologies and integrate data systems to enable cross-programme utilization of targeted NGS data.
- Ensure separate recording of true failures and unclassified mutations, and monitor trends over time as an essential component of result reporting.
- Regularly monitor performance data, including overall resistance rates, resistance rates by specific drugs or targets and turnaround times (both total and in-laboratory).
- Incorporate quality monitoring measures, such as tracking indeterminate rates, sequencing coverage and depth, and participation in external quality assurance programmes.
- Establish an external quality assurance programme for sequencing that covers all relevant targets of interest.

- Integrate the sequencing data generated into existing surveillance systems to monitor the prevalence and trends in drug resistance effectively. Share the data to update the WHO mutation catalogue.
- Collect cost data to address important questions, such as the costs associated with introducing and scaling up targeted NGS in different settings, the trade-offs between turnaround time and batching, and the optimal balance in various settings.
- Assess the impact of multi-disease testing on program operations and costs, including disease-specific testing volumes, turnaround times, costing, resource sharing, and resource requirements.
- Evaluate the impact of time to treatment initiation/modification, treatment outcomes, and overall cost-effectiveness of targeted NGS implementation.

Research priorities

Clinical research needs:

- Conduct clinical trials to assess the impact of targeted NGS on patient-important outcomes.
- Evaluate the accuracy and effectiveness of targeted NGS among populations composed of individuals diagnosed with tuberculosis (TB), without enrichment for rifampicin or other drug resistance.
- Assess the accuracy and effectiveness of targeted NGS for analyzing extra-pulmonary samples, including cerebrospinal fluid for meningitis, non-sputum samples (such as nasopharyngeal aspirate, gastric aspirate, stool) for children, and alternative sample types (e.g., tongue swabs) in both adults and children.
- Undertake additional qualitative and quantitative research to further understand the perspectives of patients and clinicians regarding the acceptability and feasibility of using targeted NGS.
- Investigate the association between specific mutations, minimum inhibitory concentration (MIC), lineage, and treatment outcomes (relapse-free cure), disaggregated by population.
- Explore the factors driving the emergence of drug resistance and examine the molecular evolution of drug resistance.
- Evaluate the role of hetero-resistance in comparison to whole-genome sequencing and phenotypic drug susceptibility testing (DST) in achieving a relapse-free cure.
- Adopt a "one health" approach to investigate the intersection between human and zoonotic TB and the correlation between antimicrobial usage in humans and the agricultural sector, particularly in relation to the development of multi-drug resistance.

Implementation research needs:

- Develop and evaluate effective and efficient implementation models by integrating targeted NGS into laboratory networks and optimizing algorithms, aiming to enhance timely access to testing, treatment initiation and improve patient outcomes.
- Develop strategies to enhance the efficiency of targeted NGS testing, including sample concentration techniques, determining optimal thresholds of bacterial load from initial tests before performing targeted NGS, utilizing pooled samples from multiple individuals, and employing molecular transport medium for ambient storage and transfer of samples to testing sites.
- Regularly update the interpretive catalogue based on WHO updates, incorporating additional genetic targets (for future tests) to enhance the sensitivity and specificity of targeted NGS and include new drugs used for TB treatment (e.g. pretomanid)
- Explore technological advancements to simplify the testing process, automate steps (especially library preparation), develop decentralized targeted NGS solutions, and investigate potential synergies with existing initial tests (e.g., utilizing leftover DNA or smear-positive slides).
- Conduct comprehensive mapping of sequencing capacity within countries and perform diagnostic network optimization exercises. Placement of the technology should consider the demand across multiple diseases, facilitating multiplex use of the machines and shared costs.
- Compile and utilize lessons learned from applying targeted NGS technology in other diseases to guide implementation strategies for TB effectively.

Question

Should targeted Next-Generation Sequencing be used to diagnose drug resistance to isoniazid (INH), flouroquinolones (FQs), pyrazinamide (PZA), bedaquiline (BDQ), linezolid (LZD), clofazimine (CFZ), amikacin (AMK), ethambutol (EMB), streptomycin (STR) in patients with bacteriologically confirmed rifampin-resistant pulmonary tuberculosis (TB) disease?	
Population:	Patients with bacteriologically confirmed rifampin-resistant pulmonary TB disease (RR-TB)
Intervention:	Targeted Next-Generation sequencing (NGS) for detecting resistance to isoniazid (INH), fluoroquinolones (FQs), pyrazinamide (PZA), bedaquiline (BDQ), linezolid (LZD), clofazimine (CFZ), amikacin (AMK), ethambutol (EMB), streptomycin (STR) compared to phenotypic drug sensitivity testing
Purpose of the test:	Used as a subsequent test for further TB drug resistance in people with known RR-TB
Role of the test:	An add-on test following initial diagnosis with rifampicin-resistant TB disease with a molecular WHO-approved rapid diagnostic test
Linked treatments:	Correct treatment for RR/MDR-TB based on drug resistance patterns.
Anticipated outcomes:	Improved treatment outcomes based on drug resistance detected.
Setting:	TB programmes worldwide
Perspective:	Public health perspective
Subgroups:	N/A
Conflict of interests:	All guideline panel members completed declaration of interest forms

Judgement	Research evidence	Additional considerations
Problem **Is the problem a priority?**		
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Drug-resistant tuberculosis (DR-TB) is a major threat to global TB control. There are estimated to be over 10 million cases and 1.5 million deaths of TB annually, including 450,000 cases of DR-TB, only a third of whom are diagnosed and treated appropriately (WHO Global TB Report 2022). WHO recommendations include initial tests for resistance to RIF and INH among all TB patients and tests for resistance to fluoroquinolones (FQs) among those with RIF-resistant or RIF-susceptible INH-resistant TB. Phenotypic drug susceptibility testing (pDST), remains the reference standard for drug resistance detection to most drugs. It is a culture-based approach that requires several weeks for results to be available and performed at specialized sites with limited access. In recent years, nucleic-acid amplification tests (NAATs) have offered options for molecular detection of drug resistance, including line-probe assays and rapid molecular tests that can detect drug resistance in a fraction of the time required for culture-based methods. However, they have limitations in the number of drugs they can test for resistance, the ability to distinguish mutations with differing resistance potential and how quickly they can incorporate new data on genetic information about drug resistance. The recent introduction of new drugs and repurposing of existing antimicrobial	The group notes that the emergence of resistance to BDQ is increasing, and of increasing concern, in countries where it is being used in new regimens for MDR-TB - in Pakistan, South Africa it is already over 5% among patients tested. It is also already found

	agents for the treatment of TB have generated new TB treatment regimens at a comparatively rapid rate, providing improvements to treatment options, outcomes, and quality of life during treatment among DR-TB patients. However, as resistance to these new and repurposed drugs in the community gradually increases, there is concern about the lack of options for rapid detection of resistance to these drugs by currently approved methods. Gene sequencing technologies provide an option for rapid, accurate genetic analysis and detection of mutations indicating resistance in a fraction of the time required for culture-based methods for detecting resistance. Recently several commercial "End-to-End Solutions" for targeted next-generation sequencing (tNGS) for detection of drug resistance have become available that promise a higher throughput, a significantly faster time to result, and a greater accuracy across more TB drugs than current WHO-recommended methods for DST, and offer the potential to rapidly assimilate new information on genetic markers for resistance as they become known. They also offer the potential for drug susceptibility testing for new and repurposed drugs for which there are no otehr currently available options. The question to the GDG is to evaluate the available evidence on tNGS technologies and to generate guidance on their use in programmatic management of DR-TB globally.	in Brazil, not a high MDR-TB burden country. The more you test for it, the more you find it. However, the protocol for phenotypic DST for BDQ is very difficult to conduct in labs. This is especially concerning as there are not many good options for regimens for patients who are resistant to BDQ. Also, this drug serves as the backbone for future MDR-TB regimens in development, so it is essential to protect it.
Test accuracy How accurate is the test?		
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't	**Test accuracy** INH (pDST): Sensitivity: 0.96 (95% CI: 0.94 to 0.99) Specificity: 0.96 (95% CI: 0.92 to 1.00) LFX (pDST) Sensitivity: 0.96 (95% CI: 0.90 to 1.00)	Though the group notes the less than desired sensitivity for some drugs (BDQ, LZD, CFZ, AMK) and less than desired specificity for

know	Specificity: 0.96 (95% CI: 0.93 to 0.99) MFX (pDST) Sensitivity: 0.97 (95% CI: 0.94 to 1.00) Specificity: 0.95 (95% CI: 0.91 to 0.99) PZA (comp) Sensitivity: 0.90 (95% CI: 0.87 to 0.93) Specificity: 0.99 (95% CI: 0.97 to 1.00) BDQ (pDST) Sensitivity: 0.68 (95% CI: 0.43 to 0.93) Specificity: 0.97 (95% CI: 0.94 to 1.00) LZD (pDST) Sensitivity: 0.69 (95% CI: 0.39 to 0.99) Specificity: 1.00 (95% CI: 1.00 to 1.00) CFZ (pDST) Sensitivity: 0.70 (95% CI: 0.35 to 1.00) Specificity: 0.96 (95% CI: 0.93 to 0.99) AMK (pDST) Sensitivity: 0.87 (95% CI: 0.75 to 1.00) Specificity: 0.99 (95% CI: 0.98 to 1.00) EMB (comp) Sensitivity: 0.97 (95% CI: 0.95 to 0.98) Specificity: 0.98 (95% CI: 0.96 to 1.00) STR (pDST) Sensitivity: 0.98 (95% CI: 0.96 to 1.00) Specificity: 0.75 (95% CI: 0.59 to 0.91)	STR, the group still deems to classify the test as a whole "accurate" given that the test can still be useful for clinical decision-making, depending on the clinical and epidemiological context in which it is used. The group notes that the data are based on sites using specific critical concentrations that might not be the most appropriate. We have very little data to say that the critical concentrations used reflect treatment outcomes. The group also notes the accuracies seen are similar to smear microscopy, which is not considered highly accurate but still useful for clinical decision-making. Lastly, the group notes that the accuracy of

<table>
<tr><td></td><td></td><td>genetic testing is likely to change quickly in the future as the catalog of mutations expands.</td></tr>
<tr><td colspan="3">Desirable Effects
How substantial are the desirable anticipated effects?</td></tr>
<tr><td>○ Trivial
○ Small
○ Moderate
● Large
○ Varies
○ Don't know</td><td>
<table>
<tr><th rowspan="2">Test result</th><th colspan="3">Number of results per 1000 patients tested (95% CI)</th><th rowspan="2">№ of participants (studies)</th><th rowspan="2">Certainty of the evidence (GRADE)</th></tr>
<tr><th>Prevalence 60%</th><th>Prevalence 75%</th><th>Prevalence 90%</th></tr>
<tr><td>True positives patients with drug resistance to isoniazid (INH) (pDST)</td><td>576 (564 to 594)</td><td>720 (705 to 742)</td><td>864 (846 to 891)</td><td>1440 (12)</td><td>⨁⨁⨁⨁ High[a]</td></tr>
<tr><td>False negatives patients incorrectly classified as not having drug resistance to isoniazid (INH) (pDST)</td><td>24 (6 to 36)</td><td>30 (8 to 45)</td><td>36 (9 to 54)</td><td></td><td></td></tr>
<tr><td>True negatives patients without drug resistance to isoniazid (INH) (pDST)</td><td>384 (368 to 400)</td><td>240 (230 to 250)</td><td>96 (92 to 100)</td><td>517 (12)</td><td>⨁⨁⨁⨁ High[a]</td></tr>
<tr><td>False positives patients incorrectly classified as having drug resistance to isoniazid (INH) (pDST)</td><td>16 (0 to 32)</td><td>10 (0 to 20)</td><td>4 (0 to 8)</td><td></td><td></td></tr>
</table>
</td><td>The group agrees that the desirable effects are "large" given the treatment decision-making possible from the results, and the rapid turnaround time that tNGS offers compared to culture-based testing.</td></tr>
</table>

a. Prevalence of resistance to isoniazid across data used in the model was 74% (CI 72% to 76%)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%	Prevalence 30%	Prevalence 50%		
True positives patients with drug resistance to levofloxacin (LFX) (pDST)	96 (90 to 100)	288 (270 to 300)	480 (450 to 500)	654 (6)	⨁⨁⨁◯ Moderate[a, b]
False negatives patients incorrectly classified as not having drug resistance to levofloxacin (LFX) (pDST)	4 (0 to 10)	12 (0 to 30)	20 (0 to 50)		
True negatives patients without drug resistance to levofloxacin (LFX) (pDST)	864 (837 to 891)	672 (651 to 693)	480 (465 to 495)	913 (7)	⨁⨁⨁⨁ High[a]
False positives patients incorrectly classified as having drug resistance to levofloxacin (LFX)	36 (9 to 63)	28 (7 to 49)	20 (5 to 35)		

(pDST)					

a. Prevalence of resistance to levofloxacin across data used in the model was 42% (CI 39% to 44%)
b. One outlying study for sensitivity

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%	Prevalence 30%	Prevalence 50%		
True positives patients with drug resistance to moxifloxacin (MFX) (pDST)	97 (94 to 100)	291 (282 to 300)	485 (470 to 500)	652 (6)	⨁⨁⨁⨁ High[a]
False negatives patients incorrectly classified as not having drug resistance to moxifloxacin (MFX) (pDST)	3 (0 to 6)	9 (0 to 18)	15 (0 to 30)		
True negatives patients without drug resistance to moxifloxacin (MFX) (pDST)	855 (819 to 891)	665 (637 to 693)	475 (455 to 495)	921 (8)	⨁⨁⨁⨁ High[a]
False positives patients incorrectly classified as having drug resistance to moxifloxacin (MFX) (pDST)	45 (9 to 81)	35 (7 to 63)	25 (5 to 45)		

a. Prevalence of resistance to moxifloxcin across data used in the model was 41% (CI 39% to 44%)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 30%	Prevalence 50%	Prevalence 90%		
True positives patients with drug resistance to pyrazinamide (PZA) (composite)	270 (261 to 279)	450 (435 to 465)	810 (783 to 837)	346 (3)	⨁⨁⨁⨁ High[a]
False negatives patients incorrectly classified as not having drug resistance to pyrazinamide (PZA) (composite)	30 (21 to 39)	50 (35 to 65)	90 (63 to 117)		
True negatives patients without drug resistance to pyrazinamide (PZA) (composite)	693 (679 to 700)	495 (485 to 500)	99 (97 to 100)	269 (3)	⨁⨁⨁⨁ High[a]
False positives patients incorrectly classified as having drug resistance to pyrazinamide (PZA) (composite)	7 (0 to 21)	5 (0 to 15)	1 (0 to 3)		

a. Prevalence of resistance to pyrazinamide (composite) across data used in the model was 56% (CI 52% to 60%)

Test result	Number of results per 1000 patients tested (95% CI)	№ of participants	Certainty of the

	Prevalence 1%	Prevalence 3%	Prevalence 5%	(studies)	evidence (GRADE)
True positives patients with drug resistance to bedaquiline (BDQ) (pDST)	7 (4 to 9)	20 (13 to 28)	34 (22 to 47)	31 (3)[a]	⊕⊕◯◯ Low[b,c,d]
False negatives patients incorrectly classified as not having drug resistance to bedaquiline (BDQ) (pDST)	3 (1 to 6)	10 (2 to 17)	16 (3 to 28)		
True negatives patients without drug resistance to bedaquiline (BDQ) (pDST)	960 (931 to 990)	941 (912 to 970)	922 (893 to 950)	519 (4)[e]	⊕⊕⊕⊕ High[d]
False positives patients incorrectly classified as having drug resistance to bedaquiline (BDQ) (pDST)	30 (0 to 59)	29 (0 to 58)	28 (0 to 57)		

a. This model is not controlled for cycle threshold (CT) value as that variable was collinear in the original model, but we did not downgrade for risk of bias.
b. One study had very low sensitivity but it only had 3 resistant samples. It identified 0/3. We did not downgrade for inconsistency.
c. Very wide 95% confidence intervals for sensitivity. We downgraded by two levels for imprecision.
d. Prevalence of resistance to bedaquiline across data used in the model was 6% (CI 4% to 8%)

e. This model is not controlled for rifampicin resistance as this variable was collinear in the original model. Instead, the data have been restricted to isolated that are resistant to rifampicin by Xpert, and then controlled for CT value.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 3%	Prevalence 5%		
True positives patients with drug resistance to linezolid (LZD) (pDST)	7 (4 to 10)	21 (12 to 30)	34 (20 to 50)	31 (4)[a]	⨁⨁◯◯ Low[b,c,d]
False negatives patients incorrectly classified as not having drug resistance to linezolid (LZD) (pDST)	3 (0 to 6)	9 (0 to 18)	16 (0 to 30)		
True negatives patients without drug resistance to linezolid (LZD) (pDST)	990 (990 to 990)	970 (970 to 970)	950 (950 to 950)	1093 (6)[e]	⨁⨁⨁⨁ High[d]
False positives patients incorrectly classified as having drug resistance to linezolid (LZD) (pDST)	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)		

a. This model is restricted to isolates that were resistant to rifampicin by Xpert, and controls for CT value. We did not downgrade for serious risk of bias.
b. One study was an outlier for sensitivity but only had 1 resistant sample (0/1 detected). We did not downgrade

for serious inconsistency.

c. Very wide 95% confidence intervals; we downgraded by two levels for imprecision.
d. Prevalence of resistance to linezolid across data used in the model was 3% (CI 2% to 4%)
e. This model is restricted to isolates that were resistant to rifampicin by Xpert, and does not control for CT value as both variables were collinear in the original model

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	Prevalence 3%	Prevalence 5%		
True positives patients with drug resistance to clofazimine (CFZ) (pDST)	7 (3 to 10)	21 (10 to 30)	35 (17 to 50)	36 (4)[a]	⨁⨁◯◯ Low[b,c,d]
False negatives patients incorrectly classified as not having drug resistance to clofazimine (CFZ) (pDST)	3 (0 to 7)	9 (0 to 20)	15 (0 to 33)		
True negatives patients without drug resistance to clofazimine (CFZ) (pDST)	950 (921 to 980)	931 (902 to 960)	912 (884 to 941)	789 (6)	⨁⨁⨁⨁ High[d]
False positives patients incorrectly classified as having drug resistance to clofazimine (CFZ) (pDST)	40 (10 to 69)	39 (10 to 68)	38 (9 to 66)		

a. Model not controlled for CT value as this was collinear in the original model
b. The two smaller studies are outliers for sensitivity. Downgraded by one level for inconsistency as it's more than one small study.
c. Very wide 95% confidence intervals for sensitivity. We rated down one level for serious imprecision.
d. Prevalence of resistance to clofazimine across data used in the model was 3% (CI 2% to 4%)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%	Prevalence 10%	Prevalence 15%		
True positives patients with drug resistance to amikacin (AMK) (pDST)	44 (38 to 50)	87 (75 to 100)	131 (112 to 150)	115 (5)[a]	⊕◯◯◯ Very low[b,c,d,e]
False negatives patients incorrectly classified as not having drug resistance to amikacin (AMK) (pDST)	6 (0 to 12)	13 (0 to 25)	19 (0 to 38)		
True negatives patients without drug resistance to amikacin (AMK) (pDST)	941 (931 to 950)	891 (882 to 900)	842 (833 to 850)	1003 (8)[a]	⊕⊕⊕◯ Moderate[d,e]
False	9 (0 to 19)	9 (0 to 18)	8 (0 to 17)		

positives patients incorrectly classified as having drug resistance to amikacin (AMK) (pDST)					

a. The model is restricted to isolated that were resistant to rifampicin by Xpert, as this was collinear in the original model, but controls for CT value
b. Two outlying studies for sensitivity, albeit small studies, rated down by one level for serious risk of bias.
c. wide 95% confidence intervals for sensitivity
d. Prevalence of resistance to amikacin across data used in the model was 10% (CI 9% to 12%), rated down by one level for serious inconsistency
e. Non WHO recommended CC used

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%	Prevalence 30%	Prevalence 50%		
True positives patients with drug resistance to ethambutol (EMB) (composite)	97 (95 to 98)	291 (285 to 294)	485 (475 to 490)	431 (4)	⨁⨁⨁◯ Moderate[a, b]
False negatives patients incorrectly classified as not having drug resistance to ethambutol (EMB) (composite	3 (2 to 5)	9 (6 to 15)	15 (10 to 25)		

)					
True negatives patients without drug resistance to ethambutol (EMB) (composite)	882 (864 to 900)	686 (672 to 700)	490 (480 to 500)	123 (4)[c]	⨁⨁⨁◯ Moderate[a, b]
False positives patients incorrectly classified as having drug resistance to ethambutol (EMB) (composite)	18 (0 to 36)	14 (0 to 28)	10 (0 to 20)		

a. Prevalence of resistance to ethambutol (composite) across data used in the model was 78% (CI 74% to 81%)
b. Different samples tested for index and reference tests, therefore downgraded by one level for serious risk of bias.
c. The model is restricted to isolated that were resistant to rifampicin by Xpert, as this was collinear in the original model, but controls for CT value

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%	Prevalence 30%	Prevalence 50%		
True positives patients with drug resistance to streptomycin (STR) (pDST)	98 (96 to 100)	294 (288 to 300)	490 (480 to 500)	493 (5)	⨁⨁⨁⨁ High[a]
False negatives patients	2 (0 to 4)	6 (0 to 12)	10 (0 to 20)		

incorrectly classified as not having drug resistance to streptomycin (STR) (pDST)					
True negatives patients without drug resistance to streptomycin (STR) (pDST)	675 (531 to 819)	525 (413 to 637)	375 (295 to 455)	250 (5)	⨁⨁◯◯ Low[a,b,c]
False positives patients incorrectly classified as having drug resistance to streptomycin (STR) (pDST)	225 (81 to 369)	175 (63 to 287)	125 (45 to 205)		

a. Prevalence of resistance to streptomycin across data used in the model was 66% (CI 63% to 70%), but not rated down for risk of bias.
b. One study was an outlier; rated down by one level for inconsistency.
c. Wide 95% confidence intervals for specificity; rated down by one level for serious imprecision.

NOTE:

- A True Positive test indicates that the patient is correctly treated with appropriately modified regimen for resistance pattern; risk of treatment failure or developing further resistance are minimized.
- A True Negative test indicates that the patient is correctly treated with appropriate regimen; treatment burden minimized.

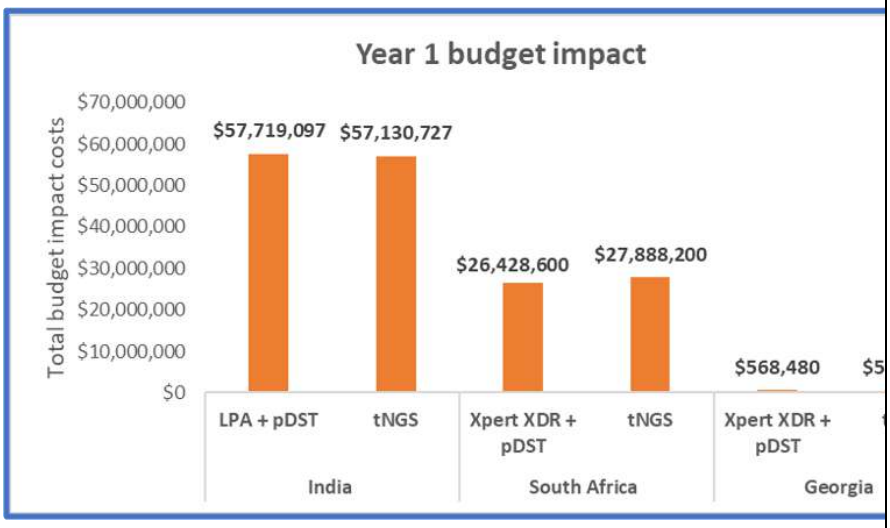
Undesirable Effects
How substantial are the undesirable anticipated effects?

○ Large ● Moderate ○ Small ○ Trivial ○ Varies ○ Don't know	See tables above for numbers of false positive and false negative test results expected for each drug (see "Desirable Effects"). NOTE: • A False Positive test may result in incorrect and inappropriate treatment regimens, and put people with TB being at risk of unnevessary adverse effects. • A False Negative test may result in people with TB receiving incorrect treatment and causing delays in receiving appropriate treatment, putting them at risk of treatment failure, mortality, developing further resistance, and transmission of DR-TB to others. Indeterminate rates: INH (pDST) = 14.6% (13.0-16.2%) LFX (pDST) = 9.2% (7.8-10.7%) MFX (pDST) = 9.3% (7.9 - 10.9%) PZA (comp) = 17.6% (14.6 - 20.8%) BDQ (pDST) = 16.7% (13.7-20.1%) LZD (pDST) = 15.1% (13.1-17.3%) CFZ (pDST) = 11.6% (9.5-14.1%) AMK (pDST) = 17.8% (15.6-20.2%) EMB (comp) = 20.6% (17.3-24.2%) STR (pDST) = 18.8% (16.1-21.8)	The group notes that the high indeterminate rates affect this decision significantly, as the lack of clinically actionable results from the test from 9 to 21% of the time reduces its clinical utility and increases the effective per-patient test cost.
Certainty of the evidence of test accuracy What is the overall certainty of the evidence of test accuracy?		
○ Very low ● Low ○ Moderate ○ High ○ No included studies	Certainty of test accuracy: - HIGH for INH, MFX, PZA - MODERATE for LFX, EMB - LOW for BDQ, LZD, CFZ, STR - VERY LOW for AMK	Given the range of data included in this combined PICO, the composite measure of certainty of evidence for test accuracy is LOW.
Certainty of the evidence of test's effects What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?		
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No included studies. tNGS is an "in vitro" test, and therefore no adverse effects are expected for patients beyond discomfort from producing sputum. tNGS is likely to have a faster turn-around time than culture-based tests as it can be completed in a few days' time versus several weeks to months required for culture growth. However, placement of the test in the health system and batching/multiplexing of the technology will impact the turn-around time experienced in a particular setting.	
Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?		
● Very	Treatment regimen depends on the results of drug susceptibility testing.	This decision is

low ○ Low ○ Moderate ○ High ○ No included studies	· WHO suggests the use of a 6-month treatment regimen, composed of Bedaquiline, Pretomanid, Linezolid (600 mg) and Moxifloxacin (BPaLM), rather than the 9-month or longer (18-month) regimens in MDR/RR-TB patients (**Conditional recommendation, very low certainty evidence**) · WHO suggests the use of the 9-month all-oral regimen rather than longer (18-month) regimens in patients with MDR/RR-TB and in whom resistance to fluoroquinolones has been excluded. (**Conditional recommendation, very low certainty evidence**) · In multidrug- or rifampicin-resistant tuberculosis (MDR/RR-TB) patients on longer regimens, all three Group A agents and at least one Group B agent should be included to ensure that treatment starts with at least four TB agents likely to be effective, and that at least three agents are included for the rest of the treatment if bedaquiline is stopped. If only one or two Group A agents are used, both Group B agents are to be included. If the regimen cannot be composed with agents from Groups A and B alone, Group C agents are added to complete it. (**Conditional recommendation, very low certainty evidence**)	based on the strength of evidence underpinning the latest WHO-recommended treatment regimens for MDR/RR-TB patients.
Certainty of the evidence of test result/management How certain is the link between test results and management decisions?		
○ Very low ○ Low ○ Moderate ○ High ● No included studies	No included studies for this question about whether those tested access management (linkage to care). If clinicians receive the results of the test, there is a high likelihood that the test results would be used and treatment decisions will be based on the test results for resistance detection. TB programmes have processes in place to link laboratory test results with clinicians treating patients. Linkage of laboratory results to patients in a timely manner impacts on loss to follow-up of patients and retention in care. tNGS is likely to have a faster turn-around time than culture-based tests as it can be completed in a few days' time versus several weeks required for culture growth. However, placement of the test in the health system and batching/multiplexing of the technology will impact the turn-around time experienced in any particular setting. In most contexts, TB medicines are available. Availability of TB medicines will impact the ability to treat patients according to the test results. In a study considering second-line DST, the authors conclude that, in most settings, second-line DST could substantially improve treatment outcomes for patients with rifampin-resistant TB, reduce transmission of drug-resistant TB, prevent amplification of drug resistance, and be affordable or even cost-saving. Given the large investment made in each patient treated for rifampin-resistant TB, these payoffs would come at a relatively small incremental cost. These anticipated benefits likely justify addressing the real challenges faced in implementing second-line DST in most high-burden settings (Kendall EA, Cohen T, Mitnick CD, Dowdy DW. Second line drug susceptibility testing to inform the treatment of rifampin-resistant tuberculosis: a quantitative perspective. International Journal of Infectious Diseases. 2017;56:185-9: doi.org/10.1016%2Fj.ijid.2016.12.010).	As with the previous PICO, the group notes that there are many unknowns here - it is not yet known what the uptake would be for this technology if approved. In addition, while targeted NGS is expected to have a faster turnaround time to results compared to culture-based DST methods, there are many health system factors that will affect how it will impact on patient care. Experiences with implementation of other rapid DST options (mWRDs, LPA) have illustrated how health

		system issues affect turnaround time for supposedly rapid tests.
Certainty of effects What is the overall certainty of the evidence of effects of the test?		
○ Very low ○ Low ○ Moderate ○ High ● No included studies	There are no available comparative intervention studies on the effectiveness of TNGS compared to current standard of care on patient-important outcomes. <u>Indirect evidence from other studies with faster turn around of results:</u> A comparative cohort study from China found that patients with early available molecular DST results had a more rapid culture conversion (aHR1.94 95% CI: 1.37-2.73; median,12 vs 24 months, respectively; *P* < 0.001) and a higher rate of treatment success (68% vs 47%, *P* < 0.01) (1). Similarly, a pragmatic trial from Brazil showed that compared to the MGIT group, culture conversion after 6 months was higher for Xpert in arm 1 (90.9% vs 79.3%, p=0.39) and LPA in arm 2 (80.0% vs 83.0%, p=0.81) (2). In contrast, a study from Ethiopia did not show any difference in treatment outcomes between Xpert, LPA and MGIT used for detection drug-resistant TB (3). References 1. Shi W, Davies Forsman L, Hu Y, Zheng X, Gao Y, Li X, et al. Improved treatment outcome of multidrug-resistant tuberculosis with the use of a rapid molecular test to detect drug resistance in China. International Journal of Infectious Diseases. 2020;96:390-7.doi.org/10.1016/j.ijid.2020.04.049 2. Kritski A, Oliveira MM, Almeida IN de, Ramalho D, Andrade MK de N, Carvalho M, et al.. Clinical Impact of the Line Probe Assay and Xpert® MTB/RIF Assay in the Presumptive Diagnosis of Drug-Resistant Tuberculosis in Brazil: A Pragmatic Clinical Trial. Rev Soc Bras Med Trop [Internet]. 2022;55:e0191–2021. Available from: https://doi.org/10.1590/0037-8682-0191-2021 doi.org/10.1590%2F0037-8682-0191-2021 3. Kassa GM, Merid MW, Muluneh AG, Wolde HF. Comparing the impact of genotypic based diagnostic algorithm on time to treatment initiation and treatment outcomes among drug-resistant tuberculosis patients in Amhara region, Ethiopia. PLOS ONE. 2021;16(2):e0246938. doi.org/10.1371/journal.pone.0246938	It is noted that there are no specific comparative effectiveness data for this question.
Values Is there important uncertainty about or variability in how much people value the main outcomes?		
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probabl	WHO commissioned a qualitative evidence synthesis (QES) which did not find any included studies. A primary qualitative study of user experiences (technicians and implementers in the FIND studies) was conducted but did not explicitly explore how much people value the outcomes or their preferences with respect to the intervemtion compared to the comparison. <u>Indirect evidence:</u> A qualitative evidence synthesis of recipient and provider perspectives on rapid molecular tests for TB and drug resistance found that people with tuberculosis valued reaching diagnostic closure with an accurate diagnosis, avoiding diagnostic delays, and keeping diagnostic-associated costs low. Similarly, healthcare providers valued aspects of accuracy and the resulting	The group notes that there is probably no important uncertainty in how much patients and providers value the main outcomes

y no important uncertainty or variability ○ No important uncertainty or variability	confidence in low-complexity NAAT results, rapid turnaround times, and low costs to people seeking a diagnosis. (Engel N, Ochodo EA, Karanja PW, Schmidt B-M, Janssen R, Steingart KR, Oliver S. Rapid molecular tests for tuberculosis and tuberculosis drug resistance: a qualitative evidence synthesis of recipient and provider views. Cochrane Database of Systematic Reviews 2022, Issue 4. Art. No.: CD014877. DOI: 10.1002/14651858.CD014877.pub2. Accessed 02 October 2023).	based on precedent on this topic.
Balance of effects Does the balance between desirable and undesirable effects favor the intervention or the comparison?		
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know	The balance of desirable and undesirable effects **probably favors** the intervention versus the comparison.	Here the group notes especially the potential for faster turnaround time for TNGS technology, and the rapid potential for TNGS technology to assimilate new information as evidence becomes available.
Resources required How large are the resource requirements (costs)?		
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings	Literature reviews suggest unit test costs for tNGS are consistently higher than phenotypic drug susceptibility testing (pDST). The largest cost components were test kit reagents, ancillary consumables, and costs associated with sequencing. Key cost drivers include specific sequencer used, depth and breadth of coverage, inefficiencies in initial sample runs, the economics of scale via batching or cross-batching, operational efficiency, availability of trained personnel, sequencers being used to full capacity, bulk purchases, and complexity of the infectious pathogen. Empirical costing estimates for tNGS unit cost per sample for Deeplex Myc-TB Genoscreen tNGS ranged from: (unit costs includes consumables, equipment,	Here the group notes that the resources required will vary significantly based on different in-country models of laboratory

○ Large savings ● Varies ○ Don't know	staffing and overgead where available, costs assume tNGS testing for all drugs) **- $134 to $257 in South Africa,** **- $120 to $198 in Georgia and** **- $121 to $175 in India,** depending on patient volume, batching and negotiated tNGS kit cost.  From published BIA (Cates 2022) : For all NGS scenarios, the majority (55–80%) of costs were devoted to reagent kits and start-up costs of NGS were small relative to routine costs borne each year.	infrastructure and different ranges of prevalences of drug resistance across the various drugs: - For smaller populations, it is more feasible to implement TNGS for RR-TB patients because the numbers are small, however then the cost becomes high per patient. - Accumulated costs for phenotypic testing over a long period of time will add up; cost savings can be envisaged from the test itself or from the treatment side, depending on the results. - For the population specified in this PICO, as it is already Rif-resistant, there will be significant costs for testing no matter what. - The budget impact assessment shows very little difference between the overall costs of phenotypic testing and TNGS testing. - However the group notes

		that there are uncertainties in the real world implementation of the models; hence the group settles on "varies".
Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
● Very low ○ Low ○ Moderate ○ High ○ No included studies	The systematic review contained 10 manuscripts with very limited economic data and no cost-effectiveness analyses. Data on total implementation costs are also limited with only one budget impact assessment retrieved in the systematic review. Several key scenarios were assessed in the empirical costing to derive unit cost ranges.	
Cost effectiveness Does the cost-effectiveness of the intervention favor the intervention or the comparison?		
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the intervention ● Varies	We found tNGS used as the test of drugs resistance in patients with bacteriologically confirmed rifampin-resistant pulmonary TB disease has higher cost and fewer health gains compared with universal pDST, as pDST has high diagnostic accuracy and we have assumed there is no difference in the loss to follow-up between tNGS and pDST arms. When tNGS was compared with current in-country DST practice as the test of drug resistance in patients with bacteriologically confirmed rifampin-resistant pulmonary TB we found tNGS to be cost-effective in South Africa (ICER=$15,619; 95% UR: Cost Saving-$114,782) at a willingness-to-pay (WTP) threshold of $21,165. In India, tNGS dominated the in-country DST practice with less cost and more health gains (95% UR: Cost Saving-$60,083). In Georgia, tNGS was not cost-effective (ICER=$18,375, 95% UR: Cost saving-$158,972) at the WTP threshold of $15,065. Key drivers of cost-effectiveness included rate of contamination of pDST, probability of repeat testing among pDST(proxy for LTFU) and per unit cost of tNGS. Reduced LTFU for tNGS leads to improved cost-effectiveness. TAT impacted by test placement which is impacted by test volume and cost. Multiplexing may reduce unit test costs and improve turn around time, if volume of eligible for testing is low. Batching may reduce unit test cost but also increase turn around time (more delay), leading to increasing LTFU. Lack of batching may lead to increased unit test costs, decreased likelihood of cost-effectiveness. This is the first study done to assess the cost-effectiveness of tNGS used for	As with the discussion of resources, the group notes that there are many uncertainties inherent in this data. The cost effectiveness analysis can change quickly as the technology gets cheaper and better. However the group notes the need to be realistic about the true cost of implementing the technologies in the current laboratory

○ No included studies	diagnosis of drug resistance in patients with bacteriologically confirmed rifampin-resistance pulmonary TB diseases. This cost-effectiveness result only looks at tNGS doing DST of FQ, BDQ and LZD only among RR-TB individuals. Not including other drugs used in the treatment of DR-TB in this model might have underrepresented the potential cost-effectiveness of tNGS.	systems: - We are not used to having a range of options in tests. Introduction of new tests brings with it confusion about how the tests go together. - The future will involve other tests and other drugs as well. - Need to apply creativity to problem solving - Cost-effectiveness analysis (CEA) modeling does not take into account reduction in transmission of MDR-TB that tNGS could bring - Need to keep in mind that it took a long time for Xpert to be taken up and implemented in an appropriate way, and a lot of costs to be incurred along the way - Moving forward we need data from pragmatic trials on aspects of costs and cost-effectiveness.
Equity What would be the impact on health equity?		
○ Reduced ○ Probabl	From a qualitative study of relevant stakeholders, the following considerations on the impact of tNGS on equity were found:	- The group shared positive

y reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	**- Centralized vs decentralized placement may have equity implications for access.** Given high-level specialised laboratory infrastructure, specialized human resources and technical complexity, tNGS technology is only suitable for placement at centralized, reference laboratories. This may have equity access considerations as it may mean less access for some regions of a country without reference labs. **- Affordability and cost-effectiveness are major concerns:** There was a major concern about financial costs of the tNGS technology and the affordability for LMICs. Participants were worried about not only the cost of the equipment, but also the costs of ongoing specialist supplies, especially for reagents, as well as the cost of maintaining equipment. They noted that costing calculations should be comprehensive and should include the cost of specialist consumables, extra general laboratory consumables, and the additional infrastructure needs (such as the extra space, temperature control, and internet connectivity needs). Cost-effectiveness calculations should also include assessment of the impact of the use of tNGS testing on improving TB outcomes in comparative studies. **- The MDR-TB case burden of a country could influence equitable access at centralized level**s. In some settings with high caseloads, the tNGS technology in central laboratories may not be sufficient for processing large caseloads in good time, and in settings with low caseloads, waiting for sufficient samples to batch will also cause delays.	views regardinghow this technology might improve equity but important limitations - The group noted that programmes have access challenges for tNGS but also equally for other pDST methods - The group considered that tNGS is going to be more scalable than phenotypic DST, in which case it would increase equity, but this remains a research question. - However, low bacillary load issues will impact equity for populations that often have paucibacillary disease, including , children and people living with HIV. For example, we know that for Xpert there is a higher indeterminate rate for children due to low bacillary load. . How to deal with this issue is also a research question; performing the

		test on culture may reduce indeterminate rates. -
Acceptability Is the intervention acceptable to key stakeholders?		
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	From a qualitative study of relevant stakeholders (laboratory staff and management who were involved with testing tNGS platforms in the three FIND trial sites, India, Georgia, and South Africa) **acceptability of tNGS technology was high.** There was an overwhelmingly positive sentiment for the potential utility of tNGS, and it was seen as a 'major advancement' in molecularly MDR TB diagnostics. - The main reasons for the **high level of acceptability were the comprehensiveness** (resistance diagnosis for more drugs and for newest and repurposed drugs), **the convenience of using sputum sample** (as compared to culture samples), **and the rapidness** (quick results compared to phenotypic testing times; 3-5 days as compared to 4-6 weeks). - There was also the sense that there is **a good window of opportunity for the utility of tNGS technology;** that the technology is arriving at the right time given that resistance to newer TB drugs is likely to increase as use if these drugs become routine.	The group noted that: - For a TB patient, knowing they are receiving the correct treatment is very important. - However, currently we are depending on personnel who are overburdened and have little time to explain results to patients. This will work differently in programmatic use than in clinical studies, where patients are lost and separated from their results in the overall. - We acknowledge that we don't yet have data from patients and doctors about how they would perceive this test/technology. Given the importance of understanding perspectives from key affected communities,

		efforts will need to be made to translate and inform those communities about this technology and the recommendations. - The group feels that the technology is probably acceptable, if a setting has the resources required to do it. The question will be whether it be acceptable to the funders of health programmes.
Feasibility Is the intervention feasible to implement?		
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	From a qualitative study of stakeholders from the FIND studies, the following considerations regarding the feasibility of TNGS were reported: **- Start-up and setting up challenges:** There were multiple starting and setting up problems. Some related to the newness of the technology and the trial setting, problems with importing technology and specialist supplies, problems related to absence of in-country technical assistance for problem-solving, as well as need for more hands-on training practice. **- High technical complexity of the test is a challenge:** tNGS technology was viewed as a high complexity molecular test that was technically challenging. For example, preparing the sample for sequencing involves multiple steps, that require attention to detail, precision, and with little room for error. The complexity of the library preparation phase was more particular for the Deeplex platform, though both the Deeplex and the Nanopore platforms were thought to have different pros and cons in terms of complexity. Both platforms were thought to have insufficient opportunities for early error recognition and error correction, and this increased the risk of failed runs. **- Specialized laboratory infrastructure and human resources are required which are potentially challenging**: As tNGS is a molecular-based testing platform, the platform requires highly specialised laboratory infrastructure that includes multiple rooms to prevent contamination and specialized cold storage facilities. Highly specialized molecular/medical scientists are needed to perform the tests. In these LMIC settings, such specialized laboratory infrastructure and staff may only be available at centralized laboratories and not necessarily at regional laboratories. **- Specialist requirements for operating the test are potentially challenging:** In addition to highly specialized laboratory infrastructure and staff, the testing	There is discussion of what implementation guidance will be provided to countries to help countries implement these tools? Many countries adopt WHO guidelines directly, larger countries adapt and adopt as it fits them, and with many delays. Many countries feel that WHO is moving too quickly and that hampers countries' ability to

	technology also requires uninterrupted supply of electricity, high internet connectivity, high computer capacity, clean water, and temperature controls - requirements that may pose challenges in some LMIC settings. **- Supply chain challenges was an obstacle:** A major concern was the supply chain challenges - procurement bottle-necks and delays jeopardized continuous access to specialist supplies. **- Data management and storage requirements presented challenges:** There were concerns that data analysis and data storage requirements were not fully developed, including systems for backing up data, data ownership and data security considerations. Consideration is needed for how tNGS and routine laboratory information systems would be interlinked. **- Continuous updating of the WHO mutations reference library would be required:** There is the sense that the usefulness of the tNGS technology is dependent on the informational support provided by the WHO mutations reference library, which allows for meaningful interpretation of resistance data; and thus, there is a need for the WHO reference library to be continuously updated. **- There are different feasibility concerns for the different tNGS platforms**: The overall sentiment that is that all three the tNGS platforms needed to be further developed before being fully ready for operational use, some more than others. The high level of technical complexity of the sample preparation stages (mainly the library preparation stage) was considered a key challenge for the Deeplex platform, and the need for improved computer analysis and storage capacity was a challenge for the Oxford Nanopore (ONP) platform, though both required a high level of precision and attention to detail, and more steps for early error recognition. The third platform was not ready for testing in two sites. Participants did not want to express explicit preference for one tNGS platform over the other, noting that both Deeplex and ONT had their pros and cons, and that both needed further development to be fit for purpose.	implement WHO recommendations. However it provides an opportunity for countries to procure through processes that rely on WHO recommendations – it is a gatekeeper. The group notes that this is probably feasible, depending on resources available.

Summary of judgements

	JUDGEMENT						
Problem	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Test accuracy	Very inaccurate	Inaccurate	**Accurate**	Very accurate		Varies	Don't know
Desirable Effects	Trivial	Small	Moderate	**Large**		Varies	Don't know
Undesirable Effects	Large	**Moderate**	Small	Trivial		Varies	Don't know
Certainty of the evidence of test accuracy	Very low	**Low**	Moderate	High			No included studies
Certainty of the evidence of test's effects	Very low	Low	Moderate	High			**No included studies**
Certainty of the evidence of	**Very low**	Low	Moderate	High			No included

management's effects							studies
Certainty of the evidence of test result/management	Very low	Low	Moderate	High			**No included studies**
Certainty of effects	Very low	Low	Moderate	High			**No included studies**
Values	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
Balance of effects	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know
Resources required	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
Certainty of evidence of required resources	**Very low**	Low	Moderate	High			No included studies
Cost effectiveness	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	**Varies**	No included studies
Equity	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
Acceptability	No	Probably no	**Probably yes**	Yes		Varies	Don't know
Feasibility	No	Probably no	**Probably yes**	Yes		Varies	Don't know

Type of recommendation

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

Conclusions

Recommendation

In people with bacteriologically-confirmed rifampicin-resistant pulmonary TB disease, targeted next-generation sequencing technologies may be used on respiratory samples to diagnose resistance to isoniazid, fluoroquinolones, bedaquiline, linezolid, clofazimine, pyrazinamide, ethambutol, amikacin, and streptomycin rather than culture-based phenotypic drug susceptibility testing *(conditional recommendation, certainty of evidence high [isoniazid, fluoroquinolones and pyrazinamide], moderate [ethambutol], low [bedaquiline, linezolid, clofazimine and streptomycin], very low [amikacin]).*

Subgroup considerations

In people with bacteriologically-confirmed rifampicin-resistant pulmonary TB disease, priority should be given to those at a higher risk of resistance to medications used for the treatment of rifampicin-resistant TB (RR-TB), including individuals who:

- continue to be smear or culture positive after two months or more of treatment or have experienced treatment failure,
- have a history of prior exposure to TB treatment, including the new and repurposed drugs,
- are in contact with a person known to have resistance to TB drugs, including the new and repurposed drugs,
- have pre-extensively drug-resistant (pre-XDR) TB with resistance to fluoroquinolones.

Priority should be given to samples with a high bacillary load as determined by initial bacteriological tests (e.g., semi-quantitative high/medium or smear-positive grading). In situations where the bacillary load is low (e.g., semi-quantitative grading of low/very low/trace or smear-negative), the recommendations still hold while acknowledging the higher rates of indeterminate results. Therefore, phenotypic DST is likely still required for low bacillary load samples
Similarly, the recommendations apply to children, adolescents, and people living with HIV (PLHIV), acknowledging the higher risk of indeterminate results in these populations due to higher frequency of samples with low bacterial load.
The recommendation is based on data obtained from sputum and bronchoalveolar lavage specimens and can be extrapolated to other lower respiratory tract samples (e.g., endotracheal aspirates). However, further research is needed to evaluate the use of these tests on alternative sample types for diagnosing pulmonary TB in children (such as nasopharyngeal and stool samples) and diagnosing extra-pulmonary TB.

Implementation considerations

- Targeted next-generation sequencing is a high-complexity test in its current format and is most suitable for centralised laboratories equipped with specialised skills and infrastructure.
- .
- Since sensitivity for bedaquiline, linezolid and clofazimine resistance is suboptimal, due consideration of the pre-test probability is important in interpreting the targeted NGS results for these drugs. Further testing of samples with a susceptible result, using culture-based phenotypic DST, would be warranted particularly when risk of resistance is high. Since specificity is high, a resistant result may be used to guide the therapy, particularly among those at risk for resistance.. It should also be noted that the basis of pretomanid resistance has not been fully elucidated and culture based DST is also required for this drug.

Monitoring and evaluation

- Standardize the nomenclature for result reporting across different targeted NGS technologies for integration into health information data systems.
- Ensure separate recording of true failures and unclassified mutations, and monitor trends over time as an essential component of result reporting.
- Regularly monitor performance data, including overall resistance rates, resistance rates by specific drugs or targets and turnaround times (both total and in-laboratory).
- Incorporate quality monitoring measures, such as tracking indeterminate rates, sequencing coverage and depth, and participation in external quality assurance programmes.
- Establish an external quality assurance programme for sequencing that covers all relevant targets of interest.
- Integrate the sequencing data generated into existing surveillance systems to monitor the prevalence and trends in drug resistance effectively. Share the data to update the WHO mutation catalogue.
- Collect cost data to address important questions, such as the costs associated with introducing and scaling up targeted NGS in different settings, the trade-offs between turnaround time and batching, and the optimal balance in various settings.
- Assess the impact of multi-disease testing on program operations and costs, including disease-specific testing volumes, turnaround times, costing, resource sharing, and resource requirements.
- Evaluate the impact of time to treatment initiation/modification, treatment outcomes, and overall cost-effectiveness of targeted NGS implementation.

Research priorities

Clinical research needs:

- Conduct clinical trials to assess the impact of targeted NGS on patient-important outcomes.
- Assess the accuracy and effectiveness of targeted NGS for detecting resistance to new and repurposed drugs, including pretomanid, across varied geographic and epidemiologic settings.
- Assess the accuracy and effectiveness of targeted NGS for analyzing extra-pulmonary samples, including cerebrospinal fluid for meningitis, non-sputum samples (such as nasopharyngeal aspirate, gastric aspirate, stool) for children, and alternative sample types (e.g., tongue swabs) in both adults and children.
- Undertake additional qualitative and quantitative research to further understand the perspectives of end-users and clinicians regarding the acceptability and feasibility of using targeted NGS.

Implementation research needs:

- Develop and evaluate effective and efficient implementation models by integrating targeted NGS into laboratory networks and optimizing algorithms, aiming to enhance timely access to testing, treatment initiation and improve patient outcomes.
- Develop strategies to enhance the efficiency of targeted NGS testing, including sample concentration techniques, determining optimal thresholds of bacterial load from initial tests before performing targeted NGS, utilizing pooled samples from multiple individuals, and employing molecular transport medium for ambient storage and transfer of samples to testing sites.
- Regularly update the interpretive catalogue based on WHO updates, incorporating additional genetic targets (for future tests) to enhance the sensitivity and specificity of targeted NGS and include new drugs used for TB treatment (e.g. pretomanid)
- Explore technological advancements to simplify the testing process, automate steps (especially library preparation), develop decentralized targeted NGS solutions, and investigate potential synergies with existing initial tests (e.g., utilizing leftover DNA or smear-positive slides).
- Conduct comprehensive mapping of sequencing capacity within countries and perform diagnostic network optimization exercises. Placement of the technology should consider the demand across multiple diseases, facilitating multiplex use of the machines and shared costs.
- Compile and utilize lessons learned from applying targeted NGS technology in other diseases to guide

implementation strategies for TB effectively.

For further information, please contact:
Global Tuberculosis Programme
World Health Organization
20 Avenue Appia CH-1211 Geneva 27 Switzerland
Web site: https://www.who.int/teams/
global-tuberculosis-programme/overview