WHO consolidated guidelines on tuberculosis

Module 2: Screening

Systematic screening for tuberculosis disease

Web Annex C.
GRADE Evidence to Decision Tables

World Health Organization

WHO consolidated guidelines on tuberculosis. Module 2: Screening - systematic screening for tuberculosis disease

Web Annex C. GRADE Evidence to Decision Tables

WHO consolidated guidelines on tuberculosis. Module 2: screening - systematic screening for tuberculosis disease. Web Annex C. GRADE evidence to decision tables

ISBN 978-92-4-002271-3 (electronic version)

© World Health Organization 2021

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

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

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

Suggested citation. Web Annex C. GRADE evidence to decision tables. In: WHO consolidated guidelines on tuberculosis. Module 2: screening - systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

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

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

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

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

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

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

This publication forms part of the WHO guideline entitled *WHO consolidated guidelines on tuberculosis. Module 2: screening - systematic screening for tuberculosis disease.* It is being made publicly available for transparency purposes and information, in accordance with the *WHO handbook for guideline development*, 2nd edition (2014).

Web Annex C. GRADE Evidence to Decision Tables

Table 1. Should systematic screening for TB disease, compared to passive case detection, be conducted in the general population? What tools should be used to screen for TB disease in the general population?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	The WHO's End TB Strategy envisions a 90% reduction in tuberculosis (TB) incidence and 95% reduction in TB deaths by 2035, and the Resolution adopted by the United Nations General Assembly in September 2018 commits to diagnosing and treating 40 million people with TB and treating at least 30 million people for TB infection by 2022. These targets will not be met by current practices in case detection. In order to achieve these ambitious targets there is an urgent need to deploy strategies to improve detection of people with active TB. In 2013 the WHO issued guidelines that general population screening may be considered only in defined areas or subpopulations with a very high prevalence of undetected TB (1% or greater) or in remote areas with poor access to health care, based on very low quality of evidence. Since the publication of the guidelines in 2013, there have been new studies, including reviews, randomized controlled trials, observational studies, and other research evaluating the real or potential impact of screening interventions on both individual-level and community-level outcomes related to TB. There have also been numerous prevalence surveys since 2013, which have shed new light on the magnitude of the burden of TB in several key countries.	

TB prevalence survey results for the past 10 years

Prevalence of bacteriologically-confirmed cases among people 15 years and older

Country	Year	Prevalence estimate (per 100,000)	CI lower bound	CI upper bound	
AFRO region					
Ethiopia	2011	277	208	347	0.1-0.3%
Gambia	2012	212	152	272	0.3-0.5%
Nigeria	2012	524	378	670	0.5-1%
Rwanda	2012	119	79	160	>1%
UR Tanzania	2012	590	330	860	
Ghana	2013	356	288	425	
Malawi	2013	452	312	593	
Zambia	2014	638	502	774	
Zimbabwe	2014	344	268	420	
Kenya	2015	558	455	662	
Uganda	2015	401	292	509	
Namibia	2018	465	340	590	
Eswatini	2018	352	264	440	
Lesotho	2019	581	466	696	
South Africa	2020	Pending			
Mozambique	2020	Pending			
EMRO region					
Pakistan	2011	398	333	463	
Sudan	2013	183	128	238	
SEARO region					
Thailand	2012	242	176	322	
Indonesia	2013	759	590	961	
Bangladesh	2015	287	244	330	
DPR Korea	2016	587	520	655	
Myanmar	2018	468	390	546	
Nepal	2018	374	308	441	
WPRO region					
China	2010	119	103	135	
Cambodia	2011	831	707	977	
Lao PDR	2011	595	457	733	
Mongolia	2015	560	455	665	
Philippines	2016	1159	1016	1301	
Viet Nam	2018	322	260	399	

Observational data of screening interventions among the general population from the last ten years (2010-2020) show information about the number needed to screen (NNS):

- Of 25 studies using symptom-based screening, the mean NNS was 1,296 (range 31-6,031). Of these, among 3 studies conducted in moderate burden settings, the mean NNS was 4,424 (range 2,417-6,031); in 18

studies conducted in medium burden settings, the mean NNS was 1146 (range 40-4085); in 4 studies conducted in high burden settings the mean NNS was 964 (31-1699).

- Of 4 studies using CXR-based screening, the mean NNS was 1,492 (range 186-3,016). Of these, 1 was conducted in a moderate burden setting (NNS 3,016) and 3 were conducted in medium burden settings with mean NNS 475 (range 186-695).
- 18 studies used a combined screening tool of symptoms or CXR. The mean NNS was 488 (range 23-2,857). The mean NNS for 3 studies conducted in moderate burden settings was 1,567 (range 23-2,853). The mean NNS for 15 studies conducted in medium burden settings was 444 (range 173-763). The mean NNS for 3 studies conducted in high burden settings was 165 (range 125-189).
- In 2 studies that used Xpert as a screening tool, the mean NNS was 1,002 (range 338-1,010).

Moderate (30-100/100k)				Medium (100-300/100k)				High (>300/100k)			
Symptoms	CXR	Sx/CXR	MWRD	Symptoms	CXR	Sx/CXR	MWRD	Symptoms	CXR	Sx/CXR	MWRD
4424 (2417-6031) n=3	3016 n=1	1567 (23-2853) n=3	-	1146 (40-4085) n=18	475 (186-695) n=3)	444 (173-763) n=15	1010 n=1	964 (31-1699) n=4	-	165 (125-189) n=3	338 n=1

Results presented as NNS (range), n = number of studies

For the current revision of the guidelines for screening for active TB, the question to the Guideline Development Group (GDG) is whether this recommendation should be updated, in light of current evidence, and whether the group would like to make specific recommendations for screening tests and algorithms to be used.

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS

○ Very inaccurate
○ Inaccurate
● Accurate
○ Very accurate
○ Varies
○ Don't know

Chest radiography (CXR - any abnormality):

Sensitivity: 0.94 (95% CI: 0.92 to 0.96) Specificity: 0.89 (95% CI: 0.85 to 0.92)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%	Prevalence 1%	Prevalence 2%		
True positives patients with active TB	5 (5 to 5)	9 (9 to 10)	19 (18 to 19)	4.243 (22)	⊕◯◯◯ VERY LOW[a,b,c]
False negatives patients incorrectly classified as not having active TB	0 (0 to 0)	1 (0 to 1)	1 (1 to 2)		
True negatives patients without active TB	884 (848 to 912)	880 (844 to 908)	871 (835 to 899)	1012752 (22)	⊕⊕◯◯ LOW[d,e,f]
False positives patients incorrectly classified as having active TB	111 (83 to 147)	110 (82 to 146)	109 (81 to 145)		

a. Only 2 studies had low risk of bias in the reference standard domain. Less than half of the studies had low risk in the flow-and timing domain
b. Moderate range in sensitivity (70%-100%) with some overlap in CIs. Variables that may explain observed variation are WHO region (Africa vs Asia/Pacific/other), prevalence of tuberculosis in the study population, and prevalence of smoking in the population (10% or more vs. lower).
c. CIs around the FN are narrow (relative to the point estimate)
d. Due to the low prevalence in the studies the Reference standard and Flow and Timing issues do not affect specificity as much as sensitivity.
e. Moderate in specificity (71%-99%). Variable that may explain observed variation is whether the CXR was read of any abnormality including other visible organs (82.4%, 95% CI 73.8%- 88.6%) vs. pulmonary abnormalities (91.1%, 95% CI 87.8%-93.5%).

The GDG considered that the accuracy of the tests could be ranked as follows:

- CXR - very accurate (is the only screening tool that meets the WHO 2014 Targte Product Profile, or TPP, for a screening test - very high sensitivity and specificity)
- Molecular WHO-approved Rapid Diagnostic Tests (mWRDs) - accurate (relatively high sensitivity, very high specificity)
- Symptoms - any TB symptom - accurate (higher sensitivity and lower specificity)
- Symptoms - prolonged cough - inaccurate (low sensitivity and higher specificity)

The GDG notes that factors related to patient selection, flow and timing may affect the

f. The CI around the FP is as such that the proportion of the population requiring follow up testing can vary by almost a factor two, which has serious resource implications.

measures of accuracy seen here for CXR for TB.

Chest radiography (CXR - suggestive for TB):

Sensitivity: 0.85 (95% CI: 0.77 to 0.90) Specificity: 0.96 (95% CI: 0.93 to 0.97)

The GDG notes that different symptom screening approaches have varying trade-offs in sensitivity and specificity. The feasibility of implementing symptom screening makes it a much more programmatically accessible option.

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%	Prevalence 1%	Prevalence 2%		
True positives patients with active TB	4 (4 to 5)	8 (8 to 9)	17 (15 to 18)	2.152 (19)	⊕⊕◯◯ LOW[a,b,c]
False negatives patients incorrectly classified as not having active TB	1 (0 to 1)	2 (1 to 2)	3 (2 to 5)		
True negatives patients without active TB	951 (922 to 969)	946 (917 to 964)	937 (908 to 954)	464818 (19)	⊕⊕⊕⊕ HIGH[d,e,f]
False positives patients incorrectly classified as having active TB	44 (26 to 73)	44 (26 to 73)	43 (26 to 72)		

Screening with CXR among people who are asymptomatic has additional ethical implications.

a. Only 3 of the 19 studies had low risk of bias in the Reference standard domain. Only 3 of 19 the studies had low risk in the Flow-and Timing domain. The sensitivity in studies with low risk in domain 3 or domain 3 is lower compared to studies with high or unknown risk.
b. Wide range in sensitivity (37%-100%) with some overlap in CIs. Variables that may explain observed variation are WHO region (Africa vs Asia/Pacific/other), and HIV prevalence although the latter was not statistically significant (p 0.074)
c. CIs around the FN are narrow (relative to the point estimate)
d. Due to the low prevalence in the studies the Reference standard and Flow and Timing issues do not affect specificity as much as sensitivity.
e. Range in specificity fairly narrow (84%-100%). None of the examined variables significantly modified the pooled specificity estimate.
f. The proportion false-positives (i.e. requiring further confirmatory testing) ranges from 2.6% to 7.2% of 1000 persons screened, which is reasonably precise, as it remains a fairly low proportion.

Prolonged cough (2 weeks or more):

Sensitivity: 0.42 (95% CI: 0.36 to 0.48) Specificity: 0.94 (95% CI: 0.92 to 0.96)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%	Prevalence 1%	Prevalence 2%		
True positives patients with active TB	2 (2 to 2)	4 (4 to 5)	8 (7 to 10)	6.737 (40)	⊕◯◯◯ VERY LOW[a,b,c]
False negatives patients incorrectly classified as not having active TB	3 (3 to 3)	6 (5 to 6)	12 (10 to 13)		
True negatives patients without active TB	938 (920 to 953)	934 (915 to 948)	924 (906 to 938)	1284181 (40)	

False positives patients incorrectly classified as having active TB	57 (42 to 75)	56 (42 to 75)	56 (42 to 74)		⊕⊕⊕⊕ HIGH[d,e,f]

a. QUADAS-2 Reference standard: more than three quarter of the studies did not require all participants to undergo bacteriological testing, but classified TB negative in those participants based on results of CXR and symptoms (incorporation bias). Flow and Timing: More than half of the studies scored high risk of bias. Of all participants who required bacteriological testing based on the protocol, less than 95% had a result. Sensitivity analysis showed that studies with low risk bias in these QUADAS-2 domains had considerably lower sensitivity (most extreme: studies with low risk for Reference standard (8 studies): sensitivity 29.3% (95% CI 19.4% - 41.7%)
b. Very wide range in point estimates (10% to 100%), with some overlap of the CIs. In stratified analysis, population level variables that significantly (p<0.05 modified the pooled estimates were economic region and higher vs. lower (<0.5%) tuberculosis prevalence among the study participants. Study design variables that significantly modified the pooled estimates were presence of incorporation bias and whether the reference standard included culture or not (but a combination of smear and Xpert MTB/RIF).
c. CIs around the FN are not very wide (relative to the point estimate)
d. Due to the low prevalence in the studies the Reference standard and Flow and Timing issues do not affect specificity as much as sensitivity.
e. Wide range in point estimates (spec 68% - 99%) but considerable overlap of CI. A few outlying values are of studies that share a quality concern in the patient selection domain. Variables that may explain heterogeneity in specificity were economic region and tuberculosis prevalence among the study participants.
f. The proportion false-positives (i.e. requiring further confirmatory testing) ranges from 4% to 7.6% of 1000 persons screened, which is reasonably precise.

Any cough:

Sensitivity: 0.51 (95% CI: 0.43 to 0.60) Specificity: 0.88 (95% CI: 0.82 to 0.92)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%	Prevalence 1%	Prevalence 2%		
True positives patients with active TB	3 (2 to 3)	5 (4 to 6)	10 (9 to 12)	2.734 (21)	⊕◯◯◯ VERY LOW[a,b,c]
False negatives patients incorrectly classified as not having active TB	2 (2 to 3)	5 (4 to 6)	10 (8 to 11)		
True negatives patients without active TB	871 (812 to 913)	867 (808 to 908)	858 (800 to 899)	768.291 (21)	⊕⊕◯◯ LOW[d,e,f]
False positives patients incorrectly classified as having active TB	124 (82 to 183)	123 (82 to 182)	122 (81 to 180)		

a. QUADAS-2 Reference standard: more than half of the studies did not require all participants to undergo bacteriological testing, but classified TB negative in those participants based on results of CXR and symptoms (incorporation bias). Flow and Timing: about one third of the studies scored high risk of bias. Of all participants who required bacteriological testing based on the protocol, less than 95% had a result. Sensitivity analysis showed that studies with low risk bias in these QUADAS-2 domains had considerably lower sensitivity (most extreme: studies with low risk for Reference standard (8 studies): sensitivity 35.6% (95% CI 18.8% - 56.8%)
b. Very wide range in point estimates (0% to 100%), with some overlap of the CIs. Some of the heterogeneity could be explained by economic region. Studies in low income countries showed higher sensitivity (64.8%, 54.8-73.6%), in upper/middle/high income studies sensitivity was lower (34.4%, 23.3-47.5%).
c. CIs around the FN are not very wide (relative to the point estimate)
d. Due to the low prevalence in the studies the Reference standard and Flow and Timing issues do no affect specificity as much as sensitivity.

e. Wide range in point estimates (spec 43% - 99%) without overlap of CI. No statistical significant variables that could explain heterogeneity, however in low income countries the sensitivity was somewhat lower (80.8%, 69.1-88.9%) than in the upper/middle/high income studies.
f. The CI around the FP is as such that the proportion of the population requiring follow up testing can vary by more than a factor two, which has serious resource implications.

Any TB symptom:

Sensitivity: 0.71 (95% CI: 0.62 to 0.79) Specificity: 0.64 (95% CI: 0.52 to 0.74)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%	Prevalence 1%	Prevalence 2%		
True positives patients with active TB	4 (3 to 4)	7 (6 to 8)	14 (12 to 16)	3915 (28)	⊕◯◯◯ VERY LOW[a,b,c]
False negatives patients incorrectly classified as not having active TB	1 (1 to 2)	3 (2 to 4)	6 (4 to 8)		
True negatives patients without active TB	634 (515 to 739)	631 (512 to 735)	625 (507 to 728)	460.878 (28)	⊕⊕◯◯ LOW[d,e,f]
False positives patients incorrectly classified as having active TB	361 (256 to 480)	359 (255 to 478)	355 (252 to 473)		

a. QUADAS-2 Reference standard: more than half of the studies did not require all participants to undergo bacteriological testing, but classified TB negative in those participants based on results of CXR and symptoms (incorporation bias). Flow and Timing: about one third of the studies scored high risk of bias. Of all participants who required bacteriological testing based on the protocol, less than 95% had a result. Sensitivity analysis showed that studies with low risk bias in these QUADAS-2 domains had considerably lower sensitivity (most extreme: studies with low risk for Reference standard (12 studies): sensitivity 62.9% (95% CI 47.4% - 76.1%) and Flow and Timing (9 studies): sensitivity 62.9% (43.5 - 78.9%)
b. Very wide range in point estimates (18% to 100%), with overlap of the CIs. Some of the heterogeneity could be explained by economic region. Studies in low income countries showed higher sensitivity (78.9%, 69.3-86.2%), in upper/middle/high income studies sensitivity was lower (56.3%, 40.6-70.8%).
c. CIs around the FN are not very wide (relative to the point estimate)
d. Due to the low prevalence in the studies the Reference standard and Flow and Timing issues do not affect specificity as much as sensitivity.
e. Wide range in point estimates (13% - 99%) without overlap of CI. No statistical significant variables that could explain heterogeneity.
f. The CI around the FP is as such that the proportion of the population requiring follow up testing can vary by almost a factor two, which has serious resource implications.

Indirect data for Molecular WHO-approved rapid diagnostics (mWRDs) for screening in the general population - the following data are pooled across different populations of high-risk individuals (miners, prisoners, contacts):

Molecular WHO-approved rapid diagnostics:

Sensitivity: 0.69 (95% CI: 0.48 to 0.86) Specificity: 0.99 (95% CI: 0.97 to 0.99)

Test result	Number of results per 1000 patients tested (95% CI)			№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%	Prevalence 1%	Prevalence 2%		
True positives patients with pulmonary tuberculosis	3 (2 to 4)	7 (5 to 9)	14 (10 to 17)	337 (5)	⊕◯◯◯ VERY LOW[a,b,c]
False negatives patients incorrectly classified as not having pulmonary tuberculosis	2 (1 to 3)	3 (1 to 5)	6 (3 to 10)		
True negatives patients without pulmonary tuberculosis	983 (967 to 990)	978 (962 to 985)	968 (953 to 975)	8619 (5)	⊕⊕◯◯ LOW[a]
False positives patients incorrectly classified as having pulmonary tuberculosis	12 (5 to 28)	12 (5 to 28)	12 (5 to 27)		

a. 'General population' is a broad category. Studies contributing to this pooled estimate included adults residing in prisons, household contacts of persons with TB, and miners. There is uncertainty associated with applicability to the general population. Additionally, one of the studies included a small number of children (age < 15) in the screened population, which deviates from the intended study population. We downgraded two levels for indirectness.
b. Sensitivity estimates ranged from 33% to 100%. We thought this variability could partly be explained by the different high-risk groups in this analysis. We downgraded one level for inconsistency.
c. The 95% CrI is wide. We thought the 95% CrI around true positives and false negatives would likely lead to different decisions depending on which limits are assumed. As we had already downgraded for inconsistency, we did not downgrade further for imprecision.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ○ Moderate ○ Large ● Varies ○ Don't know	**Specific to screening tests:** **Desirable: Identifying true positive and negatives** The anticipated desirable effect is the correct classification of people with active TB as screening positive for TB (true positive), resulting in appropriate referral for further evaluation (as indicated by the screening algorithm in use), and increasing the probability of ultimately leading to timely diagnosis and treatment. Another anticipated desirable effect is correctly ruling out disease in people who do not have TB (true negative), avoiding unnecessary further diagnostic evaluation (and associated resources for the person and the health system), providing reassurance to the person undergoing screening, allowing for the pursuit of an alternative diagnosis if they have respiratory symptoms, and allowing for the determination of eligibility for TB preventive therapy, if indicated. **Summary table**	The GDG notes that the scale and balance of desirable and undesirable effects will vary according the epidemiological conditions and intensity of the screening intervention being implemented (coverage of the population and sensitivity of the screening test and algorithm).

Test	Test accurac	Studies (persons	CoE	0.5% prevalence	1% prevalence	2% prevalence
Chest radiography (any abn)	Se 0.94	22 (4243)	Very low	TP:5 / FN:0	TP:9 / FN:1	TP:19 / FN:1
Chest radiography (any abn)	Sp 0.89	22 (1012752)	Low	TN:884 / FP:111	TN:880 / FP:110	TN:871 / FP:109
Chest radiography (sugg abn)	Se 0.85	19 (2152)	Low	TP:4 / FN:1	TP:8 / FN:2	TP:17 / FN:3
Chest radiography (sugg abn)	Sp 0.96	19 (464818)	High	TN:951 / FP:44	TN:946 / FP:44	TN:937 / FP:43
Prolonged cough (≥2 weeks)	Se 0.42	40 (6737)	Very low	TP:2 / FN:3	TP:4 / FN:6	TP:8 / FN:12
Prolonged cough (≥2 weeks)	Sp 0.94	40 (1284181)	High	TN:938 / FP:57	TN:934 / FP:56	TN:924 / FP:56
Any cough	Se 0.51	21 (2734)	Very low	TP:3 / FN:2	TP:5 / FN:5	TP:10 / FN:10
Any cough	Sp 0.88	21 (768291)	Low	TN:871 / FP:124	TN:867 / FP:123	TN:858 / FP:122
Any TB symptom	Se 0.71	28 (3915)	Very low	TP:4 / FN:1	TP:7 / FN:3	TP:14 / FN:6
Any TB symptom	Sp 0.64	28 (460878)	Low	TN:634 / FP:361	TN:631 / FP:359	TN:625 / FP:355
mWRDS	Se 0.69	5 (337)	Very low	TP:3 / FN:2	TP:7 / FN:3	TP:14 / FN:6
mWRDS	Sp 0.99	5 (8619)	Low	TN:983 / FP:12	TN:978 / FP:12	TN:968 / FP:12

Individual-level effects

No randomized trials addressed this question.

Treatment outcomes - success

Three observational studies found that the proportion of cases with treatment success may be similar between patients found through screening and through standard case detection (very low certainty evidence).

Treatment outcomes - fatality

Four observational studies found there may be no difference in the proportion of cases who died between patients found through screening and through standard case detection (very low certainty of evidence).

Time to diagnosis/disease severity at diagnosis

Three observational studies found that there may be fewer smear grade 2+ and 3+ in those TB cases identified through systematic screening compared to those detected through standard case detection (very low certainty evidence).

Community-level effects

TB prevalence

Data from two cluster randomized trials showed inconsistent findings. One trial from Zambia and South Africa showed there might be no impact of systematic screening using community mobilization and mobile sputum collection points on adult TB disease prevalence (adjusted risk ratio [aRR] 1.09 [95% CI: 0.86 – 1.40]) (low certainty evidence).

By contrast, a trial in Vietnam showed that systematic screening using three years of annual door to door sputum collection and testing using Xpert MTB/Rif reduced adult TB disease prevalence (aRR 0.56 [95% CI: 0.40 – 0.78]); and a non-controlled study in Zimbabwe reported five rounds of systematic screening interventions over three years (including community mobilization, mobile vans, and door to door screening campaigns) reduced TB prevalence (aRR 0.59 [95% CI: 0.40 - 0.89]), however this study was at a serious risk of bias due to unobserved and unmeasured secular trends (moderate certainty evidence).

One observational study in China reported on three rounds of systematic screening between 2013 and 2015, consisting of door-to-door symptom screening followed by CXR. The study found serial reductions in the absolute number of TB cases detected (very low certainty evidence).

Case detection/case notification

There is inconsistent evidence that systematic screening for TB may improve the detection of TB within the general population (very low certainty)

No randomized trials addressed this question. In four observational studies, the certainty of evidence was very low and with mixed findings including that systematic screening may initially increase TB case notification.

TST/IGRA positivity in children (TB incidence)

Both ZAMSTAR and ACT3 measured the effect of ACF on TB infection (defined as a positive TST or IGRA test) in children living in clusters with and without ACF (incidence of TB infection in children known to be TST negative at baseline in ZAMSTAR and prevalence of TB infection in ACT3).

In ZAMSTAR the mean TST positivity among children who had been TST negative at baseline was 1.41 per 100 person years in ACF clusters and 1.05 in non-ACF clusters, adjusted rate ratio 1.36 (95% CI: 0.59 – 3.14). In ACT3, among children born in 2012 (age 1 – 2 years when intervention started in 2014), prevalence ratio for QuantiFERON positivity

For screening tests (symptoms, CXR, mWRD):

The desirable effects will depend greatly on the diagnostic test used following the screening test. Across the screening tests, there is not much difference in absolute numbers of true positives and negatives, within a given prevalence setting. The data only relate to one round of screening; population-level screening using repeated screening rounds is not captured in this data.

It was mentioned in the GDG discussions that there is a risk that a positive mWRD on screening may not be followed by a definite diagnostic evaluation in a programmatic setting, thus programmatically using mWRDs as a screening test may risk more people being started wrongly on treatment compared to screening algorithms using CXRs and symptom screening.

- Chest radiography : Large desirable
- mWRDs: Moderate desirable
- Symptoms - any TB symptom: Moderate desirable
- Symptoms - prolonged cough: Moderate desirable

For effectiveness data on screening of the general population:

- Desirable effects of screening on population-level effects varies with

was 1.29 (95% CI: 0.70 – 2.36). Among children born between 2004 and 2011, prevalence was 4.1% in intervention group areas and 8.3% in control group areas (prevalence ratio 0.50 (95% CI: 0.32 – 0.78).

Other outcomes

No studies reported outcomes related to TB disease incidence, TB-specific mortality or all-cause mortality, or knowledge and health-seeking behavior in the general population.

the population being screened and the epidemiology of TB there, including: the prevalence of TB and portion of prevalent cases arising from recent transmission (eg TB in children transmitted in families), the sensitivity of the screening tool, the intensity/coverage of the screening approach

- The research presented here on individual- and community-level effects may not capture the longer-term morbidity and mortality averted through community screening.

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ○ Small ○ Trivial ● Varies ○ Don't know	**Specific to screening tests:** **Undesirable: Identifying false positive and negatives:** The anticipated undesirable effect is the incorrect classification of people without active TB as screening positive for TB (false-positive), with resulting anxiety, resources required for further unnecessary diagnostic evaluation (for both the patient and the health system), and the risk of a false-positive diagnosis, leading to inappropriate treatment and related social, economic, or health consequences (including adverse effects of treatment). Another anticipated undesirable effect is the incorrect classification of people who have active TB as screening negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, and potentially inappropriate preventive therapy treatment with associated risk of development of drug resistance. **Summary table**	*For screening tests (symptoms, CXR, mWRD)*: Varies for all. Mostly would be in terms of false-positive screening results, which could be expected to be mitigated by diagnostic steps later in the screening algorithm. Also it could result in "over-treatment" of latent TB infection, which would still have an effect on reducing TB burden moving forward. *For effectiveness data on screening of the general population*: Varies for all. - Undesirable effects of screening on population-level effects also vary with the population being screened and the epidemiology of TB there, including: the prevalence of TB, the sensitivity of the screening tool, the

Test	Test accurac	Studies (persons	CoE	0.5% prevalence	1% prevalence	2% prevalence
Chest radiography (any abn)	Se 0.94	22 (4243)	Very low	TP:5 / FN:0	TP:9 / FN:1	TP:19 / FN:1
Chest radiography (any abn)	Sp 0.89	22 (1012752)	Low	TN:884 / FP:111	TN:880 / FP:110	TN:871 / FP:109
Chest radiography (sugg abn)	Se 0.85	19 (2152)	Low	TP:4 / FN:1	TP:8 / FN:2	TP:17 / FN:3
Chest radiography (sugg abn)	Sp 0.96	19 (464818)	High	TN:951 / FP:44	TN:946 / FP:44	TN:937 / FP:43
Prolonged cough (≥2 weeks)	Se 0.42	40 (6737)	Very low	TP:2 / FN:3	TP:4 / FN:6	TP:8 / FN:12
Prolonged cough (≥2 weeks)	Sp 0.94	40 (1284181)	High	TN:938 / FP:57	TN:934 / FP:56	TN:924 / FP:56
Any cough	Se 0.51	21 (2734)	Very low	TP:3 / FN:2	TP:5 / FN:5	TP:10 / FN:10
Any cough	Sp 0.88	21 (768291)	Low	TN:871 / FP:124	TN:867 / FP:123	TN:858 / FP:122
Any TB symptom	Se 0.71	28 (3915)	Very low	TP:4 / FN:1	TP:7 / FN:3	TP:14 / FN:6
Any TB symptom	Sp 0.64	28 (460878)	Low	TN:634 / FP:361	TN:631 / FP:359	TN:625 / FP:355
mWRDS	Se 0.69	5 (337)	Very low	TP:3 / FN:2	TP:7 / FN:3	TP:14 / FN:6
mWRDS	Sp 0.99	5 (8619)	Low	TN:983 / FP:12	TN:978 / FP:12	TN:968 / FP:12

		intensity/coverage of the screening approach - We have very limited data to look at individual and community-level undesirable outcomes from TB screening, especially longer-term. - In terms of acceptibility of TB screening qualitative reviews reported mixed results, including benefits of having access to health care closer to home, but at times resulting in entering a disorganised or chaotic health service with some expressing doubts regarding the reluability of diagnostic testing. Note - children should be considered separately.

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 varies for sensitivity and specificity for the various screening tests. For CXR (suggestive of TB) - low certainty. For prolonged cough, any cough, any TB symptoms, CXR (any abnormality) , mWRDs- very low certainty.	

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; the direct benefits and adverse effects of the screening tests are unknown. The possible adverse effects of any screening test include additional burden of anxiety and time for the patient. The possible benefits of symptom screening for TB including detection of other conditions. Direct CXR is a safe technology using a radiation dose of 0.1 mSv, which corresponds to 1/30 of the average annual radiation dose from the environment (3 mSv) and 1/10 of the annual accepted dose of ionizing radiation for the general public (1 mSv). Innovations in x-ray technology in recent years have substantially reduced the radiation exposure levels. A proportion of the X-rays used in radiography are absorbed by the body. The potential effects from ionizing radiation depend on the dose. The long term risks from ionizing radiation include an increased risk of cancer. Therefore, exposure to the low radiation doses delivered to patients during a CXR poses a small risk. Children and pregnant women are especially vulnerable to ionizing radiation. Also, children have a longer life expectancy, resulting in a larger window for developing long-term radiation-induced health effects. The direct benefit or adverse effect of molecular WRD testing are unknown. Producing sputum is not easy or pleasant for patients.	The GDG considered that infection control is a potential concern when conducting an mWRD test, through the process of producing a sputum sample which may produce aerosols, which in turn may lead to the transmission of TB (or other respiratory pathogens). Therefore if using mWRDs with sputum sample, infection control measures need to be followed.

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	The primary objective of screening for active TB is to ensure that active TB is detected early and treatment is initiated promptly, with the ultimate aim of reducing the risk of poor treatment outcomes, health sequelae and the adverse social and economic consequences of TB, as well as helping to reduce TB transmission. Children aged < 5 years who are household contacts of people with bacteriologically confirmed pulmonary TB and who are found not to have active TB should be given TB preventive treatment (Strong recommendation, high certainty in the estimates of effect). Adults and adolescents who are household contacts of people with bacteriologically confirmed pulmonary TB who are found not to have active TB may be given TB preventive treatment. (Conditional recommendation, low certainty in the estimates of effect). People living with HIV who are unlikely to have active TB should receive TB preventive treatment as part of a comprehensive package of HIV care (Strong recommendation, high certainty in the estimates of effect). Treatment of drug sensitive TB is highly effective (approximately 90%). Treatment of MDR TB can be effective as well (approximately 70%) (WHO consolidated treatment guidelines 2020, Strong recommendation, high certainty in the estimates of effect). Effectiveness of preventive therapy is between 60-90% (WHO preventive treatment guidelines 2020). See here for evidence - https://tuberculosis.evidenceprime.com/	

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	*Linkage to care/initial presentation for treatment* Two observational studies reported that there may be higher proportions of initial default among people diagnosed through screening compared to standard case detection - very low quality of evidence.	The GDG considered that evidence for good linkage to care and very little initial loss to follow-up from trials not captured by the PICOs explored here (eg trials by Corbett, not yet published from Marks), suggesting this data may not capture all existing data on linkage to care among cases detected through screening. However, trial conditions do not reflect programmatic conditions. Hence linkage to care in programmatic conditions are expected to be lower compared to trials. It is paramount that screening is accompanied by good communication and support systems to ensure that those diagnosed with TB are also rapidly initiated on treatment. Also the GDG noted that the published literature is lacking in adequately measuring this outcome.

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with standard case detection	Risk difference with L systematic screening for active TB
Linkage to care - initial default	0 (2 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	**ACF n/N (%; 95%CI) vs PCF n/N (%; 95%CI)** • **Gopi 2005:** 57/243 (23%; 18-29%) vs 156/1049 (15%; 13-17%) • **Balasubramanian 2004:** 68/231 (29%; 24-36%) vs 120/833 (14%; 12-17%) [e]	

a. Population: the study population in both studies were smear +ve TB cases
b. All proportions similar with similar confidence intervals
c. sample sizes are relatively large.
d. None of the studies control for potential confounders. There were methodological issues and often insufficient information to determine bias domains across the studies.
e. Both studies done in the same population in South India but over different periods of time (Gopi: from January 2001 to December 2003; Balasubramanian: from December 1998 to November 2001).

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

Individual-level outcomes

No randomised trials addressed this question.

Treatment outcomes - success

Three observational studies found that the proportion of cases with treatment success may be similar between patients found through screening and through standard case detection (**very low certainty evidence**).

Treatment outcomes - fatality

Four observational studies found there may be no difference in the proportion of cases who died between patients found through screening and through standard case detection (**very low certainty of evidence**).

Time to diagnosis/disease severity at diagnosis

Three observational studies found that there may be fewer smear grade 2+ and 3+ in those TB cases identified through systematic screening compared to those detected through standard case detection (**very low certainty evidence**).

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with standard case detection	Risk difference with systematic screening for TB disease
Treatment outcome: treatment success (cured + treatment completed)	0 (3 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	**ACF n/N (%; 95%CI) vs SCD n/N (%; 95%CI)** • **denBoon 2008:** 16/20 (80%; 56-94%) vs 379/473 (80%; 76-84%) • **Santha 2003:** 45/65 (69%; 57-80%) vs 225/330 (68%; 63-73%) • **Harper 1996:** 50/64 (78%; 66-87%) vs 997/1272 (78%; 76-81%)	
Treatment outcome: case fatality	0 (4 observational studies)	⊕◯◯◯ VERY LOW[a,e,f,g]	-	**ACF n/N (%; 95%CI) vs SCD n/N (%; 95%CI)** • **denBoon 2008:** 2/27 (7%; 1-24%) vs 18/473 (4%; 2-6%) • **Santha 2003:** 4/65 (6%; 2-15%) vs 23/330 (7%; 4-10%) • **Cassels 1982:** 9/111 (8%; 4-15%) vs 17/159 (11%; 6-17%) • **Harper 1996:** 5/64 (8%; 3-17%) vs 104/1272 (8%; 7-10%) [h]	
Earlier case detection: severity at diagnosis - smear grade (proportion 2+ and 3+)	0 (3 observational studies)	⊕◯◯◯ VERY LOW[a,b,d,i]	-	**ACF n/N (%; 95%CI) vs SCD n/N (%; 95%CI)**	

The GDG felt that the level of certainty of the data for this recommendation should reflect the highest-quality trials addressing the PICO question, privileging that over observational data.

				• **Abdurrahman 2016:** 268/480 (56%; 51-60%) vs 151/208 (73%; 66-79%) • **denBoon 2008 :** 10/18 (56%; 31-78%) vs 314/446 (70%; 66-75%) • **Santha 2003:** 39/96 (41%; 31-51%) vs 228/330 (69%; 64-74%) j

a. None of the studies control for potential confounders. There were methodological issues and often insufficient information to determine bias domains across the studies.
b. All proportions similar with similar confidence intervals
c. Population: study population were smear-positive TB cases in 2 studies (Santha and Harper) and smear/culture-positive TB cases in 1 study (den Boon). Intervention: In 1 study (den Boon), there was no screening test applied. All individuals in the community survey were eligible for sputum smear and culture examination.
d. 1 study has a low number of TB cases in the ACF group (den Boon). But the remaining studies have relatively large numbers in both the ACF and PCF groups. This is reflected in the width of the CIs
e. Population: study population were smear-positive TB cases in 3 studies (Santha, Cassels and Harper) and smear/culture-positive TB cases in 1 study (den Boon). Intervention: In 1 study (den Boon), there was no screening test applied. All individuals in the community survey were eligible for sputum smear and culture examination.
f. All studies (proportions and CIs) are similar. The exception is den Boon - the total number of TB cases and events in the ACF group in this study is low, resulting in a very wide CI.
g. The number of events (deaths) is low.
h. 2 studies (den Boon, Cassels) includes initial defaulters in the ACF group alone.
i. There is no gold standard for severity diagnosis of TB. Smear grade is an indirect and imperfect measure of severity, especially in the context of high HIV prevalence.
j. 2 studies (den Boon, Santha) includes initial defaulters in the ACF group alone.

Community-level outcomes:

TB prevalence

Data from two cluster randomized trials showed inconsistent findings. One trial from Zambia and South Africa showed there might be no impact of systematic screening using community mobilization and mobile sputum collection points on adult TB disease prevalence (adjusted risk ratio [aRR] 1.09 [95% CI: 0.86 – 1.40]) (**low certainty evidence**).

By contrast, a trial in Vietnam showed that systematic screening using annual door to door sputum collection and testing using Xpert MTB/Rif over 3 consecutive years reduced adult TB disease prevalence (aRR 0.56 [95% CI: 0.40 – 0.78]); and a non-controlled study in Zimbabwe showed five rounds of systematic screening interventions over three years (including community mobilization, mobile vans, and door to door screening campaigns) reduced TB prevalence (aRR 0.59 [95% CI: 0.40 - 0.89]), however this study was at a serious risk of bias due to unobserved and unmeasured secular trends. (**moderate certainty evidence**).

One observational study in China reported on three rounds of systematic screening between 2013 and 2015, consisting of door-to-door symptom screening followed by CXR. The study found serial reductions in the absolute number of TB cases detected (**very low certainty evidence**).

Case detection/case notification

There is inconsistent evidence that systematic screening for TB may improve the detection of TB within the general population. **(very low certainty**)

No randomised trials addressed this question. In four observational studies, the certainty of evidence was very low and with mixed findings including that systematic screening may initially increase TB case notification.

TST/IGRA positivity in children

Both ZAMSTAR and ACT3 measured the effect of ACF on of TB infection (defined as a positive TST or IGRA test) in children living in clusters with and without ACF (incidence of TB infection in children known to be TST negative at baseline in ZAMSTAR and prevalence of TB infection in ACT3). In ZAMSTAR the mean TST positivity among children who had been TST negative at baseline was 1.41 per 100 person years in ACF clusters and 1.05 in non-ACF clusters, adjusted rate ratio 1.36 (95% CI: 0.59 – 3.14). In ACT3, among children born in 2012 (age 1 – 2 years when intervention started in 2014), prevalence ratio for QuantiFERON positivity was 1.29 (95% CI: 0.70 – 2.36). Among children born between 2004 and 2011, prevalence was 4.1% in intervention group areas and 8.3% in control group areas (prevalence ratio 0.50 (95% CI: 0.32 – 0.78). (**moderate certainty**)

Other outcomes

No studies reported outcomes related to TB-specific mortality or all-cause mortality, or knowledge and health-seeking behavior in the general population.

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with standard case detection	Risk difference with R v2 systematic screening for active TB
TB disease prevalence (ZAMSTAR) follow up: 4.5 years	90601 (1 RCT)[a]	⊕⊕⊕◯ MODERATE[b,c,d]	**RR 1.09** (0.86 to 1.40)	Study population	
				878 per 100,000[a]	**79 more per 100,000** (123 fewer to 351 more)[a]
TB disease prevalence (ACT3) follow up: 3 years	83830 (1 RCT)[e]	⊕⊕⊕◯ MODERATE[b,d]	**RR 0.55** (0.39 to 0.77)	Study population	
				226 per 100,000[e]	**101 fewer per 100,000** (138 fewer to 52 fewer)[e]
TB disease prevalence (DETECTB)	21303 (1 observational study)[f]	⊕◯◯◯ VERY LOW[g]	**RR 0.59** (0.40 to 0.89)	Study population	
				654 per 100,000[f]	**268 fewer per 100,000** (392 fewer to 72 fewer)[f]
TB disease prevalence (other non-randomised studies)	0 (1 observational study)	⊕◯◯◯ VERY LOW[h,i]	-	One study (**Liu et al**) among general population in China undertook three annual rounds of TB prevalence survey (the prevalence survey also met our definition of an ACF intervention) in three clusters (two rural, one urban) 2013-2015, People were assessed for TB by door to door symptom screening (everyone) and CXR (for people who had symptoms or were "high risk" for TB). Mean number of people screened each year was 91,754 (population denominator). In 2013, 35 people with TB identified. In 2014, 25 people with TB identified. In 2015, 15 people with TB identified. [j]	
Case notification rate (DETECTB)	0 (1 RCT)	⊕◯◯◯ VERY LOW[k,l]	-	**DETECTB** compared two different types of ACF interventions in Harare, Zimbabwe: door to door symptom screening vs. sputum collection in mobile vans with community mobilization (no standard case detection comparison). Mobile van ACF detected more TB cases than door to door ACF, risk ratio 1.48 (1.11 to 1.96). Very indirect evidence that ACF may have some effect on TB case notifications.	

Case notification rate (non randomised studies)	0 (4 observational studies)	⊕◯◯◯ VERY LOW[m,n,o,p,q]	-	Four observational studies using before-after design (with and without control groups). In general population. **Kan et al (2012)** showed CNR ratio in 24 counties in Anhui regio of China with ACF was 3.47 (comparing pre-ACF baseline CNR to during ACF endline CNR), in control counties in same region with no ACF case notifications also increased with CNR 3.14. Ratio of CNR ratios 1.19. Intervention counties population size 15 million people, control counties 29 million people. Co-intervention of financial incentives to local primary care doctors. **Cegielski et al (2013)** showed CNR in two neighborhoods in Texas, USA was 0 comparing before and after ACF (as ACF detected no cases). In the rest of the county, excluding the two neighborhoods that received ACF, CNR ratio baseline to endline was 0.66. 3000 people in ACF communities, not stated population of rest of county. Cointervention of LTBI treatment. **Parija et al (2014)** showed CNR ratio in 203 'sectors' in Odisha, India who were provided with ACF was 1.11 comparing baseline vs. endline CNRs. In 202 sectors without ACF CNR ratio was 1.01. Ratio of CNR ratios 1.10. Estimated 6 million people in contol sectors and 6 million in intervention sectors. No co-interventions. **Chen et al (2019)** showed CNR ratio in 10 communities in Yunnan Province, China provided with ACF was 0.86 comparing baseline to endline CNRs. In 136 communities in Yunnan Province without ACF, CNR ratio baseline to endline was 0.79. Ratio of CNR ratios 1.01. 35,000 people in intervention communities and 243,000 in control communities. No co-interventions.	
TST positivity in children (ZAMSTAR)	10103 (RCTs)	⊕⊕⊕◯ MODERATE[b,r]	**RR 1.36** (0.59 to 3.14)	Study population	
				66 per 1,000	**24 more per 1,000** (27 fewer to 142 more)
IGRA positivity in children (ACT3)	1484 (RCTs)	⊕⊕⊕◯ MODERATE[s]	**RR 0.50** (0.32 to 0.78)	Study population	
				41 per 1,000	**21 fewer per 1,000** (28 fewer to 9 fewer)

a. Denominator refers to number of adults who gave informed consent, completed questionnaire and provided a sputum sample that was evaluable.
b. Indirectness not strictly relevant as only one study per row (therefore not marked down). However, the approach taken by ZAMSTAR and ACT3 are very different to each other. ZAMSTAR used community mobilization, education and sputum drop off points (mobile sputum collection points and "fast track" at permanent facilities). ACT3 used annual door to door sputum collection (regardless of symptoms).
c. Downgraded by one level for serious imprecision. Confidence interval includes the null and substantial harm as well as modest benefit.
d. Some concerns of bias in measurement of outcome as relatively large numbers of enumerated individuals weren't approached, didn't consent, didn't produce sputum or didn't have a valid sputum result.
e. Denominator refers to number of adults enumerated as living in subcommands, contacted to give consent, capable to consent and actually consented to take part in survey. No requirement to actually provide sputum.
f. Denominator is number of adults (selected at random from intervention areas) who were located, consented to be surveyed and provided sputum.

	g. Doesn't control for secular trends in TB prevalence over time. TB prevalence is a before-after observational secondary outcome from a randomized trial. DETECTB had a larger proportion of adults enumerated who were found, consented, produced sputum and had a sputum result (81% of enumerated sample in baseline prevalence survey and 71% in endline prevalence survey) than ACT3 or ZAMSTAR. h. No confidence interval provided i. Assessed using ROBINS-i. Multiple issues identified, including no accounting for confounding or temporal trends in TB case notifications. j. Not possible to give a confidence interval due to no estimate of clustering available to adjust for. No adjustment for confounding (by secular trends or any other potential confounder). Authors report p value for each pairwise comparison in each of three sites (i.e. 2013 vs. 2014 site A, 2013 vs 2015 site B etc.). The difference in people with TB identified 2013 vs 2015 was reported to be statistically significant (p<0.05) in one of three sites. k. TB case detection through ACF methods was the primary outcome of DETECTB. l. Trial compared two methods of ACF (door to door symptom screening and mobile vans for sputum collection) to each other. No comparison to standard case detection. Additionally, the primary outcome is TB cases detected and notified directly through the two ACF interventions, not total number of TB cases notified from people living in intervention areas. m. Different studies used different methods of ACF n. We are aware of a body of unpublished literature around ACF interventions. o. Differences in effect size and direction of effect p. In general, no measures of uncertainty (confidence intervals) available. q. Risk of bias assessed using ROBINS-i (slightly modified), 3 studies at moderate ROB, 5 at serious ROB and 5 at critical ROB. r. 65% of children who had negative TST (0mm induration) in 2005 were identified in 2009 for repeat TST s. ACT3 presents two comparisons of IGRA positivity in children. Children born in 2012 (originally secondary outcome) had non-statistically significant more IGRA positives in ACF areas (p=0.42) and children born 2004 - 2011 (post hoc outcome) had statistically significantly fewer IGRA positives. Downgraded by one for inconsistency.	

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	Screening aims to identify people with active TB earlier and therefore ensure earlier treatment and better health outcomes for individuals and lower TB transmission to community. No direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is important uncertainty in what we know about how patients perceive and value the outcomes explored above.	In the absence of research data, the GDG provided their own views that given the likelihood of detecting more TB cases and ultimately reducing TB deaths and overall TB burden, there is probably no important uncertainty or variability about how much the community would value screening.

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	<u>Effectiveness</u> The panel determined that the balance between desirable and undesirable effects **favors** systematic screening for active TB in the general population. <u>Testing strategies</u> The panel determined that the balance between desirable and undesirable effects **favors** systematic screening with symptoms (any TB symptoms, any cough, cough > 2 weeks) for active TB among in the general population. The panel determined that the balance between desirable and undesirable effects **favors** systematic screening with radiography (chest X-ray any abnormality) for active TB in the general population.	*For the intervention:* The group felt that on the whole the balance of effects **probably favors the intervention**, but noted that the balance depends on many factors:

● Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know	The panel determined that the balance between desirable and undesirable effects **favors** systematic screening with mWRDs for active TB in the general population.	- TB prevalence, particularly undetected TB, in the setting; - The accuracy of the screening and the diagnostic confirmatory tests being used; - The current performance and level of case detection being provided by the existing health system/TB programme; - The vulnerability of population being screened - for populations at higher risk the balance of effects would lean more strongly toward "favors" Several GDG members noted that their judgement of this was "varies" but a majority noted that it "probably favors". *For the screening tests:* ● CXR (any abnormality): **favours the intervention** ● mWRDs: **probably favours the intervention** ● Symptoms - any TB symptom: **probably favours the intervention** ● Symptoms - prolonged cough: **probably favours the intervention** The balance of effects for symptom screening, and the best choice of symptoms to use, depends on the prevalence of TB symptoms and the local understanding of symptoms.

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	***Costing systematic review about health system resources required:*** Five studies directly addressed the question of the resources required for systematic screening for TB in the general population community setting. Evidence was available from studies conducted in Sub-Saharan Africa, China, India, Pakistan and Tajikistan. Average screening costs per person ranged from: US$0.10-$164 per person using a range of different screening tools, algorithms and approaches (Kranzer 2012, Mupere 2013, Sekandi 2014, Andre 2018, Machekera 2019). Mupere (2013) reported the total program costs (US$1,535,000) for an ACF program using door-to-door screening in urban Uganda, screening 536,844 persons over a 5-year time frame. Sekandi (2015) reported total program costs (US$41,160) of door-to-door ACF in urban setting in Uganda screening 1000 persons. ***Systematic review on patient costs of TB evaluation and treatment and risk of catastrophic costs:*** A review was conducted to investigate costs and occurrence of catastrophic costs for patients, families, and communities associated with TB investigation and treatment among people with active TB diagnosed through TB screening compared to standard case detection. The review found 6 studies published since 2015, all from low or middle-income countries that measured costs and prevalence of catastrophic costs borne to TB patients who were diagnosed through active TB case finding (ACF) compared to that of those who were diagnosed through standard TB diagnosis practices. Pooled analysis was not possible but all studies found evidence of lower total costs associated with TB evaluation and treatment for patients diagnosed through screening compared to standard case detection. Some studies also found evidence of lower prevalence of catastrophic costs among patients diagnosed through screening compared to standard case detection. • Shewade et al. (2018) found that, when compared with patients detected through PCF, ACF patients incurred lower total median costs (US$ 4.6 and 20.4 for ACF and PCF respectively, $p < 0.001$). When adjusted for the catastrophic costs prevalence, the analysis showed that patients detected through ACF had a 32% lower prevalence of catastrophic costs compared to patients detected by PCF. • Sekandi et al (2015) performed a static decision modelling framework to examine the costs and effectiveness of three TB case detection strategies (PCF alone, PCF+ACF, and PCF plus household contact investigation (HCI)). The study demonstrated that when compared to PCF alone: US$28.88 (US$14.44-US$43.32), ACF patients incurred lower costs: US$4.76 (US$2.38-US$7.14). • Hussain et al (2019) reported patient's costs at 3 stages of the illness: pre-diagnosis, diagnosis and treatment. The study demonstrates that ACF patients incurred lower costs (US$59) compared to patients found through standard case detection (US$71). • Munyandi et al (2019) comparing patients diagnosed through ACF and standard case detection in rural India found that patients who were diagnosed through ACF incurred lower costs at all stages of the illness (diagnosis (US$30 vs US$130; p-value=0.001), treatment (US$39 vs US$97; p-value=0.004) and entire illness (US$69 vs US$227; p-value=0.001); 9% of patients detected through ACF experienced catastrophic costs due to TB compared to 29% of patients diagnosed through standard case detection. • Gurung et al (2019) in Nepal found lower direct (medical and non-medical) costs but no significant difference in the indirect cost incurred by ACF patients compared to standard case detection (US$252.80 vs US$315.3; P-value=0.161). The study showed that patients detected through ACF had lower (though not significantly) catastrophic costs (45% vs 61%; p-value=0.143). • Morishita et al (2016) found that ACF patients incurred lower pre-treatment costs (US$5.10 vs US$22.4; p-value<0.001) though total costs did not differ significantly (US$233.2 vs US$235, respectively; p=0.367); catastrophic cost prevalence was lower among ACF patients (36.1% vs 45.0%, p-value = 0.244), though not significant.	The GDG noted the savings to patients costs, including a reduction in catastrophic costs. Overall the GDG noted that costs were generally large and varied from large to very large, according to specific interventions. The GDG also noted that judgments on costs depends on the perspective taken - that of the funder or that of the population being screened. The GDG emphasized that reduced patient costs, in particular a reduced risk of catastrophic costs, should be considered a critical outcome.

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	Health system review: There were five studies contributing to the evidence for this question, results varied across studies due to the differences in methodological approaches and screening programmes limiting comparability. Of the five studies that were included, three were high quality and two were downgraded because they did not identify and measure alternative outcomes clearly. Patient consequences review: Study quality varied and may impact the certainty of the evidence. Evidence on economic consequences of ACF and ICF programmes is still limited, but the reviewed studies demonstrated that ACF can, arguably, reduce the economic burden of TB to patients and their families. The available data suggest that ACF strategies may provide better financial risk protection for TB patients and their households than PCF. ACF may also limit catastrophic costs for patients and their families compared to PCF.	

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 intervention ○ Varies ● No included studies	Six studies (6/12) reported cost-effectiveness of screening in the general population and the average cost per person diagnosed with TB among these studies ranged from US$44-$28,825 in the general population (Daftary 2019, Bogdanova 2019, Hinderaker 2011, Machekera 2019, Jo 2020, Kranzer 2012). Community-based programs that primarily included personnel-related costs (i.e. healthcare workers or community volunteers to conduct screening) had lower costs on average per person diagnosed, ranging from US$44-$565 (Hinderaker 2011, Andre 2018, Daftary 2019, Machekera 2019) compared to mass CXR screening in the hospital setting or with mobile clinic units which ranged from US$343-$12,549 per person diagnosed with TB (Jo 2020, Kranzer 2012, Bogdanova 2019). Mobile clinic screening units were drivers of increased cost due to additional transportation and staffing needs and higher consumable costs attributable to the program. ACF interventions were determined to be cost-effective compared to standard case detection (i.e. PCF) in the general population in five studies (5/12) (Sekandi 2015, Mupere 2013, Hussain 2019, Kranzer 2012, Azman 2014) with ICERs ranging from US$43-$9,400 per DALY averted, or US$109 per QALY. One study (Nishikiori 2013) found ACF using CXR as an initial screening tool was not cost-effective in the general population with an ICER ranging from US$67-$495 per additional TB case detected.	The GDG noted that the studies at hand are mostly modelled data with heterogeneous evidence. Also there is undoubtedly a gap in cost-effectiveness data for the specific recommendation that this EtD will result in. They thus voted for the judgement 'no included studies'. The GDG noted that the time horizon of most available studies was too short to adequately capture the full "effectiveness" potential from screening, and studies to demonstrate cost effectiveness of screening would require a much longer time horizon to show the potential benefits or savings. Costs for screening are often the most at the start of a screening programme, while cost savings only materialize in the longer term (when TB incidence decreases). Under these circumstances cost-effectiveness analysis with short time horizons will always favor standard case detection. It is thus important to conduct studies with longer time horizons. The data available on cost-effectiveness mostly

		report on individual-level outcomes and do not incorporate community-level outcomes such as reduced transmission, future reduction in prevalence, etc. Also, the "effectiveness" measure (eg the willingness-to-pay threshold) is set specific to each study and may not be applicable across other settings.

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	***Stigma, social and mental health consequences of screening*** A systematic review of qualitative evidence was conducted to explore and understand stigma, social and mental consequences of systematic screening. **Summary of qualitative findings for stigma, social amd mental consequences** **CERQual Summary of Qualitative Findings Table** AIM: Patient perspective social and mental health consequences as a result of TB screening and/or passive case finding **Summary of Review Findings** / **Studies** / **CERQual confidence in the evidence** / **Explanation of CERQual** The contact at the household level away from the health facility setting that accompanies ACF and some forms of ECF had advantages. These included: reducing opportunity costs of extended treatment seeking; some aspects of mental health (feeling cared for, acceptance of TB patients); and enhanced knowledge. — Sathar 2020, Nordstoga 2019, Lorent 2015 — Moderate confidence — Some Limitations in adequacy, relevance, and coherence. ECF still required accessing the health facility and in more authoritarian, hegemonic and lower income contexts, particularly for migrants and women, this could create anxiety and other challenges (for example, financial, language), more especially if TB was confirmed. — Nordstoga 2019, Kumwenda,2017 — Low Confidence — Limitations in coherence, adequacy, relevance, and methodology. ECF = Enhanced case finding ACF = active case finding Detailed summarized qualitative synthesis statements: • Active case finding reduces barriers associated with passive case finding such as extended waiting times at clinics and costs associated with treatment seeking. (Moderate confidence) • Active case finding led to increased knowledge about TB. (Very low confidence) • Active case finding was not perceived to increase stigma. While enhanced case finding in the form of screening of migrants had mixed reviews in relation to stigma. (Very low confidence) • Increased anxiety as a result of enhanced case finding due to fear of consequences of a positive test results including (stigma, deportation for migrants, or punishment). • Anxiety also resulted from authoritative perception of health system. (Low confidence) • Enhanced case finding strategies that required multiple trips to the health facility led to conflict in relationships for women who had to seek consent from their spouses. Husbands were reported as being aggressive, angry, and suspicious of infidelity as the freedom to go to the facility on consecutive days, is perceived as having freedom to be with other men. (Very low confidence) • Feelings of dependence and vulnerability resulted from active or enhanced case finding strategies that required assistance from a third person (financial assistance or assistance navigating systems such as language barriers). (Low confidence) • Positive mental health implications of systematic screening included feeling cared for by those carrying out the case finding activities and increased understanding and acceptability of people living with TB. (Low confidence)	The GDG notes that community screening, if implemented well, would likely increase equity.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	A systematic review of qualitative studies was conducted to understand individual patient and community responses to active case finding and systematic TB screening. Key themes from 35 qualitative studies are summarised in the following statements: **Statement 1:** *People had varied preference regarding the place of screening: clinic, community or household visits, but all valued discretion, privacy and confidentiality whatever the location because they were fearful of the stigma associated with TB and HIV.* **Statement 2:** *Community ACF programmes improved the community's access to screening but led them to expect diagnosis and treatment delivered in the same way, which it often wasn't.* **Statement 3:** *Respondents across many studies reported the underlying TB service to be at best inconvenient and at worst chaotic. People were also afraid of TB medications.* **Statement 4:** *Resistance to ACF and TB contact tracing is relatively common. Communities and individuals are unclear about TB programme motives, but the story is not a straightforward one about trust alone. Respondents were especially likely to refuse ACF or contact tracing if they felt healthy or believed their symptoms were not serious. People weigh up their own priorities when deciding whether to engage with TB services.* **Statement 5:** *Uncertainty due to poor diagnostic test accuracy is well known in communities and devalues TB services in their eyes.*	The GDG notes that screening should always be voluntary - neither screening nor treatment should be mandatory.

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ○ Yes ● Varies ○ Don't know	Screening at scale may not be feasible in all settings. A survey of screening guideline users was conducted by WHO in 2019. Among 118 respondents (80% from NTP): • 36% of respondents reported the existing recommendation of screening in defined areas/populations with extremely high levels of undetected TB has been integrated into national guidelines (with or without adaptation), • 20% of respondents reported that the existing recommendation is routinely conducted. Screening for symptoms of TB is feasibe in any setting. Screening with chest X-ray requires access to radiography, in a health facility or a mobile screening unit. Basic radiography is an essential technology in primary care, and in recent years mobile digital basic X-ray systems have been developed that can provide radiology services outside the clinic setting, including in mobile vans. However, access to high-quality radiography is limited in many settings, including access to trained personnel to read images. mWRDs are recommended by WHO as the primary mode of TB diagnosis. mWRDs can be implemented at most levels of health care but resource and availability may prohibit use for screening at scale.	The GDG highlighted that the feasibility of implementing population-wide screening varies (the GDG notes all the comments about the factors that affect the intervention above). The feasibility of implementing specific screening tests differs: • Symptoms - Yes. • CXR - Probably yes, dependent on costs. • mWRDs - Varies - the group notes that it depends on the ability to produce sputum, costs, access to platforms. It was noted that it can be more feasible to collect sputum for mWRDs than to bring people to a center for CXR. Again dependent on costs.

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
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

Overall recommendation

The GDG suggests conducting systematic screening for TB disease in the general population in areas with an estimated prevalence of undetected TB of 0.5% or higher **(conditional recommendation, low certainty of the evidence)** .

Test specific recommendations

Among the general population, the GDG suggests using chest radiography (any abnormality, abnormality suggestive if TB), symptom screening (any cough, prolonged cough, any TB symptom) or molecular WHO-approved rapid diagnostic tests rather than no screening tests for detection of TB disease **(conditional recommendation, very low certainty of the evidence for test accuracy).**

Justification

Systematic screening in the general population is conducted with the joint primary objectives of benefiting individuals with TB by reducing diagnostic delay, improving treatment outcomes, and reducing costs and financial loss associated with the course of disease, as well as reducing the population prevalence of TB, thereby reducing further transmission of the mycobacterium and averting future incidence of TB disease.

Evidence from one randomized trial in Vietnam, with an estimated prevalence of 350/100,000 population, reported that systematic screening using three years of annual door to door sputum collection and testing using Xpert MTB/Rif reduced adult TB disease prevalence by more than 40% (prevalence ratio 0.56 (95% CI: 0.40 – 0.78), moderate certainty of evidence), and reduced incidence of TB infection in children aged 3-10 years by 50% (prevalence ratio 0.50 (95% CI: 0.32 – 0.78), moderate certainty). There is currently no evidence that population-wide screening using less accurate screening algorithms that begin with symptom screening are effective at reducing population prevalence or transmission of TB. There is also no direct evidence that screening improves treatment outcomes or reduces mortality in screened individuals. However, there is limited evidence that screening, including symptom-based screening interventions, may benefit individuals who are diagnosed with TB through such interventions, including earlier diagnosis with less severe TB, lower costs to the patient, and lower risk of catastrophic costs associated with the course of disease.

Subgroup considerations

- Systematic screening for TB within subgroups of the population with increased risk of TB or increased risk of poor outcomes from TB (contacts of TB patients, people living with HIV, miners, prisoners, those with clinical or structural risk factors for TB) are covered in separate recommendations.

Implementation considerations

- Screening should be conducted using the most sensitive and specific screening algorithm possible, that includes a screening test that identifies those with a higher likelihood of having TB, and a separate diagnostic test to confirm the diagnosis. CXR is the most sensitive screening test recommended, followed by MWRDs, and then symptom screening. The feasibility of more sensitive screening tests is affected by symptoms and implementation requirements; while symptom screening is much more feasible, the sensitivity of symptom screening is reduced compared to other screening tests (CXR, mWRDs).
- mWRD tests perform differently when used for screening than when used for diagnosis, especially regarding sensitivity and positive predictive value, and the results should be interpreted appropriately. People who screen positive for TB using an MWRD should receive a thorough clinical evaluation to establish a diagnosis of TB, including further tests and procedures such as CXR, additional molecular rapid diagnostic tests, patient history, etc.
- Current evidence suggests that population-wide screening needs to be conducted repeatedly in order to be effective at reducing population TB prevalence, though there is limited data on what frequency is ideal or sufficient. Evidence of reduction of prevalence and incidence of TB was achieved using annual screening over a period of 4 years.
- Implementation of population-wide screening requires a large investment of resources. Conducting population-wide screening using a highly accurate screening and diagnostic algorithm will inevitably require more resources than screening using less accurate screening tools but has more potential to reduce the population burden of TB. Symptom screening costs considerably less than screening with CXR or mWRD but has not been shown to reduce prevalence or transmission. The cost-effectiveness of population-wide screening with highly accurate algorithm is currently unknown. The higher the population prevalence of TB, the more cost-effective the intervention will be.

Monitoring and evaluation

- While is important to know the TB epidemiology in a community to be able to effectively evaluate the impact of any screening programme, without conducting a prevalence survey, it is very difficult to estimate the prevalence of undetected TB in a specific population is prior to conducting screening. Surveillance data reporting TB notifications can be used as a starting point, but must be carefully considered alongside other factors that affect the risk of TB and the likelihood of getting diagnosed, including the prevalence of other risk factors for TB (such as HIV, diabetes mellitus, other lung conditions such as silicosis or fibrotic chest lesions, smoking, malnutrition, poverty, overcrowded living conditions, older age, substance abuse disorder), and the availability of health services (see operational guidance for further discussion).
- Further suggestions for monitoring and evaluation of TB screening interventions are provided in the operational guidance.

Research priorities

- Well-designed trials in different settings (including different HIV prevalences, rural, urban, and suburban settings) are required to investigate the effectiveness of population-wide systematic screening for TB on population-level outcomes (TB prevalence, incidence, transmission) and individual-level outcomes (diagnostic delay, treatment outcomes, patient costs, social consequences, acceptability and equity), as well as to guide implementation choices including method of delivery, screening algorithms, duration and frequency of screening, and the mode of delivering the intervention.
- Research on the longer term impacts of screening, including any potential evidence of morbidity or mortality averted, is not necessarily captured in the existing data. Research on cost-effectiveness of screening using a longer time horizon to adequately capture all potential costs and longer-term effects, including potential reduced future prevalence and incidence, is needed.
- More cost-effectiveness research on population-wide screening that accounts for population and individual-level effects of screening, that factors in reduced transmission and averted TB cases, is needed across different prevalence settings.
- Observational research and programmatic evaluations reporting the impact of screening on TB case notification rates, which provide an important source of evidence of the impacts of screening under programmatic conditions, must be carefully designed to minimize bias.
- Screening intervention studies should incorporate both qualitative and quantitative assessment of the indirect effects of screening, given the importance of health-seeking behavior in TB care and the potential impact of population-wide screening to change it, and the importance of assessing for any unintended mental, social, or economic consequences of screening.

Table 2. Should systematic screening for TB disease, compared to passive case detection, be conducted among household and close contacts of individuals with TB?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	The WHO's End TB Strategy envisions a 90% reduction in tuberculosis (TB) incidence and 95% reduction in TB deaths by 2035, and the Resolution adopted by the United Nations General Assembly in September 2018 commits to diagnosing and treating 40 million people with TB and treating at least 30 million people for TB infection by 2022. These targets will not be met by current practices in case detection. In order to achieve these ambitious targets there is an urgent need to deploy strategies to improve detection of people with active TB. Systematic evaluation of people who have been exposed to potentially infectious cases of TB, or TB contacts, can be an efficient, targeted approach to intensified TB case finding. Systematic screening has been strongly recommended in contacts of pulmonary TB patients given their high risk of TB, albeit with very low quality of evidence, as they have a high prevalence of active TB - previous systematic reviews of published studies show that a pooled average of 3.5–5.5% (the equivalent of a prevalence of 3500–5500 per 100 000 population) of household members or other close contact with a person who has infectious TB are themselves found to have previously undiagnosed active TB. TB preventive treatment is recommended for TB contacts who are screened and found to not have indications of active disease. For the purpose of this guideline: A household contact is a person who shared the same enclosed living space for one or more nights or for frequent or extended periods during the day with a person diagnosed with pulmonary TB during the 3 months before commencement of the current treatment episode (definitions of 'household' vary considerably and must be adapted to the local context). A close contact is a person who is not in the household but shared an enclosed space, such as a social gathering place, workplace or facility, for extended periods during the day with a person diagnosed with pulmonary TB during the 3 months before commencement of the current treatment episode. For the remainder of this table, both household and close contacts of a person with TB will be referred to as "contacts". To inform the guideline update, a systematic review was conducted to determine the prevalence of active TB among TB contacts. The effects of systematic screening on mortality, prevalence, case notification and case detection were also reported. The mean pooled prevalence of TB among contacts was found to be: • Among contacts of bacteriologically-confirmed index patients: 3.4% (2.9-3.8) • Among contacts of MDR/XDR index patients: 3.7% (2.4-5.3) • Among contacts of all index cases: 3.5% (3.1-3.8) • Among contacts who are bacteriologically-confirmed secondary cases: 2.9% • Among contacts <5 years old: 3.8% (2.6-5.3) / among contacts 5-14 years old: 2.5% (1.7-3.5) / among contacts >14 years old: 4.9% (3.6-6.3) • Among contacts who are PLHIV: 11.7% (7.0-17.2) For the current revision of the guidelines for screening for active TB, the question to the Guideline Development Group (GDG) is whether this recommendation should be updated, in light of current evidence, and whether the group would like to make specific recommendations for screening tests and algorithms to be used.	The GDG agreed that this is a priority problem and noted that household contacts are among the risk groups with the highest level of exposure. For members of this risk group, the identification of early subclinical TB is important.

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with standard case detection	Risk difference with systematic screening
Co-prevalent TB cases detected among contacts of any bacteriologically-	615200 (107	⊕◯◯◯ VERY LOW[a,b,c,d]	-	Contacts with TB = 10,417	

confirmed index patients assessed with: Case detection	observational studies)			Contacts screened = 615,200 Weighted pooled prevalence = 3.4% (2.9-3.8%) Median NNS = 31 (18-65) (n=101)
Co-prevalent TB cases detected among contacts of MDR/XDR index patients assessed with: Case detection	273974 (19 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	Contacts with TB = 4,850 Contacts screened = 273,974 Weighted pooled prevalence = 3.7% (2.4-5.3%) Median NNS = 27 (13-50) (n=18)
Co-prevalent TB cases detected among contacts (All TB cases) assessed with: Case detection	1324081 (187 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	Contacts with TB = 19,374 Contacts screened = 1,311,666 Weighted pooled prevalence = 3.5% (3.1-3.8%) Median NNS = 35 (17-65) (n=181)
Co-prevalent TB cases detected among contacts (Micro-confirmed cases) assessed with: Case detection	988902 (112 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	Contacts with TB = 15,170 Contacts screened = 988,902 Weighted pooled prevalence = 2.9% (2.4-3.4%) Median NNS = 42 (20-90) (n=105)
Co-prevalent TB cases detected among contacts (<5 years old) assessed with: Case detection	48911 (29 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	Contacts with TB = 803 Contacts screened = 48,911 Weighted pooled prevalence = 3.8% (2.6-5.3%) Median NNS = 30 (12-62)
Co-prevalent TB cases detected among contacts (5-14 years old) assessed with: Case detection	14622 (19 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	Contacts with TB = 283 Contacts screened = 14,622 Weighted pooled prevalence = 2.5% (1.7-3.5%) Median NNS = 36 (17-61) (n=16)
Co-prevalent TB cases detected among contacts (>14 years old) assessed with: Case detection	230934 (26 observational studies)	⊕◯◯◯ VERY LOW[a,b,c,d]	-	Contacts with TB = 3,919 Contacts screened = 230,934 Weighted pooled prevalence = 4.9% (3.6-6.3%) Median NNS = 28 (16-52) (n=24)
Co-prevalent TB cases detected among HIV infected	1696 (5	⊕◯◯◯ VERY LOW[a,b,d,e]	-	Contacts with TB = 149 Contacts screened = 1,696

contacts assessed with: Case detection	observational studies)			Weighted pooled prevalence = 11.7% (7.0-17.2%) Median NNS = 24 (17-28)

a. No significant concerns regarding indirectness were identified.
b. Substantial unexplained inconsistency was identified, owing to a range of causes of heterogeneity (including variations in screening and testing strategies, timing of screening, intensity of exposure to an index case, the rate of community transmission, HIV prevalence and other factors led to significant heterogeneity). For this reason, quality was downgraded two levels for inconsistency.
c. Imprecision was not a major concern, given the large number of participants in most included studies.
d. Almost all studies lacked a control group in which screening was not performed. Therefore, these reported estimates are likely to overestimate the benefit of screening, assuming that all case detection is due to the intervention when some cases are likely to have been detected through passive case-finding. Therefore, the evidence quality was rated down two levels for risk of bias.
e. Imprecision downgraded by one on the basis of a small number of overall participants evaluated.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	***Effectiveness review*** A systematic review was conducted. The certainty of evidence was downgraded due to: 1) methodological limitations (difference in ascertainment of the outcome between arms); 2) indirectness (due to collection of an indirect outcome and study conducted in a specific population that may not be generalizable); 3) imprecision (wide confidence intervals including appreciable harm and benefit). Overall, the certainty is moderate for the critical patient important outcome mortality and moderate for the community level outcome of TB prevalence, and case notification and low for case detection. <u>In summary:</u> • Systematic screening probably results in 7 fewer deaths per 1000 people compared to passive case-finding (moderate certainty evidence). • Systematic screening probably reduces the TB prevalence ratio by 18% (95% CI 0.64 - 1.04) resulting in 2 fewer cases/ 1000 people compared to people who were part of the passive case-finding (moderate certainty evidence). • Systematic screening probably results in 2.5 fold increase in case notification resulting in 11 more patients/ 1000 detected compared to passive case-finding (moderate certainty evidence). • Systematic screening may result in a slight increase in case detection (4 more patients/ 1000 detected) compared to passive case-finding (low certainty evidence).	The GDG noted that the initiation of TPT is an additional desirable effect that should be noted, as contacts are a major group for which TPT is recommended once TB disease is ruled out. The group noted here that the benefits expected to community level measures are less, as the population targeted is relatively small and diffuse compared to the overall population, but the benefits expected to the individual are significant and of primary importance.

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with standard case detection	Risk difference with systematic screening
Death follow up: 2 years	25707 (1 RCT)[1]	⊕⊕⊕◯ MODERATE[a,b]	**RR 0.6** (0.5 to 0.8)	Study population	
				17 per 1,000	**7 fewer per 1,000** (8 fewer to 3 fewer)

TB prevalence ratio assessed with: culture confirmed TB among adults follow up: 4.5 years	89707 (1 RCT)[2]	⊕⊕⊕◯ MODERATE[c,d]	**RR 0.82** (0.64 to 1.04)	Study population	
				10 per 1,000	**2 fewer per 1,000** (4 fewer to 0 fewer)
Case notification assessed with: Cases registered with NTP follow up: 2 years	25707 (1 RCT)[1]	⊕⊕⊕◯ MODERATE[b]	**RR 2.5** (2.0 to 3.2)	Study population	
				7 per 1,000	**11 more per 1,000** (7 more to 15 more)
Case detection assessed with: Microbiologically confirmed	919 (1 RCT)[3]	⊕⊕◯◯ LOW[e,f]	**OR 1.34** (0.42 to 4.24)	Study population	
				11 per 1,000	**4 more per 1,000** (6 fewer to 35 more)

1. Fox et al. Household-contact investigation for detection of tuberculosis in Vietnam. NEJM, 2018.
2. Ayles et al. Effect of household and community interventions on the burden of tuberculosis in southern Africa: the ZAMSTAR community-randomised trial. Lancet, 2013.
3. Davis et al. Home-based tuberculosis contact investigation in Uganda: a household randomised trial. ERJ Open Res ; 2019.

a. Mortality was evaluated as part of a pot-hoc analysis in Fox 2018.
b. Downgraded by one level for serious indirectness. Fox 2018 was conducted in Vietnam in high TB prevalence population. Despite the large sample and inclusion of many sub-populations, this trial was conducted in one country setting and may not be generalizable to all other countries relevant for this recommendation.
c. Downgraded by one level for serious indirectness. Ayles 2013 was a community-randomized trial in Zambia and South Africa. The main outcome was TB prevalence after ~4 years of follow-up. It assessed the impact of active case finding on population level prevalence rather than effectiveness of contact investigation for diagnosing TB. The study setting was a high HIV prevalence context that may not reflect other settings.
d. Not downgraded by one level for imprecision. Despite the wide confidence interval which spans appreciable benefit and no effect, there were many events and a large sample informing this result.
e. Downgraded by one level for serious risk of bias. Unclear if TB testing was similar in both arms i.e. if household contacts in the standard care arm were referred for TB testing. Differences in ascertainment outcome may introduce bias.
f. Downgraded by one level for imprecision. The study primary outcome was completion of contact investigation cascade 14 days after initial household visit. There were few events and the study was not powered to address the outcome, this is shown in the wide confidence interval crossing both appreciable benefit and harm.

Indirect evidence from review of impact of active case-finding on patient costs of TB evaluation and treatment and risk of catastrophic costs

A review was conducted to investigate costs and occurrence of catastrophic costs for patients, families, and communities associated with TB investigation and treatment among people with active TB diagnosed through TB screening compared to standard case detection. The review found 6 studies published since 2015, all from low or middle-income countries that measured costs and prevalence of catastrophic costs borne to TB patients who were diagnosed through active TB case-finding (ACF) compared to that of those who were diagnosed through passive case-finding (PCF). Pooled analysis was not possible but all studies found evidence of lower total costs associated with TB evaluation and treatment for patients diagnosed through screening compared to PCF. Some studies also found evidence of lower prevalence of catastrophic costs among patients diagnosed through screening compared to standard case detection.

- Shewade et al. (2018) found that, when compared with patients detected through PCF, ACF patients incurred lower total median costs (US$ 4.6 and 20.4 for ACF and PCF respectively, $p < 0.001$). When

adjusted for the catastrophic costs prevalence, the analysis showed that patients detected through ACF had a 32% lower prevalence of catastrophic costs compared to patients detected by PCF.

- Sekandi et al (2015) performed a static decision modelling framework to examine the costs and effectiveness of three TB case detection strategies (PCF alone, PCF+ACF, and PCF plus household contact investigation (HCI)). The study demonstrated that when compared to PCF alone: US$28.88 (US$14.44-US$43.32), ACF patients incurred lower costs: US$4.76 (US$2.38-US$7.14).
- Hussain et al (2019) reported patient's costs at 3 stages of the illness: pre-diagnosis, diagnosis and treatment. The study demonstrates that ACF patients incurred lower costs (US$59) compared to patients found through standard case detection (US$71).
- Munyandi et al (2019) comparing patients diagnosed through ACF and standard case detection in rural India found that patients who were diagnosed through ACF incurred lower costs at all stages of the illness (diagnosis (US$30 vs US$130; p-value=0.001), treatment (US$39 vs US$97; p-value=0.004) and entire illness (US$69 vs US$227; p-value=0.001); 9% of patients detected through ACF experienced catastrophic costs due to TB compared to 29% of patients diagnosed through standard case detection.
- Gurung et al (2019) in Nepal found lower direct (medical and non-medical) costs but no significant difference in the indirect cost incurred by ACF patients compared to standard case detection (US$252.80 vs US$315.3; P-value=0.161). The study showed that patients detected through ACF had lower (though not significantly) catastrophic costs (45% vs 61%; p-value=0.143).
- Morishita et al (2016) found that ACF patients incurred lower pre-treatment costs (US$5.10 vs US$22.4; p-value<0.001) though total costs did not differ significantly (US$233.2 vs US$235, respectively; p=0.367); catastrophic cost prevalence was lower among ACF patients (36.1% vs 45.0%, p-value = 0.244), though not significant.

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know		The undesirable effects may depend on the testing strategy used. The concern about false-positive test results is less of an issue with this group given the high prevalence of TB. The group also notes possible risk of stigma related to the conduct of household contact investigation as a potential undesirable effect.

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		The GDG notes that the evidence for the effect of screening of contacts on reductions in mortality is strong (albeit a post-hoc analysis) but cannot be considered globally representative - it is from a single trial, albeit a large and robust trial. The group therefore suggests that the evidence should be graded as "moderate". It was noted that there probably won't be subsequent large-scale

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with standard case detection	Risk difference with systematic screening
Death follow up: 2 years	25707 (1 RCT)[1]	⊕⊕⊕◯ MODERATE[a,b]	**RR 0.6** (0.5 to 0.8)	Study population	
				17 per 1,000	**7 fewer per 1,000** (8 fewer to 3 fewer)

TB prevalence ratio assessed with: culture confirmed TB among adults follow up: 4.5 years	89707 (1 RCT)[2]	⊕⊕⊕◯ MODERATE[c,d]	**RR 0.82** (0.64 to 1.04)	Study population	
				10 per 1,000	**2 fewer per 1,000** (4 fewer to 0 fewer)
Case notification assessed with: Cases registered with NTP follow up: 2 years	25707 (1 RCT)[1]	⊕⊕⊕◯ MODERATE[b]	**RR 2.5** (2.0 to 3.2)	Study population	
				7 per 1,000	**11 more per 1,000** (7 more to 15 more)
Case detection assessed with: Microbiologically confirmed	919 (1 RCT)[3]	⊕⊕◯◯ LOW[e,f]	**OR 1.34** (0.42 to 4.24)	Study population	
				11 per 1,000	**4 more per 1,000** (6 fewer to 35 more)

trials of this sort conducted in similar settings, though there could still be equipoise in very different settings or exploring different screening algorithms or implementation strategies.

1. Fox et al. Household-contact investigation for detection of tuberculosis in Vietnam. NEJM; 2018.
2. Ayles et al. Effect of household and community interventions on the burden of tuberculosis in southern Africa: the ZAMSTAR community-randomised trial. Lancet; 2013.
3. Davis et al. Home-based tuberculosis contact investigation in Uganda: a household randomised trial. ERJ Open Res ; 2019.

a. Mortality was evaluated as part of a pot-hoc analysis in Fox 2018.
b. Downgraded by one level for serious indirectness. Fox 2018 was conducted in Vietnam in high TB prevalence population. Despite the large sample and inclusion of many sub-populations, this trial was conducted in one country setting and may not be generalizable to all other countries relevant for this recommendation.
c. Downgraded by one level for serious indirectness. Ayles 2013 was a community-randomized trial in Zambia and South Africa. The main outcome was TB prevalence after ~4 years of follow-up. It assessed the impact of active case finding on population level prevalence rather than effectiveness of contact investigation for diagnosing TB. The study setting was a high HIV prevalence context that may not reflect other settings.
d. Not downgraded by one level for imprecision. Despite the wide confidence interval which spans appreciable benefit and no effect, there were many events and a large sample informing this result.
e. Downgraded by one level for serious risk of bias. Unclear if TB testing was similar in both arms i.e. if household contacts in the standard care arm were referred for TB testing. Differences in ascertainment outcome may introduce bias.
f. Downgraded by one level for imprecision. The study primary outcome was completion of contact investigation cascade 14 days after initial household visit. There were few events and the study was not powered to address the outcome, this is shown in the wide confidence interval crossing both appreciable benefit and harm.

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	No direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is important uncertainty in what we know about how patients perceive and value the outcomes explored above.	The GDG agreed that there is probably little uncertainty about values placed on the main outcomes assessed in the reviews presented.

important uncertainty or variability ○ No important uncertainty or 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 panel determined that the balance between desirable and undesirable effects **favors** systematic screening for active TB among contacts of TB patients.	

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	Eleven studies spoke to costs and cost-effectiveness of screening among contacts. Seven studies (7/11) directly addressed the resources required for this sub-question with evidence from Cambodia, Myanmar, Russia, Democratic Republic of Congo (DRC), Peru, and Zimbabwe. Among these studies, the average screening costs per person ranged from: US$2.57-$41.00 per person using a range of different screening tools, algorithms and approaches (Shah 2017, Machekera 2019, Andre 2018, Htet 2017, Eang 2012, Bogdanova 2019). No studies were conducted in the clinical setting. All ACF of close contacts focused on some form of community-based screening (i.e. door-to-door, household visits or telephone calls) with higher costs among programs using mobile CXR or repeated visits to homes ranging from US$16-41 (Eang 2012, Shah 2017). Two studies conducted in Myanmar reported low per person costs for screening ranging from US$2.57-4.13 (Andre 2018, Htet 2017). Andre (2018) reported total program costs (US$188,578) for an ACF program using door-to-door symptom screening in the DRC, screening 73,418 over a 2-year period. Shah (2017) reported total program cost (US$81,904) of a community based ACF program with household visits to screen all HH) for TB with an initial symptom screen followed by Xpert testing in Peru, screening all HHC of 2,500 index cases over a 2-year period. James (2017) reported total program costs (US$363,257) for door-to-door screening of US$248,222 among 33,029 HHC and neighborhood contacts screened with CXR in Cambodia, screening 35,005 HHC.	The GDG noted that the costs required for contact investigation will vary considerably by the intensity of the intervention (including whether it is conducted at the patient's home or in the clinic, what screening and diagnostic algorithm is used, if repeated screening is conducted, etc.). The GDG also notes that screening represents a redistribution of costs from the patient and their family to the health system or the entity conducting the screening. However, overall the population of close contacts of TB patients is relatively smaller than other populations that can be targeted for screening, with a higher prevalence and thus a lower NNS, so relatively speaking the resources required to

		screen this population would be lower than many other risk groups under consideration.

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	There were eleven studies contributing to the evidence for this question, nine (82%, 9/11) of which included an ACF intervention which included household visits or door-to-door screening. However, there was heterogeneity regarding programmatic, screening and diagnostic costs included among the studies which limits comparability. Three of the 11 studies did not specify which programmatic costs were included. Ten of the studies were of good quality, but one was downgraded because they did not identify and measure alternative outcomes clearly.	

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 intervention ○ Varies ○ No included studies	Eight of the eleven (8/11) studies reported the cost-effectiveness of ACF among household contacts with an average cost per person diagnosed with TB, which ranged from US$44-$2,693 (Eang 2012, Htet 2017, James 2017, Andre 2018, Bogdanova 2019, Machekera 2019, Sekandi 2015, Lung 2019). Community-based programs that primarily included personnel-related costs (i.e. healthcare workers or community volunteers to conduct screening) had significantly lower costs per person diagnosed US$35-$44 (Htet 2017, Andre 2018) compared to CXR screening among contacts with mobile units UA$2,693 (Bogdanova 2019). ACF interventions among household contacts were determined to be highly cost-effective compared to standard case detection (i.e. PCF) in four studies (4/11) with ICERs ranging from US$330-$3,347 per DALY averted (Yadav 2014, Sekandi 2015, Shah 2017, Lung 2019).	The GDG notes that, given the relatively high prevalence of TB in this group, screening is more cost-effective in this risk group compared to other groups which may be targeted for screening.

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	***Stigma, social and mental health consequences of screening*** A systematic review of qualitative evidence was conducted to explore and understand stigma, social and mental consequences of systematic screening in general (not specific for household contacts). . **<u>Summary of qualitative findings for stigma, social and mental consequences</u>**	The GDG notes that screening conducted among this high-risk group will most likely increase equity.

CERQual Summary of Qualitative Findings Table

AIM: Patient perspective social and mental health consequences as a result of TB screening and/or passive case finding

Summary of Review Findings	Studies	CERQual confidence in the evidence	Explanation of CERQual
The contact at the household level away from the health facility setting that accompanies ACF and some forms of ECF had advantages. These included: reducing opportunity costs of extended treatment seeking; some aspects of mental health (feeling cared for, acceptance of TB patients); and enhanced knowledge.	Sathar 2020, Nordstoga 2019, Lorent 2015	Moderate confidence	Some Limitations in adequacy, relevance, and coherence.
ECF still required accessing the health facility and in more authoritarian, hegemonic and lower income contexts, particularly for migrants and women, this could create anxiety and other challenges (for example, financial, language), more especially if TB was confirmed.	Nordstoga 2019, Kumwenda,2017	Low Confidence	Limitations in coherence, adequacy, relevance, and methodology.

Detailed summarized qualitative synthesis statements:

- Active case finding reduces barriers associated with passive case finding such as extended waiting times at clinics and costs associated with treatment seeking. **(Moderate confidence**)
- Active case finding led to increased knowledge about TB. (**Very low confidence**)
- Active case finding was not perceived to increase stigma. While enhanced case finding in form of screening of migrants had mixed reviews in relation to stigma. (**Very low confidence**)
- Increased anxiety as a result of enhanced case finding due to fear of consequences of a positive test results including (stigma, deportation-for migrants-, or punishment). Increased Anxiety also resulted from authoritative perception of health system. (**Low confidence**)
- Enhanced case finding strategies that required multiple trips to the health facility led to conflict in relationships for women who had to seek consent from their spouses. Husbands were reported as being aggressive, angry, and suspicious of infidelity as the freedom to go to the facility on consecutive days, is perceived as having freedom to be with other men. (**Very low confidence**)
- Feelings of dependence and vulnerability resulted from active or enhanced case finding strategies that required assistance from a third person (financial assistance or assistance navigating systems such as language barriers). (**Low confidence**)
- Positive mental health implications of ECF/ACF included feeling cared for by those carrying out the case finding activities and increased understanding and acceptability of people living with TB. (**Low confidence**)

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	A systematic review of qualitative studies was conducted to understand individual patient and community responses to active case finding and systematic TB screening. Key themes from 35 qualitative studies are summarized in the following statements: **Statement 1:** *People had varied preference regarding the place of screening: clinic, community or household visits, but all valued discretion, privacy and confidentiality whatever the location because they were fearful of the stigma associated with TB and HIV.* **Statement 2:** *Community ACF programmes improved the community's access to screening but led them to expect diagnosis and treatment delivered in the same way, which it often wasn't.* **Statement 3:** *Respondents across many studies reported the underlying TB service to be at best inconvenient and at worst chaotic. People were also afraid of TB medications.* **Statement 4:** *Resistance to ACF and TB contact tracing is relatively common. Communities and individuals are unclear about TB programme motives, but the story is not a straightforward one about trust alone. Respondents were especially likely to refuse ACF or contact tracing if they felt healthy or believed their symptoms were not serious. People weigh up their own priorities when deciding whether to engage with TB services.* **Statement 5:** *Uncertainty due to poor diagnostic test accuracy is well known in communities and devalues TB services in their eyes.*	The GDG notes that other research has shown stigma to be an important factor in health-seeking behavior among high risk populations.

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Contact investigation is a long standing practice in TB control and, albeit incompletely carried out, data collected by WHO for its Global TB Report indicates that close to 10 million contacts eligible for screening were identified by 109 countries in 2019, of whom >5 million were reported as having been investigated for both TB disease and TB infection. Interventions conducted via home visit are more labor-intensive but may prove more effective at finding all close contacts with TB and can help reduce barriers to accessing care. A survey of screening guideline users was conducted by WHO in 2019. Among 118 respondents (80% from NTP): • 83% of respondents report the existing recommendation of screening of contacts of TB patients has been integrated into national guidelines (with or without adaptation), • 56% of respondents report that the existing recommendation for screening of contacts is routinely conducted. Feasibility of screening by different approaches: • Screening with symptoms is very feasible • Screening with CXR requires access to radiography, in a health facility or a mobile screening unit. Basic radiography is an essential technology in primary care, and in recent years mobile digital basic X-ray systems have been developed that can provide radiology services outside the clinic setting, including in mobile vans. However, access to high-quality radiography is limited in many settings, including access to trained personnel to read images • mWRDs are recommended by WHO as the primary mode of TB diagnosis. mWRDs are able to be implemented at most levels of health care but resource and availability may prohibit use for screening at scale	The GDG notes that in general this intervention is feasible to implement. Feasibility varies depending on the screening and diagnostic algorithm used and the site of screening (in the community/household or at the clinic). Symptom screening is feasible in almost all settings. CXR screening is less feasible, particularly in the household setting. Screening with mWRDs can be feasible using remote sputum collection at the household and sample transport to the location of the testing unit, given resources are available for the test.

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	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

	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

Among household and other close contacts of TB patients, the GDG **recommends** systematic screening for TB **should** be conducted **(strong recommendation, moderate certainty of evidence).**

Related recommendation(s)

Test specific recommendations

Among the general population, the GDG suggests using chest radiography (any abnormality, abnormality suggestive if TB), symptom screening (any cough, prolonged cough, any TB symptom) or molecular WHO-approved rapid diagnostic tests rather than no screening tests for detection of TB disease **(conditional recommendation, very low certainty of the evidence for test accuracy).**

Justification

- Household and close contacts of people with TB are at high risk of TB infection and developing TB disease. A systematic prevalence review conducted for this guideline meeting found the weighted pooled prevalence of TB disease among all close contacts of TB patients was 3.6% (95% CI 3.3-4.0), with a median NNS of 35 (95% CI 17-65).
- Systematic screening has been strongly recommended in contacts of pulmonary TB patients since 2012, given the high prevalence of disease. For this guideline meeting, there was trial evidence on individual- and community-level effects of screening of contacts of TB patients (contact investigation).
- The GDG maintained a strong recommendation for screening of contacts for TB. This was based on the reduction in death and increase in case detection and longer term prevalence of TB. Certainty of evidence has increased to moderate certainty based on new clinical trial results. The GDG also considered that screening of contacts may increase equity related to better access to healthcare and knowledge regarding TB.
- Screening of contacts was considered to be both acceptable and feasible, although it should be noted that feasibility may be impacted by the choice of screening algorithm and intensity of screening.

Subgroup considerations

The recommendations here apply to both household contacts and close contacts. Contact investigation should always be done when the TB patient has any of the following characteristics: bacteriologically confirmed pulmonary TB, proven or suspected MDR-TB or extensively drug-resistant TB, is a person living with HIV, or is a child younger than 5 years.

Implementation considerations

- Screening of contacts can take place at the household of the TB patient or in the clinic setting. The GDG emphasized that screening should be conducted in the household if at all possible, to bring the intervention to the population at risk and ensure a high proportion of contacts are reached for screening.
- The GDG also emphasized that, given the high risk of disease in this group, the most accurate screening and diagnostic algorithm possible should be used.

Monitoring and evaluation

- Continued monitoring is helpful for the programme managers to assess the performance of TB screening interventions.
- Standard monitoring and evaluation procedures may be complemented by operational research aimed at improving the performance of screening in the local setting as well as research aimed at improving the global evidence base on screening.

Research priorities

- Further research is needed to guide the implementation of screening of contacts, including evaluating approaches to maximize the reach and effectiveness of screening interventions to reach all contacts, ensure contacts who screen positive reach health facilities for further diagnostic evaluation, and ensure initiation of preventive treatment for contacts in whom TB disease is ruled out.
- More research is needed to evaluate the accuracy and effectiveness of complete screening and diagnostic algorithms for screening of contacts, including symptom screening, CXR, and mWRDs in various combinations with diagnostic evaluation.

Table 3. Should systematic screening for TB disease, compared to passive case detection, be conducted in prison settings?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	People in prisons (including those being held by the criminal justice system or those who work at such institutions) are at an increased risk for TB; the estimated incidence of TB among people residing in prisons has previously been estimated to be 23 times higher than the general population (Baussano et al 2010), with a median incidence of 1943/100,000 population in middle-income and low-income countries, and an incidence of 238/100,000 in high-income countries. In 2013 the WHO issued a conditional recommendation that systematic screening for active TB should be considered in prisons and other penitentiary institutions, based on very low quality of evidence. For the current revision of the guidelines for screening for active TB, the question to the Guideline Development Group (GDG) is whether this recommendation needs further updating, in line with recent evidence, and whether the group would like to make specific recommendations for screening tests and algorithms to be used. Observational data of screening interventions in prisons from the last ten years (2010-2020) show: • Of 40 eligible studies, 2 were in low burden settings, 10 were in moderate burden settings, 19 were in medium burden settings, and 9 were in high burden settings. • For 24 studies that used symptom-based screening, the mean NNS was 248 (range 5-775); for 7 studies in moderate burden settings the mean NNS was 115 (range 17-560); for 13 studies in medium burden settings the mean NNS was 302 (range 30-775); for 4 studies in high burden settings the mean NNS was 142 (range 5-188). • For 14 studies that included CXR in the primary screen, the mean NNS was 401 (range 7-923). For 5 studies conducted in low/moderate burden settings the mean NNS was 416 (range 7-923); for 5 studies conducted in medium burden settings the mean NNS was 102 (range 41-178); for 4 studies conducted in high burden settings the mean NNS was 437 (range 18-505). • For 3 studies conducted in moderate/medium burden settings that used Xpert as a screening strategy, the mean NNS was 27 (range 16-62). Of these, 1 was conducted among HIV-positive prisoners (NNS 16).	

Low (<30)				Moderate (30-100/100k)				Medium (100-300/100k)				High (>300/100k)			
Symptoms	CXR	Sx/CXR	MWRD	Symptoms	CXR	Sx/CXR	MWRD	Symptoms	CXR	Sx/CXR	MWRD	Symptoms	CXR	Sx/CXR	MWRD
	923 n=1	7 n=1 (HIV-infected inmates)		115 (98-445) n=7	244 (35-577) n=4		17 (16-17) n=2		147 (135-178) n=2	86 (41-98) n=2	62 n=1		304 (18-445) n=2	494 (35-505) n=2	

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	**Individual-level effects** *Time to diagnosis/disease severity at diagnosis* No randomized trials were identified that addressed this question. One observational study found that the proportion of patients diagnosed at an advanced stage of disease may be lower among those diagnosed through screening programmes (10%, CI 3-24%) compared to those diagnosed through standard case detection (51%, CI 37-65%) (**very low quality of evidence**). **Population-level effects** *TB prevalence* Two observational studies in prisons, including a symptom- and CXR-screening study in Brazil and a symptom-screening study in Ethiopia, found screening in prisons may reduce TB prevalence (**very low certainty of evidence**). *Case detection/case notification*	The GDG noted that bringing health screening programmes into prisons can bring many benefits beyond TB-related outcomes to prisoners, including improved living conditions and enhanced nutrition. Some GDG members felt the desirable effects could be considered large, but the group

One randomized trial in Ethiopia found that screening in prisons may increase case detection (defined as the number of notified cases divided by the estimated number of incident cases) 52.9% (95% CI 17.5-88.3%) (**low certainty evidence**).

Four observational studies in Zambia, India, Uganda, and USA found that screening may increase case detection (**very low certainty evidence**).

Knowledge, attitudes, practices

One randomized trial in Ethiopia found that screening may increase prisoner knowledge about TB (aOR 2.54, 1.93 - 3.94) (**low certainty evidence**).

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Impact
Earlier case detection: severity at diagnosis - smear positivity (smear positive among culture positive cases)	0 (1 observational study)	⊕◯◯◯ VERY LOW[a,b,c]	**ACF n/N (%; 95%CI) vs PCF n/N (%; 95%CI)** Paiao 2016: 4/40 (10%; 3-24%) vs 27/53 (51%; 37-65%)
TB disease prevalence (non-randomised studies) (Sanchez et al 2013 in Brazil and Tsegaye Sahle et al 2019 in Ethiopia)	0 (2 observational studies)	⊕◯◯◯ VERY LOW[d,e,f,g]	Sanchez 2013: TB prevalence before ACF was 8 cases / 1374 people (6040 per 100,000) and after ACF was 8 cases / 954 (2900 per 100,000). Tsegaye 2019: study prevalence before ACF was 3 cases / 3024 people (99 per 100,000) and after ACF was 10 cases / 2551 (392 per 100,000).
TB case notification rates (randomised studies) (Adane et al 2019)	0 (1 RCT)	⊕⊕◯◯ LOW[h]	Mean case detection rate, defined as "the number of new smear positive cases detected divided by the estimated number of incident smear positive cases, expressed as a percentage"; **mean difference in case detection rate +52.9 percentage points (95% CI 17.5-88.3)**. CNR ratio= 1.78 (no uncertainty estimate available).
Knowledge, attitudes and practices (Adane 2019)	0 (1 RCT)	⊕⊕◯◯ LOW[i]	Odds of having good composite knowledge score about TB increased in those who received ACF (aOR 2.54, 1.93 - 3.94). Odds of having survey-reported good practice similarly increased (aOR 1.84, 1.17 - 2.96). No statistically significant difference between groups in attitude scores (aOR 0.80, 0.52 - 1.25).

a. No adjustment for potential confounders (downgraded by 1 level for methodological limitations).
b. There is no gold standard for assessing severity, although increased smear positivity within a mostly non-immunosuppressed population could be suggestive of more severe disease. This is however not the case in immunosuppressed populations (rated down by 1 level for indirectness).
c. One small study, low event rates (rated down by 1 level for imprecision).
d. High risk of bias due to unaccounted for confounding by temporal trends (downgraded by 2 levels for very serious risk of bias).
e. Different direction of effect between two studies (downgraded by 1 level for serious inconsistency).
f. Different methods of ACF evaluated (downgraded by 1 level for serious indirectness).
g. Measures of uncertainty not available. Small numbers of events (downgraded by 2 levels for very serious imprecision).
h. Only measured in one study (downgraded by 2 levels for very serious imprecision).

agreed on the judgement of "moderate".

	i. Measured by survey rather than observation (downgraded by 2 levels for very serious indirectness).	

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ○ Small ● Trivial ○ Varies ○ Don't know		The GDG felt the undesirable effects of screening in prisons were trivial but potentially included stigma and isolation resulting from a diagnosis of TB. One possible undesirable effect, overtreatment resulting from false-positive screening leading to false-positive diagnosis, would have a possible benefit of treating infection as well.

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	Certainty of evidence for the effects of screening in prisons on time to case detection is very low. Certainty of evidence for the effects of screening in prisons on case notification and health knowledge is low; certainty of evidence for the effects of screening on TB prevalence is very low.	

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	Screening aims to identify people with active TB earlier and therefore ensure earlier treatment and better health outcomes for individuals and lower TB transmission, which is of high importance in prison settings. No direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is important uncertainty in what we know about how patients perceive and value the outcomes explored above.	

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 panel determined that the balance between desirable and undesirable effects **favors** systematic screening for TB disease in prison settings.	The GDG felt the balance of effects clearly favors the intervention.

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	Four studies directly addressed this question of resource requirements for systematic screening for TB among prisoners. Evidence was available from studies conducted in Belgium, Former Soviet Union (FSU), South Africa and Zimbabwe. The average screening costs per prisoner/inmate ranged from US$2.16-$35.00 (Machekera 2019, Smit 2017, Winetsky 2012, Zishiri 2014) and varied based on the screening and diagnostic algorithm used (i.e. annual symptom, CXR or SSM screening and/or Xpert for diagnosis). Smit (2017) reported total program costs (€113,325 (95% UI: €102,238-€125,148)) of screening 7,901 prisoners using mobile CXR units in Belgium. Zishiri (2014) reported total program costs (US$730,000) of an ACF program using CXR screening for all inmates, screening 20,7000 inmates at four correctional facilities in South Africa of over 12-months. Evidence from one study conducted among male prisoners (>18 years old) in Brazil was added, as it was published after the search date (da Silva Santos 2020). Da Silva Santos (2020) reported both fixed costs (i.e. purchasing, maintenance and software for CAD4TB (v5)) along with variable costs (i.e. human resources and evaluation of digital CXR) for screening among prisoners. The average cost per inmate screened for each of the components was as follows: WHO 4SS – US$1.90; clinical evaluation – US$2.60; CAD4TB – US$6.28; Xpert – US$19.20. Study authors noted that the costliest component of Xpert was the consumables, while human resources, equipment and CAD4TB score analysis contributed equally to those costs	The costs should be considered in relation to the budget of the entity paying for the screening (correctional services, ministry of health, donors, etc.). The group notes that the costs of prison screening are typically not borne by the TB programme. The GDG noted that the costs of screening in prisons will vary considerably depending on the algorithm used.

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	Four studies (4/4) provided direct evidence, but comparison was limited due to differences in setting (i.e. Europe, Africa and FSU) as well as the heterogeneity of ACF approaches among inmates; comparability and generalizability of study results are therefore limited. All four studies were assessed to be of good quality.	

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 intervention ○ Varies ○ No included studies	Two studies (2/4) reported the average cost per person diagnosed with TB which ranged from US$451-$1,513 per prisoner/inmate (Machekera 2019, Zishiri 2014). Machekera (2019) only included operational costs of personnel and laboratory consumables necessary for screening, which resulted in a lower cost per prisoner screened. Zishiri (2014) found a significantly higher cost per inmate screened due to an extensive program which screened 87% of all newly admitted inmates with WHO 4SS followed by Xpert and staffing costs attributable to screening were a major driver of cost. Two studies (2/4) assessed the cost-effectiveness of systematic screening for active TB in prisoners (Winetsky 2012, Smit 2017). Both studies found mass screening with CXR among prisoners to cost-effective as compared with standard of care (i.e. PCF) with ICERs of $543 per QALY gained and €11,603 (95% UI: €9,010 €14,909) per additional case of active TB detected. Da Silva Santos (2020) reported the following costs per inmate diagnosed with TB based on the screening algorithm: 1) Sputum Xpert for all inmates – US$249; 2) WHO 4SS followed by Xpert among inmates with symptoms – US$255; 3) Triage with CAD4TB (v5) (score >=60) followed by Xpert – US$336; 4) WHO 4SS, if positive symptoms then Xpert, if no symptoms then CAD4TB – US$393.	The GDG noted that if prison screening reduces transmisison and averts future cases, screening could be cost-effective or even cost-saving (but the evidence does not yet clearly show this).

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	We have no research evidence regarding equity for those residing in prisons. TB transmission rates are high in prisons because there is often a high prevalence of TB among people who are incarcerated, and living conditions are often crowded. It can be theorized that screening for TB would increase equity if it increases access to health services to those who have limited access.	The GDG noted that screening in prisons probably increases equity as it would increase access to health services, which can be very limited in prison settings.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	A systematic review of qualitative studies was conducted to understand individual patient and community responses to systematic TB screening, but did not find any data relevant to screening in prisons. A study comparing screening algorithms in prisons in Brazil (Da Silva Santos 2020) found some reasons for not participating included: • Lack of interest, • Lack of clothing to leave the cell, and • Fear of meeting members of rival groups From the previous guideline - A quantitative systematic review from 2012 (Mitchell et al) found that, in 16 studies, the weighted average of eligible persons who consented to undergo TB screening in prisons was 72%; the range was 18–98%; and the median proportion was 86%.	The GDG noed that acceptability to all relevant stakeholders should be considered, including prison administrators, staff, funders, etc. Screening can introduce challenges for prison administrators including moving of prisoners,

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	*Population feasibility considerations* A survey of screening guideline users was conducted by WHO in 2019. Among 118 respondents (80% from NTP):	The GDG noted that, in many settings, there may be several potential challenges in implementation of TB screening in prisons,

	• 73% of respondents report the existing conditional recommendation for screening in prisons has been integrated into national guidelines (with or without adaptation), • 47% of respondents report that the existing recommendation is routinely conducted. A study in Brazil (Santos 2020) compared the yield and sensitivity of four screening algorithms among prisoners. They found: • Screening with Xpert MTB/RIF among all individuals able to produce sputum found 74% of prevalent cases (however only 27.2% of participants were able to produce a sputum sample in initial visit) • Screening with CXR-CAD found 73% of prevalent cases • Symptom screening followed by Xpert MTB/RIF found 65% of prevalent cases; • Symptom screening followed by Xpert for those with sympoms, and followed by CXR then Xpert for those without symptoms, had the highest yield at 78% of cases A study in Brazil (Paiao 2016) found that annual screening did not reduce subsequent incidence of TB among prisoners, perhaps due to the high levels of ongoing transmission; the authors suggested more frequent screening in this setting might be required to effectiveley reduce incidence. *Test-specific feasibility considerations* Screening for symptoms of TB is feasible in any setting. Screening with CXR requires access to radiography, in a health facility or a mobile screening unit. Access to high-quality radiography is limited in many prison settings, including access to trained personnel to read images; CXR can be brought via mobile vans. mWRDs are recommended by WHO as the primary mode of TB diagnosis. mWRDs are able to be implemented at most levels of health care but resource and availability may prohibit use for screening at scale.	including logistics of separation and infection control for patients detected. Never the less, the group felt that TB screening is feasible in most all settings. In line with this, the group also notes that, in many settings, entry screening is already conducted.

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	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

	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

Among people in prisons, the GDG **recommends** systematic screening for TB should be conducted (**strong recommendation, very low certainty of evidence**).

Related recommendation(s)

Test specific recommendations

Among the general population, the GDG suggests using chest radiography (any abnormality, abnormality suggestive if TB), symptom screening (any cough, prolonged cough, any TB symptom) or molecular WHO-approved rapid diagnostic tests rather than no screening tests for detection of TB disease **(conditional recommendation, very low certainty of the evidence for test accuracy).**

Justification

- People in prisons and other penitentiary institutions are at an increased risk for TB compared to the general population, and often have limited access to healthcare services. The estimated incidence of TB among people residing in prisons is 23 times higher than the general population (Bassano et al 2010). There was previously a conditional recommendation for screening in this population.
- Data reviewed for this guideline meeting suggests that screening in prisons may improve early case detection, increase overall TB case detection and reduce TB prevalence.
- Based on this new evidence and the high risk of TB seen in this population, the GDG felt that this risk group now merited a strong recommendation for TB screening. The group felt that implementing TB screening in prisons has the potential to increase equity in access to healthcare, particularly in settings where existing health services in prisons are suboptimal.

Implementation considerations

- A prisoner is anyone held in a criminal justice facility or correctional facility during the investigation of a crime, anyone awaiting trial and anyone who has been sentenced.
- In addition, people residing in a correctional facility are almost always in close contact with several other inmates, thus whenever a person is diagnosed with TB within a prison, prisoners who have been in close contact with that person should be investigated, as per recommendations on screening of contacts.
- People who work in prisons and other penitentiary institutions are also at high risk of exposure to TB and should also be eligible for screening.
- When starting screening, it is important to ensure that good treatment and case management, as well as effective mechanisms for continuing treatment after transfer or release, are in place.
- The GDG felt that, at a minimum, screening in prisons and other penitentiary institutions should always include screening when a person enters a detention facility, annual screening, and screening upon exiting the facility in order to prevent reintroduction of TB into the broader community among people leaving detention. Treatment and follow up after release should also be ensured.
- Screening in prisons should be combined with efforts to improve living conditions and infection control. If possible, TB screening in prisons and other penitentiary institutions should be combined with screening for other diseases and health-promotion activities targeting this group.

Monitoring and evaluation

- The GDG noted that there are some settings in which this recommendation may not always apply. The Group suggested that careful monitoring and surveillance to assess the burden of TB, including prevalence and incidence, should always be done alongside prison TB screening. If the burden of TB is found to be non-existent and sustained at such levels for a period of time, TB screening can be discontinued.
- However, monitoring and surveillance to ensure TB disease and transmission do not re-emerge should be conducted.

Research priorities

- Further research on the full effectiveness and cost-effectiveness of TB screening in prisons is needed.
- Further research on equity and acceptability of screening among prisoners is needed.
- More data on implementation strategies for screening in prisons, including the accuracy and effectiveness of different screening algorithms, and more data on frequency of screening, are needed.

Table 4. Should chest X-ray with CAD software interpretation, compared to human reader interpretation, be used to screen for TB disease in people eligible for TB screening, using a bacteriologic reference standard?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	CXR has a long history of use as a TB screening tool. However, the inter-reader variability of human interpreters is substantial and access to trained radiologists is limited in resource-constrained high TB burden countries. Various computer assisted detection (CAD) systems have been developed and introduced in the market with the increasing progress in imaging and machine learning techniques. There is a need to understand their potential for use for TB screening in place of human readers.	

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	An evaluation of the range of accuracies of CAD tools to detect pulmonary TB adults set at 90% sensitivity, compared to a human reader with a microbiological reference standard. **Test accuracy** CXR with CAD software Sensitivity: 0.90 to 0.92 Specificity: 0.23 to 0.66 CXR with human reader Sensitivity: 0.82 to 0.93 Specificity: 0.14 to 0.63	• The GDG considered that when comparing accuracy of CAD to that of human reader interpretation of CXR, considering range of accuracies from different evaluations presented, the data suggest there is little difference between the two ranges. • Therefore the GDG endorsed the evidence was indicative of 'accurate'. • The GDG noted that the data for human reader interpretation of CXR indicates substantial variability. There is substantial overlap between ranges for CAD and human reader performance. • The GDG also notes that the human readers represented here are expert

radiologists, likely to perform better than many readers in the field.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT

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

RESEARCH EVIDENCE

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%		Prevalence 5%		Prevalence 10%			
	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)		
True positives patients with active TB	5 to 5	4 to 5	45 to 46	41 to 49	90 to 92	82 to 98	1325 (3)	⊕⊕⊕◯ MODERATE[a]
	1 more to 0 fewer TP in CXR with CAD software		4 more to 3 fewer TP in CXR with CAD software		8 more to 6 fewer TP in CXR with CAD software			
False negatives patients incorrectly classified as not having active TB	0 to 0	0 to 1	4 to 5	1 to 9	8 to 10	2 to 18		
	1 fewer to 0 fewer FN in CXR with CAD software		4 fewer to 3 more FN in CXR with CAD software		8 fewer to 6 more FN in CXR with CAD software			
True negatives patients without active TB	229 to 658	136 to 622	219 to 628	130 to 594	207 to 595	123 to 563	8391 (3)	⊕⊕◯◯ LOW[a,b]
	93 more to 36 more TN in CXR with CAD software		89 more to 34 more TN in CXR with CAD software		84 more to 32 more TN in CXR with CAD software			
False positives patients incorrectly classified as having active TB	337 to 766	373 to 859	322 to 731	356 to 820	305 to 693	337 to 777		
	93 fewer to 36 fewer FP in CXR with CAD software		89 fewer to 34 fewer FP in CXR with CAD software		84 fewer to 32 fewer FP in CXR with CAD software			

a. There may be important differences in the expertise of human reader (comparator test) because performance depends greatly on years of experience and educational qualifications. However, 2 of the3 studies had highly qualified radiologists with more than 10 years of experience, we decided to not downgrade for indirectness.

ADDITIONAL CONSIDERATIONS

- GDG felt desirable effects for CAD were moderate compared to a human reader performance.
- The GDG noted that the performance of human readers, and thus desirable effects, varies greatly across settings and within settings, moreso than CAD.
- The GDG noted that other desirable effects beyond accuracy (TP, TN) would include scalability of the intervention itself. In focusing on absolute effects (TP, TN), even with the best-performing human reader for CXR, the CAD technology demonstrates a moderate desirable effects. In the field, availability of radiologists is often scarce, and general practitioners would usually be made to read, who may not be as highly-skilled readers as used for comparison here, so this data may represent an underestimate of the true difference in accuracy.

b. The range around true negatives and false positives is wide, however the difference of the ranges between index test and comparator test is not large. We downgraded one level for imprecision.

Desirable: Identifying true positive and negatives: The anticipated desirable effect is the correct classification of people with active TB as screening positive for TB (true positive), resulting in appropriate referral for further evaluation (as indicated by the screening algorithm in use), and increasing the probability of ultimately leading to timely diagnosis and treatment and reducing further transmission of the mycobacterium. Another anticipated desirable effect is correctly ruling out disease in people who do not have TB (true negative), avoiding unnecessary further diagnostic evaluation (and associated resources for the person and the health system), providing reassurance to the person undergoing screening, allowing for the pursuit of an alternative diagnosis if they have respiratory symptoms, and allowing for the determination of eligibility for TB preventive therapy, if indicated.

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT

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

RESEARCH EVIDENCE

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%		Prevalence 5%		Prevalence 10%			
	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)		
True positives patients with active TB	5 to 5	4 to 5	45 to 46	41 to 49	90 to 92	82 to 98	1325 (3)	⊕⊕⊕◯ MODERATE[a]
	1 more to 0 fewer TP in CXR with CAD software		4 more to 3 fewer TP in CXR with CAD software		8 more to 6 fewer TP in CXR with CAD software			
False negatives patients incorrectly classified as not having active TB	0 to 0	0 to 1	4 to 5	1 to 9	8 to 10	2 to 18		
	1 fewer to 0 fewer FN in CXR with CAD software		4 fewer to 3 more FN in CXR with CAD software		8 fewer to 6 more FN in CXR with CAD software			
True negatives patients without active TB	229 to 658	136 to 622	219 to 628	130 to 594	207 to 595	123 to 563	8391 (3)	⊕⊕◯◯ LOW[a,b]
	93 more to 36 more TN in CXR with CAD software		89 more to 34 more TN in CXR with CAD software		84 more to 32 more TN in CXR with CAD software			
False positives	337 to 766	373 to 859	322 to 731	356 to 820	305 to 693	337 to 777		

ADDITIONAL CONSIDERATIONS

- In comparison to the absolute number of TP and FP, the GDG felt that the reported number of FP and FN (undesirable effects) is small.
- The GDG notes that, in terms of absolute numbers, CAD will identify more people who have TB and less people without TB who need follow-up diagnostics compared to human reader, which provides benefit for individuals as well as programs.
- Beyond the diagnostic accuracy data, the GDG notes that another undesirable effect of CAD (as presented here) is that it may not detect other chest conditions/issues whereas a human reader can. The capacity to

patients incorrectly classified as having active TB	93 fewer to 36 fewer FP in CXR with CAD software	89 fewer to 34 fewer FP in CXR with CAD software	84 fewer to 32 fewer FP in CXR with CAD software		

a. There may be important differences in the expertise of human reader (comparator test) because performance depends greatly on years of experience and educational qualifications. However, 2 of the3 studies had highly qualified radiologists with more than 10 years of experience, we decided to not downgrade for indirectness.
b. The range around true negatives and false positives is wide, however the difference of the ranges between index test and comparator test is not large. We downgraded one level for imprecision.

Undesirable: Identifying false positive and negatives: The anticipated undesirable effect is the incorrect classification of people without active TB as screening positive for TB (false-positive), with resulting anxiety, resources required for further unnecessary diagnostic evaluation (for both the patient and the health system), and the risk of a false-positive diagnosis, leading to inappropriate treatment and related social, economic, or health consequences (including adverse effects of treatment). Another anticipated undesirable effect is the incorrect classification of people who have active TB as screening negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, and potentially inappropriate preventive therapy treatment with associated risk of development of drug resistance.

simultaneously screen for multiple pulmonary or thoracic conditions potentially could be added to CAD technologies but would require distinct validations and is not included in the data presented to the GDG.

Certainty of the evidence of test accuracy

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

JUDGEMENT

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

RESEARCH EVIDENCE

The certainty of the evidence is low: The review consisted of 3 studies with 9,716 participants.

1. There may be important differences in the expertise of human reader (comparator test) because performance depends greatly on years of experience and educational qualifications. However, 2 of the 3 studies had highly qualified radiologists with more than 10 years of experience, we decided to not downgrade for indirectness.

2. The range around true negatives and false positives is wide, however the difference in the ranges between index test and comparator test is not large. We downgraded one level for imprecision.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0%		Prevalence 5%		Prevalence 10%			
	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)		
True positives patients with active TB	5 to 5	4 to 5	45 to 46	41 to 49	90 to 92	82 to 98	1325 (3)	⊕⊕⊕◯ MODERATE[a]
	1 more to 0 fewer TP in CXR with CAD software		4 more to 3 fewer TP in CXR with CAD software		8 more to 6 fewer TP in CXR with CAD software			
False negatives patients incorrectly classified as not	0 to 0	0 to 1	4 to 5	1 to 9	8 to 10	2 to 18		
	1 fewer to 0 fewer FN in CXR with CAD software		4 fewer to 3 more FN in CXR with CAD software		8 fewer to 6 more FN in CXR with CAD software			

ADDITIONAL CONSIDERATIONS

- The GDG notes that the data presented here only indirectly answer the question about the accuracy of CAD for screening people irrespective of symptoms of TB, as most individuals included in the three studies were pre-screened before they underwent a CXR (and thus inclusion in the analysis for diagnostic accuracy) based on presence of symptoms or abnormal CXR.
- The GDG also notes that in this analysis the accuracy of CAD in comparison to a human reader is likely underestimated, because the comparator was a highly qualified

having active TB								
True negatives patients without active TB	229 to 658	136 to 622	219 to 628	130 to 594	207 to 595	123 to 563	8391 (3)	⊕⊕◯◯ LOW[a,b]
	93 more to 36 more TN in CXR with CAD software		89 more to 34 more TN in CXR with CAD software		84 more to 32 more TN in CXR with CAD software			
False positives patients incorrectly classified as having active TB	337 to 766	373 to 859	322 to 731	356 to 820	305 to 693	337 to 777		
	93 fewer to 36 fewer FP in CXR with CAD software		89 fewer to 34 fewer FP in CXR with CAD software		84 fewer to 32 fewer FP in CXR with CAD software			

a. There may be important differences in the expertise of human reader (comparator test) because performance depends greatly on years of experience and educational qualifications. However, 2 of the3 studies had highly qualified radiologists with more than 10 years of experience, we decided to not downgrade for indirectness.
b. The range around true negatives and false positives is wide, however the difference of the ranges between index test and comparator test is not large. We downgraded one level for imprecision.

and trained specialist human reader. The human reader used as comparison in these studies does not represent the routine human readers in most settings. Bias is against CAD, therefore may be under-estimating benefit of CAD compared to CXR.

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. No direct benefits, adverse effects or burden to the person being screened resulting directly from the use of CAD technologies are anticipated. There may be concerns regarding privacy and security of patient data.	The GDG noted that direct effects of a human reader of CXR can include identification of other pathologies, compared to CAD.

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	The primary objective of screening for active TB is to ensure that active TB is detected early and treatment is initiated promptly, with the ultimate aim of reducing the risk of poor treatment outcomes, health sequelae and the adverse social and economic consequences of TB, as well as helping to reduce TB transmission. Treatment of drug sensitive TB is highly effective (approximately 90%, strong recommendation, high certainty in the estimates of effect). Treatment of MDR TB can be effective as well, if quality assured (approximately 70%) (WHO consolidated treatment guidelines 2020).	

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	TB outreach screening, may increase tuberculosis case detection (RR 1.24, 95% CI 0.86 to 1.79; 4 trials, 6,458,591 participants in 297 clusters, low-certainty evidence); and probably increases case detection in areas with tuberculosis prevalence of 5% or more (RR 1.52, 95% CI 1.10 to 2.09; 3 trials, 155,918 participants, moderate-certainty evidence). These interventions may lower the early default (prior to starting treatment) or default during treatment (RR 0.67, 95% CI 0.47 to 0.96; 3 trials, 849 participants, low-certainty evidence). However, this intervention may have little or no effect on treatment success (RR 1.07, 95% CI 1.00 to 1.15; 3 trials, 849 participants, low-certainty evidence), and we do not know if there is an effect on treatment failure or mortality. (Cochrane review, Mhimbira et al 2017).	The GDG felt there was no reason to believe the linkage to care in the screening cascade would be different between a human reader and CAD for screening with CXR, except perhaps a potential for higher linkage with CAD if it makes the capacity to read the image quicker and more readily available.

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	There was no research evidence available for effectiveness of CAD compared to CXR for people being screened for TB.	Research gap - further studies are indicated.

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 direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is important uncertainty in what we know about how patients perceive and value the outcomes explored above.	Although the GDG felt that most people affected by TB would probably value the main outcomes (being diagnosed earlier, starting TB treatment more promptly), for the specific intervention there was important uncertainly about how much people valued human vs CAD reading of CXRs. This includes the consideration of a human reader's ability to detect additional conditions.

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 panel determined that the balance between desirable and undesirable effects favors **using** CAD as a screening test for detection of TB.	The GDG notes that CAD probably favors intervention, particularly in settings where experienced readers are not available.

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 the costing systematic review, no included studies directly addressed this subquestion in which CXR with CAD software were used to screen for active TB in people eligible for TB screening. Data was available from one study that provided indirect evidence on the resources required for screening with CXR with CAD software for screening in the context of a passive case finding (PCF) program, not in an active case finding (ACF) program (Philipsen 2015). <u>Indirect evidence:</u> One study (Philipsen, 2015) conducted among persons who self-referred for TB testing (PCF) in South Africa, used CAD4TB as a <u>triage test</u> compared to Xpert alone. This study found the CAD4TB automated CXR (ACR) cost US$5-$13 per person screened, compared to US$13 per person screened with Xpert alone. Program costs reported for CAD4TB in this study included: 1) Equipment (direct digital CXR, PACS, shipment, CAD4TB software and training): US$263,003.40 2) Annual running costs: US$46,847.14 3) Annual HR costs (i.e. technician and training/monitoring): US$28,500 The average ACR cost in first 10 years (assuming 50,000 tests run per year – high volume): US$1.46. Major drivers of programmatic costs included the throughput (volume) of persons being screened using CAD4TB, underlying prevalence of TB and HIV, as well as equipment, maintenance and software purchase costs. The capital expenditures drive up the cost per person if the machine and software is not used frequently or by a large number of people, which also may be driven by the setting (i.e. urban vs. rural).	● There are large costs associated with CAD including investment in a new program, the resources (software licensing, etc.), required ● X-ray equipment costs should not factor in this question as they would need to be in place regardless of the reader used. ● The expected degree of scale-up is an important factor in overall costs, as CAD is more easily scalable than CXR with human readers.

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	There were no studies that identified the resource requirements for this subquestion in which CXR with CAD software were used to screen for active TB in people eligible for TB screening. The lack of direct evidence is a limitation, only one study contributed indirect evidence but was of good quality.	

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 intervention ○ Varies ● No included studies	There were no published studies that assessed the cost-effectiveness of this sub-question in which CXR with CAD software were used to screen for active TB in people eligible for TB screening.	The GDG noted that cost-effectiveness will depend on the setting, including the current availability and salaries of human readers.

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 studies were sought for this domain.	• The GDG noted that CAD could increase equity if it increased scale-up of TB screening, and thus access to diagnosis and care for TB. • The GDG also noted that screening can impact vulnerable groups if it results in deportation or impacts migration. This should not differ with CAD compared to human reading of CXR.

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 was found.	The GDG noted this is an important research gap that could be addressed through qualitative assessment to see how patients, clinicians, other stakeholders view the intervention.

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	No published evidence found. Unpublished evidence from STOPTB partnership user-testing interviews. Sample interviewed: Principal investigators (PIs) or Co PI, Field managers, Project Directors, Data Managers, Active Case Finding Managers, Study Site Coordinators 10 Countries (Bangladesh, Cameroon, India, Malawi, Nepal, Pakistan, Peru, South Africa, Vietnam, Zambia 11 TB REACH projects 3 Non TB REACH funded 14 Different Projects Access to well trained human resources for reading X-rays is a bottleneck particularly in large scale implementation, but also mentioned in research (less than 10 radiologists in different countries). As there is no published data on user preferences/acceptability/ feasibility of these softwares, as part of 11 TB REACH projects in eight countries, information on user feedback was gathered using semi-structured interviews. <u>Reasons to use artifical intelligence (AI):</u> • Number of mobile units can be increased without need for Human Resource • The AI software allows high-throughput without reader fatigue • Consistent results and the ability to adapt quickly to changes in plans (training humans may be more time consuming and challenging) <u>Installation. training, workflow:</u> • Generally very easy to use. Reading is automatic and interpretation is based on a score for most software programmes. • Installation would require IT support, however it can be quick, especially for offline solutions • Easy to be used by non-clinical staff • Remote solutions are available, however, for large scale implementation might be an issue for some smaller companies • Some softwares mimic the radiology report, which is more helpful than the abnormality score. <u>Barriers:</u> • Internet connectivity is a key concern- especially in rural/remote areas, so access to offline boxes are critical • Cost is a concern	• The GDG noted that feasibility depends heavily on the setting, access to required equipment, stable internet, maintenance, etc. • Threshold calibration will be needed for each implementation of CAD, based on evidence of variability of performance of CAD across software and within software across different settings and populations.

	• Some AI packages charge a 'validation fee' while many do not – this seems excessive in addition to hardware and per read payments <u>Threshold scores and its use:</u> • General agreement among those interviewed that fixing threshold scores would not make sense. • Major advantage of AI: a continuous score that can be moved depending on the setting and local needs • Human readers do not use a threshold score - will definitely perform differently • Key concern: need to understand the score that should be used in each population in terms of a pilot screening approach.	

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
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**

	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

The GDG suggests that **either Computer Assisted Detection (CAD) or human expert interpretation of plain chest radiographs may be used** for the screening for pulmonary tuberculosis in individuals aged fifteen years and over (conditional recommendation, low certainty of the evidence for test accuracy)

Justification

- The judgement by the GDG for a conditional recommendation for either the intervention or the comparison is derived from analyses that found human expert reading and CAD performed similarly when interpreting plain radiography to screen for pulmonary TB in individuals aged fifteen years and over. The recommendation applies only to the interpretation of antero-posterior or postero-anterior views of plain chest radiographs for pulmonary TB; it does not apply to the interpretation of lateral or oblique views.
- This recommendation applies to the class of CAD softwares for automated detection of plain chest radiography for TB; individual softwares to be used will need to undergo external validated study showing diagnostic accuracy (sensitivity/specificity) and performance at different levels of prevalence to be non-inferior to the three products reviewed for this recommendation.

Subgroup considerations

- This recommendation applies to adolescents and adults 15 years of age and older in the context of screening for TB.
- The data assessed in this recommendation showed possible variation in performance of CAD according to certain characteristics of individuals being evaluated, including smear status, HIV status, age cohort, history of previous TB, smoking status, and gender, but the majority of differences did not reach statistical significance. It is unknown the extent to which performance of CXR with human reading varies similarly.

Implementation considerations

- The infrastructural requirements to operate the software appropriately at the site of use - computer, digital radiography equipment, internet, maintenance materials - will be indispensable to use of CAD and should be thoroughly assessed prior to implementation.
- Calibration of CAD software will need to be performed in each setting or population in which it will be used for screening, due to the observations of variability of the accuracy (sensitivity, specificity) across software, and variability of the software at given thresholds across different contexts, even at manufacturer-recommended thresholds.

Monitoring and evaluation

- After initial calibration is done, ongoing monitoring and analysis of CAD performance should be done to assess: consistency with human reader interpretation, proportion of images read as abnormal and requiring further investigation, positivity rate among images read as abnormal.

Research priorities

- The development and evaluation of CAD software for automated screening for TB in children is urgently needed, as CXR is an important tool in detection of pulmonary TB in children and adolescents, given the difficulty in bacteriological testing and diagnosis.
- Further evidence on the performance of CAD software according to the characteristics of the individual being evaluated (eg smear status, HIV status, age cohort, history of previous TB, smoking status, gender) is needed, to allow for better setting-specific and patient-specific calibration of thresholds used for CAD.

Table 5. Should chest X-ray with CAD software interpretation, compared to human reader interpretation, be used to triage for TB disease in people eligible for TB triage, using a bacteriologic reference standard?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	CXR has a long history of use as a TB screening, triage, and diagnostic tool. However, the inter-reader variability of human interpreters is substantial and access to trained radiologists is limited in resource-constrained high TB burden countries. Various computer assisted detection (CAD) systems have been developed and introduced in the market with the increasing progress in imaging and machine learning techniques. There is a need to understand their potential for use for TB triage in place of human readers. Triaging is defined as the process of deciding the diagnostic and care pathways for people, based on their symptoms, signs, risk markers, and test results. Triaging involves assessing the likelihood of various differential diagnoses as a basis for making clinical decisions. It can follow more- or less-standardised protocols and algorithms and may be done in multiple steps (WHO 2016). A triage test for TB is a test that can be rapidly conducted among people presenting to a health facility to identify those who should go on to further diagnostic evaluation for TB (those whose TB triage test is positive/abnormal) from those who should undergo other further investigation for non-TB diagnoses (for those whose TB triage testing is negative/normal). Triage of people with signs and symptoms of TB is recommended to reduce transmission of *M. tuberculosis* to health care workers, health center attendees, and others in settings with high risk of transmission (Conditional recommendation, very low certainty of estimates of effects).	

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	An evaluation of the range of accuracies of CAD tools to detect pulmonary TB adults set at 90% sensivity, compared to a human reader with a microbiological reference standard. **Test accuracy:** CXR with CAD software - Sensitivity: 0.90 to 0.91 Specificity: 0.25 to 0.79 Human reader (any TB abnormality) - Sensitivity: 0.89 to 0.96 Specificity: 0.36 to 0.63	• The GDG felt that the data demonstrated that CAD reading of digital chest radiography images has similar accuracy to human readers; therefore the GDG endorsed the evidence as indicative of "accurate". • While the data showed that CAD had substantial variability (greater than human readers represented), the GDG noted that the true variability in human readers across the global context is expected to be much greater (and less accurate on average) than the human readers represented here (given the limited selection of highly skilled human readers present in the sample).

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ● Small ○ Moderate ○ Large ○ Varies ○ Don't know		• The GDG felt that the difference in desirable effects (including true positives and true negatives) was small between the human reader and CAD technologies presented here. • As with the CAD for screening, the Group also mentioned some possible desirable effects beyond the count of true positives and true negatives included the ability to increase access to CXR for triage even in the absence of or with limited access to trained personnel to interpret CXR images.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%		Prevalence 20%		Prevalence 30%			
	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)		
True positives patients with active TB	90 to 91	89 to 96	180 to 182	177 to 192	270 to 273	266 to 288	(3)	⊕⊕⊕◯ MODERATE[a]
	1 more to 5 fewer TP in CXR with CAD software		3 more to 10 fewer TP in CXR with CAD software		4 more to 15 fewer TP in CXR with CAD software			
False negatives patients incorrectly classified as not having active TB	9 to 10	4 to 11	18 to 20	8 to 23	27 to 30	12 to 34		
	1 fewer to 5 more FN in CXR with CAD software		3 fewer to 10 more FN in CXR with CAD software		4 fewer to 15 more FN in CXR with CAD software			
True negatives patients without active TB	223 to 711	329 to 563	198 to 632	292 to 500	174 to 553	256 to 438	(3)	⊕⊕◯◯ LOW[a,b]
	106 fewer to 148 more TN in CXR with CAD software		94 fewer to 132 more TN in CXR with CAD software		82 fewer to 115 more TN in CXR with CAD software			
False positives patients incorrectly classified as having active TB	189 to 677	337 to 571	168 to 602	300 to 508	147 to 526	262 to 444		
	106 more to 148 fewer FP in CXR with CAD software		94 more to 132 fewer FP in CXR with CAD software		82 more to 115 fewer FP in CXR with CAD software			

a. Of the three studies, one study had only 56% of the patients presenting with signs and symptoms at the referral hospital. One site in this study (Japan) had patients who were going to the health care centre as part of their regular check-up for active TB. This site also included healthy individuals.

b. The range around true negatives and false positives is wide, however the difference of the ranges between index test and comparator test is not large. We downgraded one level for imprecision.

Desirable: Identifying true positive and negatives: The anticipated desirable effect is the correct classification of people with active TB as screening positive for TB (true positive), resulting in appropriate referral for further evaluation (as indicated by the screening algorithm in use), and increasing the probability of ultimately leading to timely diagnosis and treatment and reducing further transmission of the mycobacterium. Another anticipated desirable effect is correctly ruling out disease in people who do not have TB (true negative), avoiding unnecessary further diagnostic evaluation (and associated resources for the person and the health system), providing reassurance to the person undergoing screening, allowing for the pursuit of an alternative diagnosis if they have respiratory symptoms, and allowing for the determination of eligibility for TB preventive therapy, if indicated.

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT

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

RESEARCH EVIDENCE

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%		Prevalence 20%		Prevalence 30%			
	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)		
True positives patients with active TB	90 to 91	89 to 96	180 to 182	177 to 192	270 to 273	266 to 288	(3)	⊕⊕⊕◯ MODERATE[a]
	1 more to 5 fewer TP in CXR with CAD software		3 more to 10 fewer TP in CXR with CAD software		4 more to 15 fewer TP in CXR with CAD software			
False negatives patients incorrectly classified as not having active TB	9 to 10	4 to 11	18 to 20	8 to 23	27 to 30	12 to 34		
	1 fewer to 5 more FN in CXR with CAD software		3 fewer to 10 more FN in CXR with CAD software		4 fewer to 15 more FN in CXR with CAD software			
True negatives patients without active TB	223 to 711	329 to 563	198 to 632	292 to 500	174 to 553	256 to 438	(3)	⊕⊕◯◯ LOW[a,b]
	106 fewer to 148 more TN in CXR with CAD software		94 fewer to 132 more TN in CXR with CAD software		82 fewer to 115 more TN in CXR with CAD software			
False positives patients incorrectly classified as having active TB	189 to 677	337 to 571	168 to 602	300 to 508	147 to 526	262 to 444		
	106 more to 148 fewer FP in CXR with CAD software		94 more to 132 fewer FP in CXR with CAD software		82 more to 115 fewer FP in CXR with CAD software			

a. Of the three studies, one study had only 56% of the patients presenting with signs and symptoms at the referral hospital. One site in this study (Japan) had patients who were

ADDITIONAL CONSIDERATIONS

- The GDG felt that the trend in change of numbers of FP and FN in use of CAD versus human readers was small.
- The GDG notes that, on average, CAD will find more people who have TB while also finding fewer people without TB who require investigation compared to human reader, which provides benefit for individuals as well as programs.
- Beyond the diagnostic accuracy data, the GDG notes that another undesirable effect of CAD is that it may not detect other chest conditions/issues whereas a human reader can. The capacity to simultaneously evaluate for multiple pulmonary or thoracic conditions potentially could be added to CAD technologies but would require distinct validations and is not included in the data presented to the GDG.

going to the health care centre as part of their regular check-up for active TB. This site also included healthy individuals.

b. The range around true negatives and false positives is wide, however the difference of the ranges between index test and comparator test is not large. We downgraded one level for imprecision.

Undesirable: Identifying false positive and negatives: The anticipated undesirable effect is the incorrect classification of people without active TB as screening positive for TB (false-positive), with resulting anxiety, resources required for further unnecessary diagnostic evaluation (for both the patient and the health system), and the risk of a false-positive diagnosis, leading to inappropriate treatment and related social, economic, or health consequences (including adverse effects of treatment). Another anticipated undesirable effect is the incorrect classification of people who have active TB as screening negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, and potentially inappropriate preventive therapy treatment with associated risk of development of drug resistance.

Certainty of the evidence of test accuracy

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

JUDGEMENT

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

RESEARCH EVIDENCE

The certainty of the evidence is low: The review consisted of 3 studies with 9,716 participants.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%		Prevalence 20%		Prevalence 30%			
	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)	CXR with CAD software	human reader (any TB abnormality)		
True positives patients with active TB	90 to 91	89 to 96	180 to 182	177 to 192	270 to 273	266 to 288	(3)	⊕⊕⊕◯ MODERATE[a]
	1 more to 5 fewer TP in CXR with CAD software		3 more to 10 fewer TP in CXR with CAD software		4 more to 15 fewer TP in CXR with CAD software			
False negatives patients incorrectly classified as not having active TB	9 to 10	4 to 11	18 to 20	8 to 23	27 to 30	12 to 34		
	1 fewer to 5 more FN in CXR with CAD software		3 fewer to 10 more FN in CXR with CAD software		4 fewer to 15 more FN in CXR with CAD software			
True negatives patients without active TB	223 to 711	329 to 563	198 to 632	292 to 500	174 to 553	256 to 438	(3)	⊕⊕◯◯ LOW[a,b]
	106 fewer to 148 more TN in CXR with CAD software		94 fewer to 132 more TN in CXR with CAD software		82 fewer to 115 more TN in CXR with CAD software			
False positives	189 to 677	337 to 571	168 to 602	300 to 508	147 to 526	262 to 444		

ADDITIONAL CONSIDERATIONS

- The GDG notes that there may be important differences in the expertise of human reader (comparator test) because performance depends greatly on years of experience and educational qualifications.
- The GDG also notes that in this analysis the accuracy of CAD is likely underestimated in comparison to a human reader, therefore may be under-estimating benefit of CAD compared to human reader for CXR.

patients incorrectly classified as having active TB	106 more to 148 fewer FP in CXR with CAD software	94 more to 132 fewer FP in CXR with CAD software	82 more to 115 fewer FP in CXR with CAD software		

a. Of the three studies, one study had only 56% of the patients presenting with signs and symptoms at the referral hospital. One site in this study (Japan) had patients who were going to the health care centre as part of their regular check-up for active TB. This site also included healthy individuals.
b. The range around true negatives and false positives is wide, however the difference of the ranges between index test and comparator test is not large. 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 direct evidence was considered here. No direct benefits, adverse effects or burden to the person being screened resulting directly from the use of CAD technologies are anticipated. There may be concerns regarding privacy and security of patient data.	The GDG noted that direct effects of a human reader of CXR can include identification of other pathologies beyond TB, compared to CAD.

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	The primary objective of screening or triage for active TB is to ensure that active TB is detected early and treatment is initiated promptly, with the ultimate aim of reducing the risk of poor treatment outcomes, health sequelae and the adverse social and economic consequences of TB, as well as helping to reduce TB transmission. Treatment of drug sensitive TB is highly effective (approximately 90%, strong recommendation, high certainty in the estimates of effect). Treatment of MDR TB can be effective as well, if quality assured (approximately 70%) (WHO consolidated treatment guidelines 2020).	

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	No research evidence was identified.	The GDG felt there was no reason to believe the linkage to care in the triage pathway would be different between a human reader and CAD for screening with CXR, except perhaps a potential for better linkage with CAD if it increases the capacity to read the image and makes it available more quickly.

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	No published research evidence identified. An unpublished clinical trial from Malawi (MacPherson et al) provides indirect evidence regarding the effects of CAD. The trial reported an increase in case-finding among patients triaged with digital CXR using CAD technologies for same-day reading and referral for confirmatory testing. Participants undergoing triage with CAD-read CXR were more likely to be diagnosed with TB compared to two different types of clinician-led triage (3.0% compared to 1.6%, 1.1%.) – none of them included CXRs read by human readers Time to treatment initiation was shorter in the CAD triage arm compared to clinician-led triage (HR 2.84, 95%CI: 1.03-7.80).	This is identified as a research gap.

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 direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is important uncertainty in what we know about how patients perceive and value the outcomes explored above.	The GDG felt that most people affected by TB would probably value the main outcomes. However, for this specific intervention there is important uncertainty about how much people value human vs CAD reading of CXR. This includes the consideration of a human reader's ability to detect additional conditions.

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 panel determined that the balance between desirable and undesirable effects **does not favor either using or not using** CAD as a screening test for detection of TB in triage settings.	Based on the data available the GDG feels that it is not clear that the accuracy data favor either CAD or human readers.

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	The systematic review identified one study (Philipsen, 2015) conducted among persons who self-referred for TB testing (PCF) in South Africa, used CAD4TB as a triage test compared to Xpert alone. This study found the CAD4TB automated CXR (ACR) cost US$5-$13 per person screened, compared to US$13 per person screened with Xpert alone. Program costs reported for CAD4TB in this study included: 1) Equipment (direct digital CXR, PACS, shipment, CAD4TB software and training): US$263,003.40 2) Annual running costs: US$46,847.14 3) Annual HR costs (i.e. technician and training/monitoring): US$28,500 The average ACR cost in first 10 years (assuming 50,000 tests run per year – high volume): US$1.46. Major drivers of programmatic costs included the throughput (volume) of persons being screened using CAD4TB, underlying prevalence of TB and HIV, as well as equipment, maintenance and software purchase costs. The capital expenditures drive up the cost per person if the machine and software is not used frequently or by a large number of people, which also may be driven by the setting (i.e. urban vs. rural).	● There are large costs associated with CAD including investment in a new program, the resources (hardware, software licensing, etc.), required. However there is the potential for cost savings associated with less requirement for human reader interpretation. ● X-ray equipment costs should not factor in this question as they would need to be in place regardless of the reader used. ● Based on the data presented here, the GDG felt that the difference in resources required between CAD and human reader interpretation of CXR is most likely negligible.

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	There were no published studies that directly assessed the cost-effectiveness of this subquestion of comparing triage with CAD software compared to human reading of CXR. The lack of direct evidence is a limitation, only one study contributed indirect evidence but was of good quality.	

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 intervention ○ Varies ● No included studies	There were no published studies that directly assessed the cost-effectiveness of this subquestion of comparing triage with CAD software compared to human reading of CXR. Philipsen et al. (2015) compared screening all PLHIV with Xpert to a triage algorithm using CAD to screen prior to Xpert MTB/RIF. Using South African costs from the literature, this study found the average cost per CXR per patient using CAD (including equipment, running costs, human resources, investment and 25 yearly running costs) to be US$1.46. The cost per screened subject using Xpert alone was US$13.09 and US$90.70 per notified TB case, compared to a cost per patient screened using CAD software (triage before Xpert) ranging from US$5.27-$12.72 and US$81.75-$88.15 per notified TB case. MacPherson et al. performed a cost-effectiveness analysis of an RCT comparing adults (>18 years old) in Malawi who were randomized to digital CXR (DCXR-CAD) using CAD4TB version 5 as a triage test, followed by Xpert MTB/RIF for confirmation of TB among those individuals with a reading of 45 or above. The intervention group had an additional screening cost of US$19.92 per person as compared to standard of care (i.e. clinician directed TB screening). However, the triage testing approach was not found to be cost-effective with an ICER of US$4620.47 per QALY gained as compared to standard of care, which did not meet any of the a priori WTP thresholds (US$400/QALY, $800/QALY, $1200/QALY). The major driver of costs was the number of Xpert tests conducted and the running and implementation costs of CAD4TB.	The GDG notes that cost-effectiveness will depend on the setting, including the current availability and salaries of human readers, as with CAD for screening.

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 studies were sought for this domain.	• The GDG noted that CAD could increase equity if it increased access to CXR for triage for TB and referral for further diagnostic evaluation. • The GDG also noted that CAD technologies are not currently developed or approved to be used with children, which could decrease equity.

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 was found.	• The GDG noted this is an imporant research gap that could be addressed through qualitative assessment to see how patients, clinicians, other stakeholders view the intervention. • Acceptability will likely vary with context.

		• However the GDG overall felt that likely, in the triage context, that patients and users would find it acceptable to replace a human reader with CAD interpretation of CXR.

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	No published evidence found. Unpublished evidence from STPTB partnership user-testing interviews. Sample interviewed: Principal investigators (PIs) or Co PI, Field managers, Project Directors, Data Managers, active case finding Managers, Study Site Coordinators 10 Countries (Bangladesh, Cameroon, India, Malawi, Nepal, Pakistan, Peru, South Africa, Vietnam, Zambia 11 TB REACH projects 3 Non TB REACH funded 14 Different Projects Access to well trained human resources for reading X-rays is a bottleneck particularly in large scale implementation, but also mentioned in research (less than 10 radiologists in different countries). As there is no published data on user preferences/acceptability/ feasibility on these software, as part of 11 TB REACH projects in eight countries, information on user feedback was gathered using semi-structured interviews. <u>Reasons to use articificial intelligence (AI):</u> • Number of mobile units can be increased without need for Human Resource • The AI software allows high-throughput without reader fatigue • Consistent results and the ability to adapt quickly to changes in plans (training humans may be more time consuming and challenging) <u>Installation. training, workflow:</u> • Generally very easy to use. Reading is automatic and interpretation is simple based on a score. • Installation would require IT support, however it can be quick, especially for offline solutions • Easy to be used by non-clinical staff • Remote solutions are available, however, for large scale implementation might be an issue for some smaller companies • Some softwares mimic the radiology report, which is more helpful than the abnormality score. <u>Barriers:</u> • Internet connectivity is a key concern- especially in rural/remote areas, so access to offline boxes are critical • Cost is a concern • Some AI packages charge a 'validation fee' while many do not – this seems excessive in addition to hardware and per read payments <u>Threshold scores and its use:</u>	• The GDG noted that feasibility depends heavily on the setting, access to required equipment, stable internet, maintenance, etc. • Threshold calibration will be needed for each implementation of CAD, based on evidence of variability of performance of CAD across software and within software across different settings and populations. • In some settings where human readers are available, it may not be feasible or beneficial to implement CAD (e.g Netherlands).

	• General agreement among those interviewed that fixing threshold scores would not make sense. • Major advantage of AI: a continuous score that can be moved depending on the setting and local needs • Human readers do not use a threshold score - will definitely perform differently • Key concern: need to understand the score that should be used in each population in terms of a pilot screening approach.	

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**
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

The GDG suggests that **either Computer Assisted Detection (CAD) or human expert interpretation of plain chest radiographs may be used** for triage for TB disease in individuals aged fifteen years and over (conditional recommendation, low certainty of the evidence for test accuracy)

Justification

- The judgement by the GDG for a conditional recommendation for either the intervention or the comparison is derived from analyses that found human expert reading and CAD performed similarly when interpreting plain radiography for triage for pulmonary TB in individuals aged fifteen years and over. The recommendation applies only to the interpretation of antero-posterior or postero-anterior views of plain chest radiographs for pulmonary TB; it does not apply to the interpretation of lateral or oblique views.
- This recommendation applies to the class of CAD softwares for automated detection of plain chest radiography for TB; individual softwares to be used will need to undergo external validated study showing diagnostic accuracy (sensitivity/specificity) and performance at different levels of prevalence to be non-inferior to the three products reviewed for this recommendation.

Subgroup considerations

- The data assessed in this recommendation showed possible variation in performance of CAD according to certain characteristics of individuals being evaluated, including smear status, HIV status, age cohort, history of previous TB, smoking status, and gender, but the majority of differences did not reach statistical significance. It is unknown the extent to which performance of CXR with human reading varies similarly.

Implementation considerations

- The infrastructural requirements to operate the software appropriately at the site of use - computer, digital radiography equipment, internet, maintenance materials - will be indispensable to use of CAD and should be thoroughly assessed prior to implementation.
- Calibration of CAD software will need to be performed in each setting or population in which it will be used for screening, due to the observations of variability of the accuracy (sensitivity, specificity) across software, and variability of the software at given thresholds across different contexts, even at manufacturer-recommended thresholds.

Monitoring and evaluation

After initial calibration is done, ongoing monitoring and analysis of CAD performance should be done to assess: consistency with human reader interpretation, proportion of images read as abnormal and requiring further investigation, positivity rate among images read as abnormal.

Research priorities

- The development and evaluation of CAD software for automated detection of TB in children is urgently needed, as CXR is an important tool in detection of pulmonary TB in children and adolescents, given the difficulty in bacteriological testing and diagnosis.
- Further evidence on the performance of CAD software according to the characteristics of the individual being evaluated (eg smear status, HIV status, age cohort, history of previous TB, smoking status, gender) is needed, to allow for better setting-specific and patient-specific calibration of thresholds used for CAD.
- Further research on user, patient, and stakeholder perspectives on the use of CAD in place of human reader interpretation for CXR are needed to guide implementation and scale-up.

Table 6. Should C-Reactive Protein (CRP) using a cut-off of 5mg per litre, compared to the WHO-recommended 4 symptom screen, be used to screen for TB disease in people living with HIV?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	• People with HIV were 19 times more at risk of TB incidence than those without HIV in 2018 and a third of AIDS deaths were due to TB in 2018. Ensuring early detection and timely treatment of TB among people living with HIV is critical for reducing the mortality. • An estimated 44% of PLHIV with TB were not notified to have reached care in 2018. A systematic review of autopsy studies found as high as 64% of TB prevalence among people who had died from AIDS, in half of whom TB had been undetected prior to death. • Since 2011 WHO has recommended "Adults and adolescents living with HIV should be screened for TB with a clinical algorithm; those who report any one of the symptoms of current cough, fever, weight loss or night sweats may have active TB and should be evaluated for TB and other diseases" Screening with the WHO-recommended 4 symptom screen (W4SS) has been recommended at every health visit. • Designed primarily for ruling out active TB prior to the initiation of TB preventive treatment, this screen has a relatively high negative predictive value. However, evidence from symptom screening in certain sub-populations of PLHIV, namely PLHIV not on ART, with low CD4 cell counts and inpatients, the screen has low specificity, and among PLHIV on ART and among pregnant women, the screen has low sensitivity. • With the advancement of new tools such as digital X-ray, Xpert MTB/RIF, and new evidence on TB screening among PLHIV, a systematic literature review and individual patient data meta-analysis was commissioned to determine the accuracy of screening tests and approaches that lead to better outcomes in comparison to the WHO-recommended 4 symptom screen. • Recent studies have tested the accuracy of C-Reactive Protein (CRP) for screening for TB among people living with HIV and, as they met with the selection criteria for the systematic literature review, were included within the analysis.	

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	As part of the systematic review and individual patient data analysis, the accuracy of CRP with a cutoff of 5mg/L was assessed against the accuracy of the WHO-recommended 4 symptom screen, as a first screen for tuberculosis in people living with HIV, irrespective of ART status, compared to culture reference standard. **Test accuracy** A C-Reactive Protein (CRP) cutoff of 5mg/L Sensitivity: 0.90 (95% CI: 0.78 to 0.96) Specificity: 0.50 (95% CI: 0.29 to 0.71) A C-Reactive Protein (CRP) cutoff of 10mg/L Sensitivity: 0.83 (95% CI: 0.65 to 0.93) Specificity: 0.65 (95% CI: 0.43 to 0.83) WHO-recommended 4 symptom screen (current cough, weight loss, night sweats, fever) Sensitivity: 0.83 (95% CI: 0.74 to 0.89) Specificity: 0.38 (95% CI: 0.25 to 0.53) **Screening Test Accuracy for W4SS/CRP10 Combination** <u>Sequential: W4SS+ then CRP cutoff of 10mg/L</u> Sensitivity 0.75 (0.50-0.90)), specificity 0.77 (0.46-0.93) <u>Parallel: W4SS and CRP10 cutoff of 10mg/L</u> Sensitivity 0.92 (0.70-0.98), Specificity 0.24 (0.08-0.51) **TB prevalence WHO Symptom+ among CRP- TB among symptom+ CRP-** **in the population**	• With the reduction in cutoff the sensitivity increases but the specificity decreases. A cutoff of 5mg/L meets the criteria of the target product profile for a screening tool for sensitivity which is 90%, however specificity is reduced, although it remains higher than for the WHO-recommended 4 symptom screen. • The cutoff of >5mg/L was recommended

5% 62.9% 1.1%

10% 62.8% 2.4%

20% 62.5% 5.2%

Population	Test	Test accuracy	W4SS accuracy	Studies (persons)	Certainty in evidence	Lower prevalence cutoff	Middle prevalence cutoff	higher prevalence cutoff
All PLHIV (prev: 5%, 10%, 20%)	CRP (>=5 mg/L)	Se 0.90	Se 0.83	6 (3971)	Moderate	TP: 45/ FN:5	TP: 90/ FN:10	TP: 180/FN: 20
All PLHIV (prev: 5%, 10%, 20%)	CRP (>=5 mg/L)	Sp 0.50	Sp 0.38	6 (3971)	Low	TN: 475/ FP: 475	TN: 450/ FP:450	TN: 400/FP: 400
All PLHIV (prev: 5%, 10%, 20%)	CRP (>=10mg/L)	Se 0.83	Se 0.83	6 (3971)	Moderate	TP: 42/ FN:8	TP: 83/FN:17	TP: 166/FN:34
All PLHIV (prev: 5%, 10%, 20%)	CRP (>=10mg/L)	Sp 0.65	Sp 0.38	6 (3971)	Low	TN: 617/ FP: 333	TN: 585/FP:315	TN: 520/FP: 280
Outpatients not on ART (prev: 5%, 10%	CRP (>=5 mg/L)	Se: 0.89	Se 0.84	4 (3186)	High	TP: 45 / FN: 5	TP: 89/ FN: 11	TP: 178/ FN: 22
Outpatients not on ART (prev: 5%, 10%	CRP (>=5 mg/L)	Sp: 0.54	Sp 0.37	4 (1,694)	Moderate	TN:513 / FP: 437	TN: 486/ FP: 414	TN: 432/ FP: 368
Outpatients not on ART (prev: 5%, 10%	CRP (>=10mg/L)	Se: 0.82	Se 0.84	4 (3186)	Moderate	TP: 41/ FN: 9	TP: 82/ FN: 18	TP: 164/ FN: 36
Outpatients not on ART (prev: 5%, 10%	CRP (>=10mg/L)	Sp: 0.67	Sp 0.37	4 (3186)	Moderate	TN: 637/ FP: 313	TN: 603/ FP: 297	TN: 536/ FP: 504
CD4 ≤200 (prev: 5%, 10%, 20%)	CRP (>=5 mg/L)	Se: 0.93	Se 0.86	6 (1829)	High	TP: 47/ FN: 3	TP: 93/ FN: 7	TP: 186/ FN: 14
CD4 ≤200 (prev: 5%, 10%, 20%)	CRP (>=5 mg/L)	Sp: 0.40	Sp 0.30	6 (1,671)	Low	TN: 380/ FP: 570	TN: 360/ FP: 540	TN: 320/ FP: 480
CD4 ≤200 (prev: 5%, 10%, 20%)	CRP (>=10mg/L)	Se: 0.90	Se 0.86	6 (1829)	Moderate	TP: 45/ FN: 5	TP: 90/ FN: 10	TP: 180/ FN: 20
CD4 ≤200 (prev: 5%, 10%, 20%)	CRP (>=10mg/L)	Sp: 0.54	Sp 0.30	6 (1829)	Low	TN: 513/ FP: 437	TN: 486/ FP: 414	TN: 432/ FP: 368
Outpatients on ART (prev: 1%, 5%, 10%	CRP (>=5 mg/L)	Se: 0.4	Se 0.08	1 (381)	Low	TP: 4/ FN: 6	TP: 20/ FN: 30	TP: 40/ FN: 60
Outpatients on ART (prev: 1%, 5%, 10%	CRP (>=5 mg/L)	Sp: 0.8	Sp 0.88	1 (381)	High	TN: 792/ FP: 198	TN: 760/ FP: 190	TN: 720/ FP: 180
Outpatients on ART (prev: 1%, 5%, 10%	CRP (>=10mg/L)	Se: 0.2	Se 0.08	1 (381)	Moderate	TP: 2/ FN: 8	TP: 10/ FN: 40	TP: 20/ FN: 80
Outpatients on ART (prev: 1%, 5%, 10%	CRP (>=10mg/L)	Sp: 0.9	Sp 0.88	1 (381)	High	TN: 891/ FP: 99	TN: 855/ FP: 95	TN: 810/ FP: 90
Inpatients (prev: 10%, 20%, 30%)	CRP (>=5 mg/L)	Se: 0.98	Se 0.97	1 (399)	Moderate	TP: 98/FN:2	TP: 196/ FN: 4	TP: 294/ FN: 6
Inpatients (prev: 10%, 20%, 30%)	CRP (>=5 mg/L)	Sp: 0.12	Sp 0.10	1 (399)	Moderate	TN: 108/FP: 792	TN: 96/ FP: 704	TN: 84/ FP: 616
Inpatients (prev: 10%, 20%, 30%)	CRP (>=10mg/L)	Se: 0.97	Se 0.97	1 (399)	Moderate	TP: 97/FN:3	TP: 194/ FN: 6	TP: 291/ FN: 9
Inpatients (prev: 10%, 20%, 30%)	CRP (>=10mg/L)	Sp: 0.21	Sp 0.10	1 (399)	Moderate	TN:189/FP:711	TN: 168/FP: 632	TN: 147/ FP: 553

Table: All data for CRP as screening test in people living with HIV

Note that the GDG initially considered the data specific to outpatient PLHIV not on ART, and subsequently compared accuracy, benefits, and harms for all subpopulations of PLHIV.

as it is the lowest threshold indicating abnormality and to ensure clear guidance to programmes.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not already started treatment (true negative). The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in early initiation of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality **Summary for pretest probability of 5%** Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CRP with a cutoff of 5mg/L would correctly classify 39-48 people with TB as positive (true positive), 2-11 people with TB as negative (false negative), 275-675 people without TB as positive for TB (false positive), and 275-675 people without TB as negative (true negative). **Summary for pretest probability of 10%** Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CRP with a cutoff of 5mg/L would correctly classify 78-96 people with TB as positive (true positive), 4-22 people with TB as negative (false negative), 261-639 people without TB as positive for TB (false positive), and 261-639 people without TB as negative (true negative). **Summary for pretest probability of 20%** Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CRP with a cutoff of 5mg/L would correctly classify 156-192 people with TB as positive (true positive), 8-44 people with TB as negative (false negative), 232-568 people without TB as positive for TB (false positive), and 232-568 people without TB as negative (true negative).	

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not already started treatment (true negative). The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in early initiation of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality.	Only four true positives would be missed compared with 45 found in 5% prevalence, so the undesirable effect would be small.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%		Prevalence 10%		Prevalence 20%			
	a C-Reactive Protein (CRP) cutoff of 5mg per litre	WHO-recommended 4 symptom screen	a C-Reactive Protein (CRP) cutoff of 5mg per litre	WHO-recommended 4 symptom screen	a C-Reactive Protein (CRP) cutoff of 5mg per litre	WHO-recommended 4 symptom screen		
True positives patients with active TB	45 (39 to 48)	42 (37 to 45)	90 (78 to 96)	83 (74 to 89)	180 (156 to 192)	166 (148 to 178)	3971 (6)	⊕⊕⊕◯ MODERATE[a,b,c]
	3 more TP in a C-Reactive Protein (CRP) cutoff of 5mg per litre		7 more TP in a C-Reactive Protein (CRP) cutoff of 5mg per litre		14 more TP in a C-Reactive Protein (CRP) cutoff of 5mg per litre			
False negatives patients incorrectly classified as not having active TB	5 (2 to 11)	8 (5 to 13)	10 (4 to 22)	17 (11 to 26)	20 (8 to 44)	34 (22 to 52)		
	3 fewer FN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		7 fewer FN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		14 fewer FN in a C-Reactive Protein (CRP) cutoff of 5mg per litre			
True negatives patients without active TB	475 (275 to 675)	361 (238 to 503)	450 (261 to 639)	342 (225 to 477)	400 (232 to 568)	304 (200 to 424)	3971 (6)	⊕⊕◯◯ LOW[a,d,e]
	114 more TN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		108 more TN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		96 more TN in a C-Reactive Protein (CRP) cutoff of 5mg per litre			
False positives patients	475 (275 to 675)	589 (447 to 712)	450 (261 to 639)	558 (423 to 675)	400 (232 to 568)	496 (376 to 600)		

incorrectly classified as having active TB	114 fewer FP in a C-Reactive Protein (CRP) cutoff of 5mg per litre	108 fewer FP in a C-Reactive Protein (CRP) cutoff of 5mg per litre	96 fewer FP in a C-Reactive Protein (CRP) cutoff of 5mg per litre		

a. Low risk of bias in all but one study, in which included flow and timing was at high risk of bias with low risk in the other domains. We did not downgrade for serious risk of bias.
b. Sensitivity estimates ranged from 79% to 98% with overlapping CIs, except in one study which reported 40% sensitivity. The one study enrolled outpatients on ART. This could explain the variability. We did not downgrade for inconsistency.
c. We downgraded one level for serious imprecision. The CIs around true positives and false negatives may lead to different decisions depending on which credible limits are assumed.
d. We downgraded one-level for serious inconsistency. Specificity estimates ranged from 44% to 63 % in four studies in outpatients not on ART with non-overlapping CIs. We could not explain the variability. One study in inpatients reported 12% specificity, while another study in outpatients on ART reported 79% specificity.
e. We downgraded one level for imprecision. The wide CI around true negatives and false positive that may lead to different decisions depending on which limits are assumed.

Summary for pretest probability of 5%

Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CRP with a cutoff of 5mg/L would correctly classify 39-48 people with TB as positive (true positive), 2-11 people with TB as negative (false negative), 275-675 people without TB as positive for TB (false positive), and 275-675 people without TB as negative (true negative).

Summary for pretest probability of 10%

Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CRP with a cutoff of 5mg/L would correctly classify 78-96 people with TB as positive (true positive), 4-22 people with TB as negative (false negative), 261-639 people without TB as positive for TB (false positive), and 261-639 people without TB as negative (true negative).

Summary for pretest probability of 20%

Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CRP with a cutoff of 5mg/L would correctly classify 156-192 people with TB as positive (true positive), 8-44 people with TB as negative (false negative), 232-568 people without TB as positive for TB (false positive), and 232-568 people without TB as negative (true negative).

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 systematic review of the performance of CRP with cutoff of 5mg/l to screen for TB among people living with HIV, regardless of ART status included 6 studies with a total of 3971 participants. The prevalences of TB in PLHIV in the 6 included studies from Kenya: 1% (n=385), South Africa: 10% (n=425), South Africa: 12% (n=734), Uganda: 15% (n=1525), 17% South Africa: (n=502) and South Africa: 26% (n=400). The same results may not apply in lower-prevalence settings. Certainty of evidence for true positives and false negatives was moderate, and for true negatives and false positives the evidence was low. Evidence certainty was downgraded for reasons of imprecision and inconsistency. Reasons for downgrading are further explained in the footnotes listed in the evidence table.	

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%		Prevalence 10%		Prevalence 20%			
	a C-Reactive Protein (CRP) cutoff of 5mg per litre	WHO-recommended 4 symptom screen	a C-Reactive Protein (CRP) cutoff of 5mg per litre	WHO-recommended 4 symptom screen	a C-Reactive Protein (CRP) cutoff of 5mg per litre	WHO-recommended 4 symptom screen		
True positives patients with active TB	45 (39 to 48)	42 (37 to 45)	90 (78 to 96)	83 (74 to 89)	180 (156 to 192)	166 (148 to 178)	3971 (6)	⊕⊕⊕◯ MODERATE[a,b,c,d]
	3 more TP in a C-Reactive Protein (CRP) cutoff of 5mg per litre		7 more TP in a C-Reactive Protein (CRP) cutoff of 5mg per litre		14 more TP in a C-Reactive Protein (CRP) cutoff of 5mg per litre			
False negatives patients incorrectly classified as not having active TB	5 (2 to 11)	8 (5 to 13)	10 (4 to 22)	17 (11 to 26)	20 (8 to 44)	34 (22 to 52)		
	3 fewer FN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		7 fewer FN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		14 fewer FN in a C-Reactive Protein (CRP) cutoff of 5mg per litre			
True negatives patients without active TB	475 (275 to 675)	361 (238 to 503)	450 (261 to 639)	342 (225 to 477)	400 (232 to 568)	304 (200 to 424)	3971 (6)	⊕⊕◯◯ LOW[a,b,e,f]
	114 more TN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		108 more TN in a C-Reactive Protein (CRP) cutoff of 5mg per litre		96 more TN in a C-Reactive Protein (CRP) cutoff of 5mg per litre			
False positives patients incorrectly classified as having active TB	475 (275 to 675)	589 (447 to 712)	450 (261 to 639)	558 (423 to 675)	400 (232 to 568)	496 (376 to 600)		
	114 fewer FP in a C-Reactive Protein (CRP) cutoff of 5mg per litre		108 fewer FP in a C-Reactive Protein (CRP) cutoff of 5mg per litre		96 fewer FP in a C-Reactive Protein (CRP) cutoff of 5mg per litre			

a. Low risk of bias in all but one study, in which included flow and timing was at high risk of bias with low risk in the other domains. We did not downgrade.
b. Four studies had low concern for applicability. One study had high concern for the patient selection domain since the study included only inpatients. Another study had high concern for the patient selection domain since the study included only individuals with CD4 cell count ≤350 per µL.. However, the sensitivity and specificity estimates were comparable to others. We did not downgrade
c. Sensitivity estimates ranged from 79% to 98% with overlapping CIs, except in one study which reported 40% sensitivity. The one study enrolled outpatients on ART. This could explain the variability. We did not downgrade.
d. The CIs around true positives and false negatives may lead to different decisions depending on which credible limits are assumed. We downgraded one level.
e. Specificity estimates ranged from 44% to 63 % in four studies in outpatients not on ART with non-overlapping CIs. We could not explain the variability. One study in inpatients

	reported 12% specificity, while another study in outpatients on ART reported 79% specificity. We downgraded one-level. f. The wide CI around true negatives and false positive that may lead to different decisions depending on which limits are assumed. We downgraded one level.	

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 research evidence identified. No direct benefits, adverse effects or burden resulting from the test itself are anticipated. CRP detects general inflammation. An indirect benefit to consider is its ability to detect other morbidities such as diabetes, high cholesterol, high blood pressure, obesity (BMI 30+), smoking that might need addressing, some of which are also independent risk factors for TB.	● Compared to the W4SS, CRP is considered more invasive than asking questions as it requires a finger prick test or blood needs to be drawn, depending on the CRP test.

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	The next step in the diagnostic cascade after screening is WHO recommended molecular diagnostics. While Xpert® MTB/RIF can achieve a pooled sensitivity of 81% in PLHIV, the Ultra assay has a 7% increase in sensitivity. **(Strong recommendation, high certainty in the evidence)** LF-LAM is also recommended for PLHIV. Unlike other diagnostic tests, the sensitivity of this test increases with advancement of HIV disease. LF-LAM can achieve a pooled sensitivity of 62% in inpatient settings. **(Strong recommendation; moderate certainty in the evidence about the intervention effects).** TB treatment is effective among PLHIV, and successful treatment outcomes comparable to those among people not living with HIV can be achieved (Owiti et al). **(Strong recommendation, high certainty in the estimates of effect)** Conversely, undetected TB and delays in diagnosis of TB and MDR-TB diagnosis among people living with HIV are associated with increased mortality as well as increased transmission. Findings from a systematic review (Gupta et al) of post mortem studies among people who had died from AIDS, highlighted close to 50% of post mortem diagnosis of TB had not been detected prior to death, and TB was the cause of death in more 91.4% (95% CI 85.8-97.0%) of all the TB cases. Should a diagnostic test not be indicated, programmatic management of latent TB infection and initiation of antiretroviral treatment would be the next course of action if eligible. A systematic review of 12 randomized controlled trials found that TB preventive treatment reduced the overall risk of TB by 33% (RR 0.67, 95% CI 0.51-0.87) as well as a reduction in mortality. Clinical trials confirm that early use of ART keeps people living with HIV alive and healthier and reduces the risk of transmitting the virus to others. Earlier treatment has the further advantage of simplifying the operational demands on programmes.	● Compared with the W4SS CRP is a biomedical test, the results of which health staff and patients might be more motivated to undergo follow-up with a confirmatory diagnostic test.

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	No research evidence identified. Similar to the W4SS, the time from test to test result for some CRP tests is 3-5 minutes, enabling a clinician, equipped with a point-of-care test to decide immediately whether a person is eligible or not for diagnostic confirmation, or for another intervention. Otherwise, a sample would need to be sent to the nearest facility where CRP is available.	

	CRP can be used in primary health care settings although current distribution of tests is likely to be more centralised. Access to and linkage between a CRP test and diagnostic confirmation and subsequent treatment initiation is also dependent upon the distribution and colocation of respective services in relation to one another within the given health system, as well as location of where the person living with HIV first presents.	

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	No research evidence identified.	

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 identified. Screening aims to identify people with active TB earlier and therefore ensure earlier treatment and better health outcomes for individuals and lower TB transmission to community. No direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is no important uncertainty in what we know about how patients perceive and value the outcomes explored above from research evidence.	

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 interventio	The panel determines that the balance between desirable and undesirable effects favours **using** CRP with a cutoff of 5mg/L as an option to replace the WHO-recommended Four Symptom Screen for screening for TB among people living with HIV.	It was felt that the balance of effects was equivalent to the W4SS in terms of accuracy - although the use of CRP alone might save some costs due to higher specificity and fewer false positives.

n ○ Favors the interventio n ○ 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 included studies directly addressed this question. ● Indirect evidence: Indirect evidence was available from two studies, conducted among PLHIV in Uganda, using CRP as a triage test but did not compare with WHO 4SS. Only test costs and no programmatic costs were included. CRP unit costs ranged from US$2-$6 per test (Murray 2016, Yoon 2019). Murray et al. estimated CRP as a screening test with Xpert for diagnosis cost on average US$24.30 per person screened for test costs only (no health system costs included). CRP unit test costs represent a small proportion of the total implementation costs for an ICF program. ● In a setting with TB prevalence of 10%, if 1000 PLHIV are screened, the WHO4SS should lead to 641 mWRD tests. CRP(10mg) would lead to 398 Xpert tests and CRP (5mg) to 540 Xpert tests.	● The CRP test costs more than the symptom screen, but the increased specificity of CRP would result in fewer people need a confirmatory test which would result in moderate savings

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	The lack of direct evidence on the use of CRP *(cutoff 5 mg/L)* for screening is a limitation. Two studies provided indirect evidence of CRP as a triage test*(cutoff 10 mg/L)*, both of high quality.	

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 intervention ○ Varies ○ No included studies	The use of CRP (cutoff 10 mg/L) was determined to be cost-effective in two studies. Murray et al. found CRP as a triage test followed by diagnosis with Xpert to be cost-effective with an ICER of US$588/year of life gained (UR: US$221-$1746). Yoon et al. found point-of-care (POC) CRP for screening followed by the use of Xpert for diagnosis to be cost-effective at US$70 per additional patient diagnosed. Major drivers of cost were determined to be the inclusion of lifetime cost of ART, TB prevalence and types and cost of additional tests performed (i.e. TB-LAM, culture, etc.). No studies were found on repeated periodicity.	• In light of the evidence and in line with resources required, it was suggested that the cost effectiveness would probably favour the intervention. • It was noted that the quality and delivery of the symptom screen is not always consistent which may result in more undetected TB which would have implications for reduced cost effectiveness for the symptom screen in comparison to CRP. • Periodicity of screening with CRP needs to be considered together with the potential cost implications for screening among PLHIV on ART.

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 identified. CRP is available as a cheap, point of care diagnostic finger-prick test but is not yet decentralised in lower- and middle-income countries thus access to this test for people living outside of district/national centres would currently be limited, when compared with the readily available W4SS. Increased accuracy in a screen facilitates earlier detection of a debilitating impoverishing disease, whilst reducing the burden and costs of unnecessary diagnostic follow-up procedures among false positives. CRP is a quantitative test. Compared with the W4SS, the CRP result and interpretation might be less influenced by quality of delivery of the intervention, personal judgement, response bias, discrimination or stigma.	• Compared with the symptom screen, CRP is a lab test which is not currently uniformly available and may decrease health equity due to poor access. • Eventually CRP testing might become available in HIV settings but there will be some degree of non-availability at programme level at present. • If cost is a barrier, it may

		differ by setting and impact on equity. ● In some settings it is readily available everywhere and in which case equity would not be affected.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	To date the systematic review team have found very little qualitative evidence relevant for this group but drawing upon the theoretical findings from other wider body of evidence, they would suggest that perceived and enacted HIV and TB stigma will complicate the decision-making of persons living with HIV. 1 study, Tuot 2019 Cambodia Total respondents N = 120; PLHIV, N = 6 Quality assessment not yet done. One literature review found that at-risk individuals report that fear of TB stigma and the social and economic impact of stigma affects their willingness to undergo TB screening - Tuberculosis and stigmatization: pathways and Interventions Public Health Rep. 2010; 125(Suppl 4): 34 42.doi: 10.1177/00333549101250S407PMCID: PMC2882973PMID: 20626191	

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	CRP testing is available as a simple point of care test and can be used in primary health settings. A study in UK found that use of CRP within primary healthcare proved feasible for use by general practitioners and practice nurses. Reliability and feasibility of a near patient test for C-reactive protein in primary care, Clinical Trial Br J Gen Pract. 1996 Jul;46(408):395-400.	In Zambia, under ZAMBART/TREATS lay persons were trained to use CRP.

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

	JUDGEMENT						
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
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

C-reactive protein using a cutoff of >5mg/L may be used to screen for TB disease among adults and adolescents living with HIV (conditional recommendation, low certainty of the evidence for test accuracy)

Justification

- Data on accuracy of C-Reactive Protein (CRP) using a cutoff value of >5mg/L and of >10mg/L as an indication of TB disease were reviewed by the guideline group and both were considered to have similar or superior accuracy when compared with the WHO 4 symptom screen. The cutoff of >5mg/L was recommended as it is the lowest threshold indicating abnormality in many clinical settings. The choice of cutoff will also be dependent on the CRP technology available in the given setting.
- 3,909 of study participants reviewed for CRP were adults (>19) and 62 were adolescents (10-19).
- CRP is a test that detects general inflammation in the body. Data from very high TB prevalence settings (average 14% prevalence among study participants across six studies from Kenya, South Africa and Uganda) were reviewed for CRP's accuracy in detecting TB and it should be noted that the specificity and predictive value of the test for detecting inflammation due specifically to TB would be reduced in lower TB prevalence settings.

Subgroup considerations

- The individual patient data meta-analysis on CRP reported similar sensitivity, and higher or similar specificity to the W4SS in all sub-populations assessed. CRP was found to be most accurate among outpatients not on ART, (CRP with a cutoff of 5mg/L: Sensitivity 0.89 (95%CI: 0.85-0.92) Specificity: 0.54 (95%CI 0.45-0.62) and CRP with a cutoff of 10mg/L: Sensitivity 0.82 (95% CI: 0.79-0.86) Specificity: 0.67 (95% CI: 0.60-0.74)), compared with the symptom screen (Sensitivity: 0.84 (95% CI: 0.75-0.90) Specificity: 0.37 (95% CI: 0.25 to 0.50). When performed sequentially whereby a positive W4SS screen is followed by CRP, CRP with a cutoff of 5mg/L was found to be as sensitive as the W4SS (0.78 (95% CI: 0.70-0.85) with significantly higher specificity (0.73 (95% CI: 0.66-0.79) in this subpopulation.
- Specificity of CRP for screening for TB among inpatients was found to be low (CRP with a cutoff of 5mg/L: 0.12 (0.09-0.17) CRP with a cutoff of 10mg/L: 0.21 (0.17-0.26)), similar to the W4SS (0.11 (0.8-0.14)), likely due to competing comorbidities that would also result in raised CRP levels and the presence of symptoms, so alternative TB screening strategies are advised for screening for TB among this sub-population.
- CRP, combined with the W4SS, was found to have similar sensitivity and specificity to the W4SS for all populations.
- For pregnant women data were limited to 2 studies of 62 participants. Sensitivity of CRP with a cutoff of 5mg/L had 0.70 (0.12-0.97) sensitivity and 0.41 (0.12-0.78) specificity. CRP with a cutoff of 10mg/L had a 0.70 (0.12-0.97) sensitivity and specificity was 0.54 (0.18-0.86) compared with the symptom screen: Sensitivity 0.61 (0.39-0.79) and specificity 0.58 (0.39-0.75).

Implementation considerations

- Countries should position CRP within national TB screening algorithms according to feasibility, level of health facility, resources available and equity.
- The W4SS should be conducted at every encounter with a healthcare worker, as part of a comprehensive clinical evaluation, to inform the need for increased infection control and for other investigations such as LF-LAM.
- For people who are newly diagnosed and not yet on ART, CRP may be conducted as part of the initial package of HIV-related investigations.
- Although data regarding frequency were not available, the GDG suggested that more intensified screening in addition to the W4SS with screening tools such as CRP, CXR or mWRD could be conducted annually at the time of viral load testing or other investigations as this may be most pragmatic approach.
- Consideration should also be given to the role of CRP in ruling out active TB prior to the initiation of TB preventive treatment. CRP has a negative predictive value of 99.8% among outpatients not on ART in a 1% TB prevalence setting, compared with the W4SS which has a negative predictive value of 99.6% in the same setting.
- For increased sensitivity, a sequential test whereby PLHIV who are symptom screen positive would be referred for diagnostic confirmation and only those who are symptom screen negative would receive a CRP test and if found positive referred for diagnostic confirmation would be more economical than parallel screening when PLHIV found to be positive for the screen or CRP would be referred for diagnostic confirmation.

Monitoring and evaluation

- Countries are encouraged to monitor and evaluate the yield of TB screening among people living with HIV, disaggregated by screening tools to inform programming and resource planning.

Research priorities

- Well-designed clinical trials to strengthen the evidence on the accuracy, effectiveness (including impact on patient-important outcomes e.g. mortality), feasibility and cost implications of using CRP to screen for TB across all HIV sub-populations in low, medium and high HIV and TB burden settings with and without high ART coverage, compared with other screening tools.
- Subpopulations of PLHIV for further investigation would include but not be limited to inpatients, acute care service attendees, patients failing ART, newly diagnosed HIV patients enrolling in ART clinics, stable patients established on ART, pregnant women, children and adolescents living with HIV and key populations such as people who use drugs.
- Research to evaluate the effectiveness and accuracy of combining, in different sequence, the W4SS, CRP, CXR, mWRD and LF-LAM in screening and diagnostic algorithms.
- Research to evaluate the accuracy and predictive value of CRP along the cutoff-point spectrum above any cutoff level higher than 5 mg/L for TB screening, either alone or in combination with other screening tests, among PLHIV in different TB prevalent settings.
- Frequency of screening - effectiveness, cost effectiveness, feasibility and acceptability and optimal periodicity of routine regular screening with CRP among PLHIV on ART.

Table 7. Should chest X-ray (any abnormality) or WHO-recommended 4 symptom screen vs. WHO-recommended 4 symptom screen alone be used to screen for TB disease in people living with HIV?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	• People with HIV were 19 times more at risk of TB incidence than those without HIV in 2018 and a third of AIDS deaths were due to TB in 2018. Ensuring early detection and timely treatment of TB among people living with HIV is critical for reducing the mortality. • An estimated 44% of PLHIV with TB were not notified to have reached care in 2018. A systematic review of autopsy studies found as high as 64% of TB prevalence among people who had died from AIDS, in half of whom TB had been undetected prior to death. • Since 2011 WHO has recommended "Adults and adolescents living with HIV should be screened for TB with a clinical algorithm; those who report any one of the symptoms of current cough, fever, weight loss or night sweats may have active TB and should be evaluated for TB and other diseases" Screening with the WHO-recommended 4 symptom screen (W4SS) has been recommended at every health visit. • Designed primarily for ruling out active TB prior to the initiation of TB preventive treatment, this screen has a relatively high negative predictive value. However, evidence from symptom screening in certain sub-populations of PLHIV, namely PLHIV not on ART, with low CD4 cell counts and inpatients, the screen has low specificity, and among PLHIV on ART and among pregnant women, the screen has low sensitivity. • With the advancement of new tools such as digital chest X-ray (CXR), Xpert MTB/RIF, and new evidence on TB screening among PLHIV, a systematic literature review and individual patient data meta-analysis was commissioned to determine the accuracy of screening tests and approaches that lead to better outcomes in comparison to the W4SS. • As part of the development of the WHO 2018 Updated and consolidated guidelines for programmatic management of latent TB infection, the role of CXR was reviewed and was identified for more accurate rule-out of active TB when combined with the W4SS, prior to initiation of TB preventive treatment in people living with HIV who are on ART. Furthermore, recent prevalence surveys (Modi et al 2016) have been reported to have found that screening with CXR was more accurate for detecting TB than screening with symptoms. • With increased access to digital CXR and given that studies exploring the use of CXR in TB screening among people living with HIV met the selection criteria for the systematic literature review, CXR, (any abnormality and suggestive of TB) was included within the analysis, as a standalone strategy and in combination with the symptom screen.	Data from WHIP3TB trial from 3 countries with over 4000 individuals, presented at CROI2020 were shared with the group. The data demonstrated that over a two year period of trial follow-up among people living with HIV who were on ART and TB preventive treatment that there was a slow accrual of TB cases over the trial follow-up that were detected due to symptom screening or people self-reporting symptoms. However at months 12 and 24 when more intensified TB case-finding was carried out with CXR and sputum culture there was a substantial increase in TB notifications, particularly in month 12. 80% of the cases detected by the intensified screening strategy were asymptomatic. Results from this trial demonstrated that the symptom screen misses a high proportion of TB cases among PLHIV on ART and suggests the need for regular scheduled more intensified screening in addition to the W4SS.

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	As part of the systematic review and individual patient data meta-analysis, the accuracy of **any abnormality detected by CXR alone**, by CXR together with the four-symptom screen as a **parallel screening strategy (positive screen for either W4SS or CXR leads to a diagnostic test)** and a **sequential strategy (positive screen for W4SS then CXR)** were assessed for detecting tuberculosis against the accuracy of the four-symptom screen alone, compared to culture reference standard, among outpatients living with HIV on ART. The accuracy of **CXR alone, with abnormality suggestive of TB** was also assessed. The accuracy of CXR was then reviewed for all people living with HIV, and for identified sub-populations.	The parallel screening strategy of W4SS combined with CXR was assessed first among outpatients living with HIV on ART, compared to the W4SS. The GDG then reviewed the accuracy of the CXR

Test accuracy of the WHO-recommended Four Symptom Screen

Sensitivity: 0.83 (95% CI: 0.74 to 0.89) Specificity: 0.38 (95% CI: 0.25 to 0.53)

Test accuracy of chest radiograph alone (any abnormality) Sensitivity: 0.71 (95% CI: 0.64 to 0.78) Specificity: 0.61 (95% CI: 0.50 to 0.71)

Test accuracy of chest radiography alone (suggestive of TB) Sensitivity: 0.63 (95% CI: 0.56 to 0.70) Specificity: 0.78 (95% CI: 0.66 to 0.86)

Test accuracy of chest radiography (any abnormality) or WHO four symptom screen (parallel) Sensitivity: 0.93 (95% CI: 0.88 to 0.96) Specificity: 0.20 (95% CI: 0.10 to 0.38)

Test accuracy of chest radiography (any abnormality) if screened positive for WHO four symptom screen (sequential) Sensitivity: 0.63 (95% CI: 0.54 to 0.71) Specificity: 0.73 (95% CI: 0.61 to 0.82)

Population	Test	Test accuracy	W4SS accuracy	(persons)	evidence	Lower prevalence cutoff	Middle prevalence cutoff	higher prevalence cutoff
All PLHIV (prev: 5%, 10%, 20%)	CXR any abnorm	Se: 0.71	Se 0.83	8 (6247)	Moderate	TP: 71/ FN:29	TP: 142/ FN:58	TP: 213/FN: 87
All PLHIV (prev: 5%, 10%, 20%)	CXR any abnorm	Sp: 0.61	Sp 0.38	8 (6247)	Moderate	TN: 549/ FP: 351	TN: 488/ FP:312	TN: 427/FP: 273
Outpatients not on ART (prev: 5%, 10%,	CXR any abnorm	Se: 0.70	Se 0.84	8 (3525)	Moderate	TP: 35 / FN: 15	TP: 70/ FN: 30	TP: 140/ FN: 60
Outpatients not on ART (prev: 5%, 10%,	CXR any abnorm	Sp: 0.62	Sp 0.37	8 (3525)	Moderate	TN:589 / FP: 361	TN: 558/ FP: 342	TN: 496/ FP: 304
CD4 ≤200 (prev: 5%, 10%, 20%)	CXR any abnorm	Se: 0.73	Se 0.86	8 (2234)	Moderate	TP: 37/FN:13	TP: 73/ FN: 27	TP: 146/ FN:54
CD4 ≤200 (prev: 5%, 10%, 20%)	CXR any abnorm	Sp: 0.57	Sp 0.30	8 (2234)	Moderate	TN: 542/FP:408	TN: 513/ FP: 387	TN: 456/ FP: 344
Outpatients on ART (prev: 1%, 5%, 10%)	CXR any abnorm	Se: 0.7	Se 0.53	4 (2670)	Moderate	TP: 7/ FN: 3	TP: 35/ FN: 15	TP: 70/ FN: 30
Outpatients on ART (prev: 1%, 5%, 10%)	CXR any abnorm	Sp: 0.63	Sp 0.70	4 (2670)	Low	TN: 624/ FP: 366	TN: 598/ FP: 352	TN: 567/ FP: 333
Outpatients on ART (prev: 1%, 5%, 10%)	**CXR any or WHO4SS**	**Se: 0.85**	**Se 0.53**	**5 (240)**	**Moderate**	**TP: 9/ FN: 1**	**TP: 43/ FN: 7**	**TP: 85/ FN: 15**
Outpatients on ART (prev: 1%, 5%, 10%)	**CXR any or WHO4SS**	**Sp: 0.33**	**Sp 0.70**	**5 (690)**	**Low**	**TN: 327/ FP: 663**	**TN: 314/ FP: 636**	**TN: 297/ FP: 603**
Inpatients (prev: 10%, 20%, 30%)	CXR any abnorm	Se: 0.75	Se 0.90	1 (57)	Very low	TP: 75/FN: 25	TP: 150/ FN: 50	TP: 225/ FN: 75
Inpatients (prev: 10%, 20%, 30%)	CXR any abnorm	Sp: 0.44	Sp 0.17	1 (57)	Low	TN: 396/FP: 504	TN: 352/FP: 448	TN: 308/ FP: 392

Table: **All data for CXR as screening test in people living with HIV**

- Note that the GDG initially considered the data specific to outpatient PLHIV on ART, and subsequently compared accuracy, benefits, and harms for all subpopulations of PLHIV

across all PLHIV and among sub-populations compared with the W4SS and agreed that there were no significant differences in accuracy of the tool between the different populations.

With sensitivity prioritised, parallel screening with the W4SS, equivalent to providing an x-ray for all people who screen negative with the W4SS was more accurate than W4SS.

It was acknowledged that with increased sensitivity, specificity is however, compromised.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not yet started treatment (true negative). The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in start of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality. **Summary for pretest probability of 5%** Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CXR (any abnormality) and the W4SS would correctly classify 44-48 people with TB as positive (true positive), 2-6 people with TB as negative (false negative), 589-855 people without TB as positive for TB (false positive), and 95-361 people without TB as negative (true negative). **Summary for pretest probability of 10%** Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CXR (any abnormality) and the W4SS would correctly classify 88-96 people with TB as positive (true positive), 4-12 people with TB as negative (false negative), 558-810 people without TB as positive for TB (false positive), and 90-342 people without TB as negative (true negative). **Summary for pretest probability of 20%** Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among people living with HIV), in a hypothetical population of 1,000 people living	● The increase in true positives was considered to be a moderate desirable effect. ● It was acknowledged that in low TB prevalence settings the number of true positives would be reduced.

with HIV being screened, CXR (any abnormality) and the W4SS would correctly classify 176-192 people with TB as positive (true positive), 8-24 people with TB as negative (false negative), 496-720) people without TB as positive for TB (false positive), and 80-304 people without TB as negative (true negative).

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 5%		Effect per 1000 patients/year for pre-test probability of 10%		Effect per 1000 patients/year for pre-test probability of 20%		Importance
			chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone	
True positives	cross-sectional (cohort type accuracy study)	⨁⨁⨁⨁ HIGH[a,b,c]	47 (44 to 48)	42 (37 to 45)	93 (88 to 96)	83 (74 to 89)	186 (176 to 192)	166 (148 to 178)	
TP absolute difference			5 more TP in chest X-ray (any abnormality) or WHO-4 symptom screen		10 more TP in chest X-ray (any abnormality) or WHO-4 symptom screen		20 more TP in chest X-ray (any abnormality) or WHO-4 symptom screen		
False negatives			3 (2 to 6)	8 (5 to 13)	7 (4 to 12)	17 (11 to 26)	14 (8 to 24)	34 (22 to 52)	
FN absolute difference			5 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen		10 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen		20 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen		
True negatives	cross-sectional (cohort type accuracy study)	⨁⨁◯◯ LOW[a,b,d,e]	190 (95 to 361)	361 (238 to 503)	180 (90 to 342)	342 (225 to 477)	160 (80 to 304)	304 (200 to 424)	
TN absolute difference			171 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen		162 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen		144 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen		
False positives			760 (589 to 855)	589 (447 to 712)	720 (558 to 810)	558 (423 to 675)	640 (496 to 720)	496 (376 to 600)	
FP absolute difference			171 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen		162 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen		144 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen		

a. Low risk of bias in all included studies. We did not downgrade
b. Low concern about applicability in all but one study that included only people with advanced HIV disease and another study that included ~10% that were inpatients. We did not downgrade for indirectness.
c. The CIs for sensitivity is narrow. The lower limit is higher than the point estimate and lower limit of the WHO screen and similar to the upper limit. The CIs would likely not

lead to different decisions depending on which credible limits are assumed. We did not downgrade.

d. Specificity estimates ranged from 2% to 60% with non overlapping CIs. We downgraded one level.

e. The wide CI around true negatives and false positives may lead to different decisions depending on which limits are assumed. We downgraded one level

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not yet started treatment (true negative). The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in start of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality. **Summary for pretest probability of 5%** Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CXR (any abnormality) and the W4SS would correctly classify 44-48 people with TB as positive (true positive), 2-6 people with TB as negative (false negative), 589-855 people without TB as positive for TB (false positive), and 95-361 people without TB as negative (true negative). **Summary for pretest probability of 10%** Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CXR (any abnormality) and the W4SS would correctly classify 88-96 people with TB as positive (true positive), 4-12 people with TB as negative (false negative), 558-810 people without TB as positive for TB (false positive), and 90-342 people without TB as negative (true negative). **Summary for pretest probability of 20%** Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, CXR (any abnormality) and the W4SS would correctly classify 176-192 people with TB as positive (true positive), 8-24 people with TB as negative (false negative), 496-720) people without TB as positive for TB (false positive), and 80-304 people without TB as negative (true negative).	• Whilst it was observed that there were a high number of false positives with this screening strategy, it was judged that this would not be a seriously undesirable effect on the patients as they would undergo a confirmatory test with mWRD. However, this would come at a cost to the health system for more mWRD tests, and depending on the setting, it might come at extra cost and unnecessary time for the PLHIV.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%		Prevalence 10%		Prevalence 20%			
	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone		
True positives patients with active TB	47 (44 to 48)	42 (37 to 45)	93 (88 to 96)	83 (74 to 89)	186 (176 to 192)	166 (148 to 178)	6238 (8)	⨁⨁⨁⨁ HIGH[a,b,c]
	5 more TP in chest X-ray (any abnormality) or		10 more TP in chest X-ray (any abnormality) or		20 more TP in chest X-ray (any abnormality) or			

	WHO-4 symptom screen		WHO-4 symptom screen		WHO-4 symptom screen			
False negatives patients incorrectly classified as not having active TB	3 (2 to 6)	8 (5 to 13)	7 (4 to 12)	17 (11 to 26)	14 (8 to 24)	34 (22 to 52)		
	5 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen		10 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen		20 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen			
True negatives patients without active TB	190 (95 to 361)	361 (238 to 503)	180 (90 to 342)	342 (225 to 477)	160 (80 to 304)	304 (200 to 424)	6238 (8)	⊕⊕◯◯ LOW[a,b,d,e]
	171 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen		162 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen		144 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen			
False positives patients incorrectly classified as having active TB	760 (589 to 855)	589 (447 to 712)	720 (558 to 810)	558 (423 to 675)	640 (496 to 720)	496 (376 to 600)		
	171 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen		162 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen		144 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen			

a. Low risk of bias in all included studies. We did not downgrade.
b. Low concern about applicability in all but one study that included only people with advanced HIV disease and another study that included ~10% that were inpatients. We did not downgrade for indirectness.
c. The confidence intervals for sensitivity are narrow. The lower limit is higher than the point estimate and lower limit of the WHO screen and similar to the upper limit. The confidence interval would likely not lead to different decisions depending on which credible limits are assumed. We did not downgrade for imprecision.
d. We downgraded one level for inconsistency. Specificity estimates ranged from 2% to 60% with non overlapping confidence intervals.
e. We downgraded one level for imprecision. The wide confidence interval around true negatives and false positives may lead to different decisions depending on which limits are assumed.

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 systematic review of the performance of chest X-ray (Any abnormality) and the WHO four symptom screen for parallel screening for TB among people living with HIV included 8 studies with a total of 6238 participants. The average prevalence of TB in all PLHIV in the 8 included studies was 7%, ranging from 3% to 18%. Studies were based in the following countries: Benin, Botswana, Brazil, Guinea, India, Kenya, Malawi, Myanmar, Peru, South Africa, and Zimbabwe. Overall, the certainty of evidence is low reflecting the lowest level of certainty for the GRADE evidence for accuracy. Certainty of evidence was high for sensitivity and low for specificity. The reasons for downgrading included imprecision around the estimates for specificity and unexplained inconsistency.	

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%		Prevalence 10%		Prevalence 20%			
	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone	chest X-ray (any abnormality) or WHO-4 symptom screen	WHO-4 symptom screen alone		
True positives patients with active TB	47 (44 to 48)	42 (37 to 45)	93 (88 to 96)	83 (74 to 89)	186 (176 to 192)	166 (148 to 178)	6238 (8)	⊕⊕⊕⊕ HIGH[a,b,c]
	5 more TP in chest X-ray (any abnormality) or WHO-4 symptom screen		10 more TP in chest X-ray (any abnormality) or WHO-4 symptom screen		20 more TP in chest X-ray (any abnormality) or WHO-4 symptom screen			
False negatives patients incorrectly classified as not having active TB	3 (2 to 6)	8 (5 to 13)	7 (4 to 12)	17 (11 to 26)	14 (8 to 24)	34 (22 to 52)		
	5 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen		10 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen		20 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen			
True negatives patients without active TB	190 (95 to 361)	361 (238 to 503)	180 (90 to 342)	342 (225 to 477)	160 (80 to 304)	304 (200 to 424)	6238 (8)	⊕⊕◯◯ LOW[a,b,d,e]
	171 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen		162 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen		144 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen			
False positives patients incorrectly classified as having active TB	760 (589 to 855)	589 (447 to 712)	720 (558 to 810)	558 (423 to 675)	640 (496 to 720)	496 (376 to 600)		
	171 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen		162 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen		144 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen			

a. Low risk of bias in all included studies. We did not downgrade.
b. Low concern about applicability in all but one study that included only people with advanced HIV disease and another study that included ~10% that were inpatients. We did not downgrade for indirectness.
c. The confidence intervals for sensitivity are narrow. The lower limit is higher than the point estimate and lower limit of the WHO screen and similar to the upper limit. The confidence interval would likely not lead to different decisions depending on which credible limits are assumed. We did not downgrade for imprecision.
d. We downgraded one level for inconsistency. Specificity estimates ranged from 2% to 60% with non overlapping confidence intervals.
e. We downgraded one level for imprecision. The wide confidence interval around true negatives and false positives may lead to different decisions depending on which limits are assumed.

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	Direct CXR is a safe technology using a radiation dose of 0.1 mSv, which corresponds to 1/30 of the average annual radiation dose from the environment (3 mSv) and 1/10 of the annual accepted dose of ionizing radiation for the general public (1 mSv). As a point of reference, the radiation dose of one CXR is equivalent to or less than the radiation exposure received during return travel on an intercontinental flight. Therefore, exposure to the low radiation doses delivered to patients during a CXR poses a small risk of inducing tissue reactions or cancer in the years to decades following the examination. However, it should be noted that a linear non-threshold relationship is assumed between radiation exposure and the risks of effects of this nature. Based on this linear model, the probability of developing cancer is presumed to increase even following exposure to low doses of radiation, although the increase in risk is extremely small. Even though the individual risk associated with radiation exposure from CXR is low, when a large number of individuals are exposed, the associated risks may still constitute a public health issue (*Chest radiography in tuberculosis detection, WHO 2016).* CXR also detects other lung diseases. An indirect benefit to consider is its ability to detect also other morbidities other than tuberculosis that might need addressing, as part of a practical approach to lung health.	It was felt that the certainty of evidence of the tests effects were moderate weighing up the different considerations, namely: • CXR radiation risk is low. • There are direct benefits of finding more TB and other diseases. Potential additional burden on the PLHIV to get a CXR, unless mobile CXR is available.

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	The initial diagnostic test in the TB case detection cascade after screening is WHO recommended molecular diagnostics. While Xpert® MTB/RIF can achieve a pooled sensitivity of 81% in PLHIV, the Ultra assay has a 7% increase in sensitivity. Currently, Xpert® MTB/RIF and more recently Xpert® MTB/RIF Ultra are recommended by WHO as the initial TB diagnostic test in adults and children living with HIV. LF-LAM is also recommended for PLHIV. Unlike other diagnostic tests, the sensitivity of this test increases with advancement of HIV disease. LF-LAM can achieve a pooled sensitivity of 62% in inpatient settings. TB treatment is effective among PLHIV, and successful treatment outcomes comparable to those in people not living with HIV can be achieved (*Owiti et al*). Conversely, undetected TB and delays in diagnosis of TB and MDR-TB diagnosis among people living with HIV are associated with increased mortality as well as increased transmission. Findings from a systematic review (*Gupta et al*) of post mortem studies among people who had died from AIDS, highlighted close to 50% of post mortem diagnosis of TB had not been detected prior to death, and TB was the cause of death in more 91.4% (95% CI 85.8-97.0%) of all the TB cases. Should a diagnostic test not be indicated, programmatic management of latent TB infection and initiation of antiretroviral treatment would be the next course of action if eligible. A systematic review of 12 randomized controlled trials found that TB preventive treatment reduced the overall risk of TB by 33% (RR 0.67, 95% CI 0.51-0.87) as well as a reduction in mortality. Clinical trials confirm that early use of ART keeps people living with HIV alive and healthier and reduces the risk of transmitting the virus to others. Earlier treatment has the further advantage of simplifying the operational demands on programmes.	

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	No included studies. Whilst chest X-ray is infrastructure-dependent, advancements in digital technology are rapidly expanding the potential contribution of chest radiography to TB detection, and TB programmes are scaling up access to chest X-ray. However in many resource-limited settings placement is not decentralized. Depending on the patients first port of entry, costs and time for them to travel to get an X-ray can hinder access, which can in turn hamper any decision for diagnostic	No included studies but the following considerations were voiced:

studies	confirmation. Health worker perspectives on barriers to delivery of routine tuberculosis diagnostic evaluation services in Uganda: a qualitative study to guide clinic-based interventions, *BMC Health Serv Res* **15,** 10 (2015). https://doi.org/10.1186/s12913-014-0668-0	Results are usually available on the same day - but in some countries, CXR must be read by a qualified radiologist and this can significantly increase CXR turn-around-time. Lack of access to CXR should not be a barrier to TPT initiation.

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	No included studies.	

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	Screening aims to identify people with active TB earlier and therefore ensure earlier treatment and better health outcomes for individuals and lower TB transmission to community. No direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There may be important uncertainty in what we know about how patients perceive and value the outcomes explored above.	

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 panel determines that the balance between desirable and undesirable effects probably favours **using** CXR (any abnormality) to be used in parallel with the W4SS, as an additional screening test to the W4SS, alone, for detection of TB among people living with HIV.	

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 included studies directly addressed this subquestion. Indirect evidence: Indirect evidence available from six studies for the use of CXR among PLHIV (not on ART) (Shah 2008, Shah 2009, Bassett 2010, Maheswaran 2012, Murray 2016, Bogdanova 2019). Four studies from Sub-Saharan Africa, one in Vietnam (Shah 2008) and one from the Russian Federation (Bogdanova 2019). Studies provided indirect evidence because they compared CXR to passive case finding not to W4SS directly. Two studies reported unit test costs alone for CXR ranging from US$2.23-$6.46 (Shah 2009, Murray 2016) and two studies reported health system and test costs for CXR with W4SS ranging from US$2.53-$6.46 (Shah 2008, Bogdanova 2019). The unit test cost of Xpert ranged from US$18-28 (Murray 2016), From the studies reporting indirect evidence, CXR unit costs represent a small proportion of the total implementation costs for an ACF program. Limited data are available on the programmatic costs such as overheads for CXR.	Resources required were considered to be moderate weighing up the following issues: Compared with the W4SS, CXR would have additional costs, even though the unit cost of digital CXR is not high. Radiology combined with the W4SS is more sensitive than the W4SS alone so more individuals would need diagnostic confirmation with mWRD added to the screening algorithm. Unit costs of both radiology and mWRD are artificially set costs and can vary considerably by context (e.g. income setting). These costs might change.

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	All of the above evidence was indirect, five of the six studies were high quality, while one study was determined to be of lower quality. Among the six studies, unit costs for CXR was consistent but there was heterogeneity of the screening algorithms and outcomes.	

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 intervention ○ Varies ○ No included studies	Among the six studies included here, costs per patient screened ranges from US$2.53-$57 when health system and test costs were included (Shah 2008, Bassett 2010, Bogdanova 2019). Cost per patient diagnosed ranged from US$40-$549 per TB patient diagnosed (Shah 2008, Shah 2009, Bogdanova 2019). The use of CXR following W4SS was shown to be cost-effective in South Africa and Russia. Bassett et al. found an intervention of W4SS followed by smear microscopy, CXR and Xpert for diagnosis, compared to screening with cough alone among PLHIV in South Africa, to be cost-effective with an ICERs of US$360 per additional TB patient diagnosed. Bogdanova et al. found mass CXR screening with a mobile unit to be cost-effective among PLHIV in Russia with an ICER of US$549 per additional TB patient diagnosed. The use of CXR for screening following W4SS and smear microscopy in SSA was shown to be cost-effective with an ICER of US$7775 per QALY (Mahesawaran 2012). Murray et al. found CXR used as a triage test followed by diagnosis with Xpert in Uganda to be cost-effective with an ICER of US$610 per year of life saved. Major drivers of cost were determined to be the background prevalence of TB, clinic setting (i.e. urban vs. rural), patient volume and programmatic and staffing costs.	

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	CXR is not yet decentralized in many lower and middle income countries thus access to this test for people living outside of district/national centres would be limited, and could incur high costs and patient time for receiving a screen, when compared with the readily available easy-to-use W4SS. Compared with the W4SS, the CXR result and interpretation might be less influenced by response bias, discrimination or stigma. While increased sensitivity and reduced specificity increases the burden and costs of unnecessary diagnostic follow-up investigations, more TB cases would be found using this strategy, thus reducing further disease burden, mortality, and impoverishment.	The following observations were made Context matters, e.g. certain settings have easier access. Scale up is needed to increase equity. In the absence of mobile x-rays, it might reduce equity due to variability in access, and if people have to pay out-of-pocket expenses for transport, X-ray, and there might potentially be gender issues. CXR is *currently* not readily available everywhere which reduces equity. However, machine costs have reduced considerably for digital X-ray and WHO and others have been pushing for increased access to radiology within primary care hospitals and centres so

		access may increase with investment. Currently, equity may be reduced globally, but within country it would be increased. However scale-up would increase equity. Lack of access to CXR should not be a barrier to TPT initiation.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	To date the review team have found very little qualitative evidence relevant this group, but drawing upon the theoretical findings from our wider body of evidence, we would suggest that perceived and enacted HIV and TB stigma will complicate the decision-making of persons living with HIV. 1 study, Tuot 2019 Cambodia Total respondents N = 120; PLHIV, N = 6 Quality assessment not yet done.	

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Chest X-ray is not yet decentralized in part due to infrastructure and HR requirement. Datta et al (2020) have demonstrated that systematic screening with mobile digital X-ray services via a PPP model integrated into the national program is feasible and scalable. Ownership of X-ray can affect access, in terms of costs to the end-user, and as a barrier to accessing services.	The following observations were made: For countries in the European Region it is feasible everywhere, however, in many high TB burden settings it is not readily available. If investment is made in the infrastructure it would be feasible, but currently it is not feasible in many settings due to lack of infrastructure or human resources. CXR is also used for other purposes other than for TB. So it is also about the number of CXRs that can be done without compromising its use for other purposes - so scale-up according to the overall health system needs would be required to ensure feasibility.

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**
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

Chest X-ray may be used to screen for TB disease among adults and adolescents living with HIV (conditional recommendation, moderate certainty of the evidence for test accuracy)

Justification

- Chest X-ray (CXR) is currently recommended by WHO to be used in parallel with the W4SS to assist in ruling out active TB prior to initiating TB preventive treatment among people living with HIV on ART where CXR is available. The GDG agreed that, due to the increased sensitivity, the evidence supported using CXR in addition to the W4SS prior to a diagnostic test in this population.
- Chest X-ray alone was found to have similar sensitivity and similar or higher specificity than the W4SS across all of the sub-populations. When combined in parallel with the W4SS, with a positive screen for either W4SS or CXR indicating a diagnostic test, it had higher or similar sensitivity, and similar specificity to the W4SS alone.
- Data on “any abnormality” and “abnormality suggestive of TB” detected by CXR were reviewed by the GDG who recommended that either approach could be used, depending on context, availability of radiological expertise, resources and preference towards higher sensitivity or higher specificity.
- 6,155 study participants reviewed for CXR in combination with W4SS were adults (>19), 83 were adolescents (10-19).

Subgroup considerations

- Data for TB screening among outpatients living with HIV on ART with CXR, combined in parallel with the W4SS (i.e. positive for either screen), were first assessed and judgements made against all sections of the evidence to decision table. The GDG agreed that their judgements based on research evidence presented and additional considerations made would largely apply to the broader group of people living with HIV and sub-populations. Further, it was agreed that application of the recommendation to the broader population and to all sub-populations, would simplify programming.
- CXR combined in parallel with the W4SS is already recommended for ruling out active TB among people living with HIV on ART prior to initiating TB preventive treatment. This strategy was found to have the highest sensitivity 0.85 (0.69-0.94) compared with other screening tools and strategies assessed in this sub-population.
- For pregnant women and the fetus, a CXR does not pose any significant risk, provided that good practices are observed, as the primary beam is targeted away from the pelvis. (ref: Chest Radiography in Tuberculosis Detection, WHO).
- Whilst the data were limited for inpatients living with HIV, the combined parallel strategy of CXR and the W4SS had the lowest specificity (0.07 (95%CI: 0.03-0.19), compared with its specificity for other sub-populations, similar to findings with CRP and the symptom screen alone.

Implementation considerations

- Countries should position CXR within national TB screening algorithms according to feasibility, resources available, health system level and equity.
- A baseline CXR could be included as part of initial enrollment in ART services.
- The W4SS should always be conducted, regardless of placement of CXR within the algorithm, as part of a comprehensive clinical evaluation.
- Lack of access to CXR should not be a barrier to initiating TB preventive treatment.
- Data were not available for determining optimal periodicity of more intensified screening but the GDG suggested a pragmatic approach that is aligned with HIV services, with screening provided as part of annual check-ups such as viral load monitoring and other investigations including for ruling out TB prior to TB preventive treatment – depending on the context, TB prevalence, feasibility and access.

Monitoring and evaluation

- Countries are encouraged to monitor and evaluate the yield of TB screening among people living with HIV, disaggregated by screening tools to inform programming and resource planning.

Research priorities

- Well-designed clinical trials to strengthen the evidence on the accuracy, effectiveness (including impact on patient-important outcomes e.g. mortality), feasibility and cost implications of using CXR to screen for TB across all HIV sub-populations in low, medium and high HIV and TB burden settings with and without high ART coverage, compared with other screening strategies.
- Subpopulations of PLHIV for further investigation would include but not be limited to inpatients, acute care service attendees, patients failing ART, newly diagnosed HIV patients enrolling in ART clinics, stable patients established on ART, pregnant women, children and adolescents living with HIV and key populations.
- Research to evaluate the effectiveness and accuracy of combining, in different sequence, the W4SS, CRP, CXR, mWRD and LF-LAM in screening and diagnostic algorithms.
- Frequency of screening - effectiveness, cost effectiveness, feasibility and acceptability and optimal periodicity of routine regular screening with CXR among PLHIV in care.

Table 8. Should molecular WHO-recommended rapid diagnostic tests (mWRDs) vs. WHO-recommended 4 symptom screen followed by an mWRD be used to screen for TB disease in inpatients with HIV?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	• People with HIV were 19 times more at risk of TB incidence than those without HIV in 2018 and a third of AIDS deaths were due to TB in 2018. Ensuring early detection and timely treatment of TB among people living with HIV is critical for reducing the mortality. • An estimated 44% of PLHIV with TB were not notified to have reached care in 2018. A systematic review of autopsy studies found as high as 64% of TB prevalence among people who had died from AIDS, in half of whom TB had been undetected prior to death. • Since 2011 WHO has recommended "Adults and adolescents living with HIV should be screened for TB with a clinical algorithm; those who report any one of the symptoms of current cough, fever, weight loss or night sweats may have active TB and should be evaluated for TB and other diseases" Screening with the WHO-recommended 4 symptom screen (W4SS) has been recommended at every health visit. • Designed primarily for ruling out active TB prior to the initiation of TB preventive treatment, this screen has a relatively high negative predictive value. However, evidence from symptom screening in certain sub-populations of PLHIV, namely PLHIV not on ART, with low CD4 cell counts and inpatients, the screen has low specificity, and among PLHIV on ART and among pregnant women, the screen has low sensitivity. • With the advancement of new tools such as digital chest X-ray (CXR), mWRDs, and new evidence on TB screening among PLHIV, a systematic literature review and individual patient data meta-analysis was commissioned to determine the accuracy of screening tests and approaches that lead to better outcomes in comparison to the W4SS. • There has been some debate around screening for TB using mWRD among people living with HIV, particularly among sub-populations where the symptom screen has been shown to have reduced sensitivity. (Modi et al, 2016)	

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	As part of the systematic review and individual patient data analysis, the accuracy of mWRD was assessed against the accuracy of the W4SS followed by a mWRD in all inpatients with HIV, irrespective of ART status, compared to culture reference standard. **Test accuracy** molecular WHO-approved rapid diagnostics Sensitivity: 0.77 (95% CI: 0.69 to 0.84) Specificity: 0.93 (95% CI: 0.89 to 0.96) WHO-recommended 4 symptom screen followed by mWRD Sensitivity: 0.76 (95% CI: 0.68 to 0.83) Specificity: 0.93 (95% CI: 0.89 to 0.96)	For inpatients the W4SS alone has 96% sensitivity and 11% specificity with 94% of patients symptom positive, so the value of the W4SS was judged by the GDG to have limited utility in this population. This explains for the small difference between the intervention (mWRD) and the comparator (W4SS + mWRD). Although the W4SS is recommended

		for all PLHIV, it was not designed for inpatients. The GDG noted that CRP also has limited specificity (11%) in detecting TB in this sub-population.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
● Trivial ○ Small ○ Moderate ○ Large ○ Varies ○ Don't know	The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not already started treatment (true negative). The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in early initiation of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality **Summary for pretest probability of 10%** Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among inpatients with HIV regardless of ART status), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 69-84 people with TB as positive (true positive), 16-31 people with TB as negative (false negative), 36-99 people without TB as positive for TB (false positive), and 801 to 864 people without TB as negative (true negative). **Summary for pretest probability of 20%** Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among inpatients with HIV regardless of ART status), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 138-168 people with TB as positive (true positive), 32-62 people with TB as negative (false negative), 32-88 people without TB as positive for TB (false positive), and 712-768 people without TB as negative (true negative). **Summary for pretest probability of 30%** Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 207-252 people with TB as positive (true positive), 48-93 people with TB as negative (false negative), 28-77 people without TB as positive for TB (false positive), and 623-672 people without TB as negative (true negative).	It was noted that using mWRD to screen for TB among medical inpatients living with HIV generated just a few more cases than the symptom screen followed by mWRD, due to the limited added utility of the W4SS within the full screening and diagnostic algorithm, as per the current recommendation.

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 10%		Effect per 1000 patients/year for pre-test probability of 20%		Effect per 1000 patients/year for pre-test probability of 30%		Importance
			molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	
True positives	cross-sectional (cohort type accuracy study)	⊕⊕⊕◯ MODERATE[a,b,c]	77 (69 to 84)	76 (68 to 83)	154 (138 to 168)	152 (136 to 166)	231 (207 to 252)	228 (204 to 249)	
TP absolute difference			1 more TP in molecular WHO-approved rapid diagnostics		2 more TP in molecular WHO-approved rapid diagnostics		3 more TP in molecular WHO-approved rapid diagnostics		
False negatives			23 (16 to 31)	24 (17 to 32)	46 (32 to 62)	48 (34 to 64)	69 (48 to 93)	72 (51 to 96)	
FN absolute difference			1 fewer FN in molecular WHO-approved rapid diagnostics		2 fewer FN in molecular WHO-approved rapid diagnostics		3 fewer FN in molecular WHO-approved rapid diagnostics		
True negatives	cross-sectional (cohort type accuracy study)	⊕⊕⊕◯ MODERATE[a,d]	837 (801 to 864)	837 (801 to 864)	744 (712 to 768)	744 (712 to 768)	651 (623 to 672)	651 (623 to 672)	
TN absolute difference			0 fewer TN in molecular WHO-approved rapid diagnostics		0 fewer TN in molecular WHO-approved rapid diagnostics		0 fewer TN in molecular WHO-approved rapid diagnostics		
False positives			63 (36 to 99)	63 (36 to 99)	56 (32 to 88)	56 (32 to 88)	49 (28 to 77)	49 (28 to 77)	
FP absolute difference			0 fewer FP in molecular WHO-approved rapid diagnostics		0 fewer FP in molecular WHO-approved rapid diagnostics		0 fewer FP in molecular WHO-approved rapid diagnostics		

a. All but one study were considered at low risk of bias in all domains in the overall analysis. However, three studies obtained only sputum samples. This likely resulted in misclassification of the target condition by missing extrapulmonary TB. We downgraded one level for risk of bias.
b. Sensitivity estimates ranged from 25% to 83% with overlapping Cis. We did not downgrade for inconsistency.
c. Four studies were considered a possible concern for applicability in the overall analysis. Three of these studies evaluated only individuals with CD4 cell count ≤350 per µL and one study included only inpatients. However, since this assessment is for inpatients, these study populations are likely to represent common characteristics of the target population. We did not downgrade for indirectness.
d. Specificity estimates ranged from 90% to 96%. 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	The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not already started treatment (true negative). The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in early initiation of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality. **Summary for pretest probability of 10%** Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among inpatients with HIV regardless of ART status), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 69-84 people with TB as positive (true positive), 16-31 people with TB as negative (false negative), 36-99 people without TB as positive for TB (false positive), and 801 to 864 people without TB as negative (true negative). **Summary for pretest probability of 20%** Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among inpatients with HIV regardless of ART status), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 138-168 people with TB as positive (true positive), 32-62 people with TB as negative (false negative), 32-88 people without TB as positive for TB (false positive), and 712-768 people without TB as negative (true negative). **Summary for pretest probability of 30%** Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 207-252 people with TB as positive (true positive), 48-93 people with TB as negative (false negative), 28-77 people without TB as positive for TB (false positive), and 623-672 people without TB as negative (true negative).	

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 10%		Effect per 1000 patients/year for pre-test probability of 20%		Effect per 1000 patients/year for pre-test probability of 30%		Importance
			molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	
True positives	cross-sectional (cohort type accuracy study)	⊕⊕⊕◯ MODERATE[a,b,c]	77 (69 to 84)	76 (68 to 83)	154 (138 to 168)	152 (136 to 166)	231 (207 to 252)	228 (204 to 249)	
TP absolute difference			1 more TP in molecular WHO-approved rapid diagnostics		2 more TP in molecular WHO-approved rapid diagnostics		3 more TP in molecular WHO-approved rapid diagnostics		
False negatives			23 (16 to 31)	24 (17 to 32)	46 (32 to 62)	48 (34 to 64)	69 (48 to 93)	72 (51 to 96)	
FN absolute			1 fewer FN in molecular WHO-approved rapid diagnostics		2 fewer FN in molecular WHO-approved rapid diagnostics		3 fewer FN in molecular WHO-approved rapid diagnostics		

difference									
True negatives	cross-sectional (cohort type accuracy study)	⊕⊕⊕◯ MODERATE[a,d]	837 (801 to 864)	837 (801 to 864)	744 (712 to 768)	744 (712 to 768)	651 (623 to 672)	651 (623 to 672)	
TN absolute difference			0 fewer TN in molecular WHO-approved rapid diagnostics		0 fewer TN in molecular WHO-approved rapid diagnostics		0 fewer TN in molecular WHO-approved rapid diagnostics		
False positives			63 (36 to 99)	63 (36 to 99)	56 (32 to 88)	56 (32 to 88)	49 (28 to 77)	49 (28 to 77)	
FP absolute difference			0 fewer FP in molecular WHO-approved rapid diagnostics		0 fewer FP in molecular WHO-approved rapid diagnostics		0 fewer FP in molecular WHO-approved rapid diagnostics		

a. All but one study were considered at low risk of bias in all domains in the overall analysis. However, three studies obtained only sputum samples. This likely resulted in misclassification of the target condition by missing extrapulmonary TB. We downgraded one level for risk of bias.
b. Sensitivity estimates ranged from 25% to 83% with overlapping Cis. We did not downgrade for inconsistency.
c. Four studies were considered a possible concern for applicability in the overall analysis. Three of these studies evaluated only individuals with CD4 cell count ≤350 per µL and one study included only inpatients. However, since this assessment is for inpatients, these study populations are likely to represent common characteristics of the target population. We did not downgrade for indirectness.
d. Specificity estimates ranged from 90% to 96%. 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	The systematic review of the performance of mWRD to screen for TB among inpatients living with HIV, regardless of ART status included 4 studies with a total of 639 participants. The prevalences of TB in PLHIV in the 4 included studies were 7% (Myanmar), 20% (Ghana), 25% (South Africa) and 26% (South Africa). These results may not apply in lower-prevalence settings. Certainty of evidence was moderate for both sensitivity and specificity. The reason for downgrading was for serious risk of bias as three studies obtained only sputum samples which likely resulted in misclassification of the target condition by missing extrapulmonary TB. Overall, the certainty of the evidence of accuracy is moderate.	

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 10%		Prevalence 20%		Prevalence 30%			
	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by an mWRD diagnostic test		
True positives patients with active TB	77 (69 to 84)	76 (68 to 83)	154 (138 to 168)	152 (136 to 166)	231 (207 to 252)	228 (204 to 249)	639 (4)	⊕⊕⊕◯ MODERATE[a,b,c]
	1 more TP in molecular WHO-approved rapid diagnostics		2 more TP in molecular WHO-approved rapid diagnostics		3 more TP in molecular WHO-approved rapid diagnostics			
False negatives patients incorrectly classified as not having active TB	23 (16 to 31)	24 (17 to 32)	46 (32 to 62)	48 (34 to 64)	69 (48 to 93)	72 (51 to 96)		
	1 fewer FN in molecular WHO-approved rapid diagnostics		2 fewer FN in molecular WHO-approved rapid diagnostics		3 fewer FN in molecular WHO-approved rapid diagnostics			
True negatives patients without active TB	837 (801 to 864)	837 (801 to 864)	744 (712 to 768)	744 (712 to 768)	651 (623 to 672)	651 (623 to 672)	639 (4)	⊕⊕⊕◯ MODERATE[a,d]
	0 fewer TN in molecular WHO-approved rapid diagnostics		0 fewer TN in molecular WHO-approved rapid diagnostics		0 fewer TN in molecular WHO-approved rapid diagnostics			
False positives patients incorrectly classified as having active TB	63 (36 to 99)	63 (36 to 99)	56 (32 to 88)	56 (32 to 88)	49 (28 to 77)	49 (28 to 77)		
	0 fewer FP in molecular WHO-approved rapid diagnostics		0 fewer FP in molecular WHO-approved rapid diagnostics		0 fewer FP in molecular WHO-approved rapid diagnostics			
Inconclusive	undefined						(0)	-
Complications	undefined						(0)	-

a. All but one study were considered at low risk of bias in all domains in the overall analysis. However, three studies obtained only sputum samples. This likely resulted in misclassification of the target condition by missing extrapulmonary TB. We downgraded one level for risk of bias.
b. Sensitivity estimates ranged from 25% to 83% with overlapping Cis. We did not downgrade for inconsistency.
c. Four studies were considered a possible concern for applicability in the overall analysis. Three of these studies evaluated only individuals with CD4 cell count ≤350 per µL and one study included only inpatients. However, since this assessment is for inpatients, these study populations are likely to represent common characteristics of the target population. We did not downgrade for indirectness.

	d. Specificity estimates ranged from 90% to 96%. 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	No research evidence was found. No direct benefits, adverse effects or burden resulting directly from the test itself are anticipated.	

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	LF-LAM is also recommended as an additional diagnostic test for inpatient PLHIV. Unlike other diagnostic tests, the sensitivity of this test increases with advancement of HIV disease. LF-LAM can achieve a pooled sensitivity of 62% in inpatient settings.(Strong recommendation; moderate certainty in the evidence about the intervention effects). TB treatment is highly effective among PLHIV, and successful treatment outcomes comparable to those among people not living with HIV can be achieved (Owiti et al) (Strong recommendation, high certainty in the estimates of effect). Conversely, undetected TB and delays in diagnosis of TB and MDR-TB diagnosis among people living with HIV are associated with increased mortality as well as increased transmission. Findings from a systematic review (Gupta et al) of post mortem studies among people who had died from AIDS, highlighted close to 50% of post mortem diagnosis of TB had not been detected prior to death, and TB was the cause of death in more 91.4% (95% CI 85.8-97.0%) of all the TB cases. Should active TB not be identified through further investigations, programmatic management of latent TB infection and initiation of antiretroviral treatment would be the next course of action if found otherwise eligible (Strong recommendation, high quality of evidence). A systematic review of 12 randomized controlled trials found that TB preventive treatment reduced the overall risk of TB by 33% (RR 0.67, 95% CI 0.51-0.87) as well as a reduction in mortality. Clinical trials confirm that early use of ART keeps people living with HIV alive and healthier and reduces the risk of transmitting the virus to others. Earlier treatment has the further advantage of simplifying the operational demands on programmes.	

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	No research evidence was found. mWRD tests are already recommended as the first TB diagnostic test for people living with HIV so its use as a screening test should not be an additional burden in terms of linkage between test results and management decisions, if this policy is already implemented. Test results from mWRDs can in theory be available in a few hours, but due to laboratory batching and burden, they are usually available within one day. A screening test with a delay of a few hours to days' time could impact retention of patients in the screening pathway.	

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	No study of effectiveness of inpatients living with HIV was sought.	

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	Screening aims to identify people with active TB earlier and therefore ensure earlier treatment and better health outcomes for individuals and lower TB transmission to community. No direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is no important uncertainty in what we know about how patients perceive and value the outcomes explored above.	

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 panel determines that the balance between desirable and undesirable effects does not favour either **using or not using** an mWRD screening and diagnostic strategy as an alternative test to the WHO Four Symptoms Screen followed by a mWRD test for detection of TB among inpatients with HIV, regardless of ART status.	The balance of effects does not favour either as the symptom screen is ubiquitous in this population and has limited utilility and limited added value.

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	Data was available from one study that provided indirect evidence on the resources required for the use mWRDs (i.e. Xpert) to screen for active TB in inpatient PLHIV (regardless of ARV), but compared an ICF strategy of urine LAM, urine Xpert and sputum Xpert to a reference strategy of sputum Xpert alone (Reddy 2019). In this study conducted in Malawi and South Africa, the unit cost for urine and sputum Xpert ranged from: US$5-$36 (Reddy 2019). Programmatic costs were not provided for the implementation of this ICF strategy.	The difference in the number of tests implemented with mWRD alone, and thus resources required, would be limited as the most patients are WHO 4 symptom screen positive.

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	There is no study reporting directly on the comparison of interest. One study assessed the use of mWRDs (i.e. Xpert) to screen for ATB in inpatient PLHIV (regardless of ARV) and was of good quality.	

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 intervention ○ Varies ● No	There is no direct evidence for this question. The use of mWRDs alone to screen for active TB in inpatient PLHIV was not found to be cost-effective, when compared to an ICF screening intervention of urine Xpert, TB-LAM and sputum Xpert over a patient's lifetime. This is due to the fact that additional screening with TB-LAM has a high incremental diagnostic yield, but a relatively low cost-component when compared with Xpert alone, among hospitalized HIV patients in high HIV, high TB burden settings such as Malawi and South Africa.	

included studies		

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	mWRD tests are currently recommended as the first diagnostic test for TB for people living with HIV (Strong recommendation, high quality of evidence) so its use as a screening test should have limited negative impact on equity compared with the WHO four symptom screen followed by mWRD. Increased sensitivity in a screen facilitates earlier detection of a debilitating impoverishing disease. WHO approved rapid molecular diagnostics are quantitative. Compared with the 4 symptom screen, the result and interpretation might be less influenced by personal judgement, response bias, discrimination or stigma. One literature review found that at-risk individuals report that fear of TB stigma and the social and economic impact of stigma affects their willingness to undergo TB screening - Tuberculosis and stigmatization: pathways and Interventions Public Health Rep. 2010; 125(Suppl 4): 34 42.doi: 10.1177/00333549101250S407PMCID: PMC2882973PMID: 20626191	GDG members agreed on "probably no impact" as the WHO symptom screen has limited utility in this population and makes little difference to the screening and diagnostic algorithm.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	To date the review team has found very little qualitative evidence relevant for this group, but drawing upon the theoretical findings from our wider body of evidence, we would suggest that perceived and enacted HIV and TB stigma will complicate the decision-making of persons living with HIV. (1 study, Tuot 2019 Cambodia, Total respondents N = 120; PLHIV, N = 6, Quality assessment not yet done).	

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	mWRD tests are now recommended by WHO as the primary mode of TB diagnosis. mWRDs are able to be implemented at most levels of health care. A randomized controlled trial (*Theron et al*) found that Xpert MTB/RIF can be accurately administered as a diagnostic test by a nurse in primary-care clinics, resulting in more patients starting same-day treatment, more culture-positive patients starting therapy, and a shorter time to treatment.	

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**
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

Adult and adolescent inpatients with HIV in medical wards where TB prevalence is ≥10% should be tested systematically for TB with a WHO-approved rapid molecular diagnostic test. **(strong recommendation, moderate certainty of the evidence for test accuracy)**

Summary of GDG Comments:

- The W4SS is positive in most medical inpatients (94% screened positive)
- Screening with the W4SS did not add clinical value and thus all patients should be screened with mWRD
- The results are similar when comparing W4SS with mWRDS for inpatients (a few more cases of TB will be detected).
- Resource requirements similar
- Equity is not expected to be impacted
- ART component (check if the patients had advanced disease or low CD4s)
- Testing all these medical patients for TB with mWRD was considered important - but not patients who have a broken leg - so clinical assessment is also required.

Justification

- TB is the main cause of hospitalization and mortality among PLHIV.
- The W4SS had 96% sensitivity and 11% specificity in the individual patient data meta-analysis of medical ward inpatients living with HIV. 94% were symptom screen positive. The value of the W4SS was therefore judged to have limited utility in screening for TB in this population prior to an mWRD test, and the GDG thus recommended that medical inpatients should be screened and tested with the molecular test, irrespective of symptoms to inform a decision to treat.
- Data reviewed for the 634 medical ward inpatients from the following 4 studies in high TB burden settings: Bjerrum 2015, 50 subjects, Ghana, 20% TB prevalence; Heidebrecht 2016, 121 subjects, South Africa, 25% TB prevalence; Lawn 2015, 414 subjects, South Africa 26% TB prevalence; and Thit 2017, 54 subjects, Myanmar, 7% TB prevalence. In lower prevalence settings, a screen and test strategy with mWRD alone would give rise to higher numbers of false positives with overtreatment and related social and economic consequences including potential delay in start of ART. This recommendation may not be applicable in settings with a lower pre-test probability of TB.
- Only 5 study participants reviewed for mWRD among inpatients were adolescents (10-19).
- Data on mWRD from the individual patient data meta-analysis were primarily based on studies that used Xpert MTB/RIF.

Subgroup considerations

Implementation considerations

- Careful clinical assessment is recommended to ensure TB is the primary cause of illness and other conditions leading to the clinical presentation are also managed.
- Other TB investigations, including LF-LAM, as indicated, should be used as part of a comprehensive workup to help inform any decision to treat for active TB.
- A negative mWRD does not exclude TB. Patients who are mWRD-negative but are manifestly sick, may not be able to produce a quality sputum sample or may have extrapulmonary TB.
- For patients with a prior history of TB in the past 2 years, a positive result may be due to the presence of DNA detected from previously treated TB.
- Should the patient be unable to provide sputum, other biological specimens should be considered as indicated.
- TB prevalence among people living with HIV in medical wards may be calculated as the percentage of admissions of people living with HIV over a 6-12 month period that are diagnosed with TB.

Monitoring and evaluation

- Countries are encouraged to monitor and evaluate the yield of TB screening among people living with HIV, disaggregated by screening tools to inform programming and resource planning.

Research priorities

- Well-designed clinical trials to strengthen the evidence on the accuracy, effectiveness (including impact on patient-important outcomes e.g. mortality), feasibility and cost implications of using mWRD to test for TB among medical inpatients in low, medium, moderate and high HIV and TB burden settings, with and without high ART coverage, compared with other screening strategies.
- Screening with mWRD using specimens other than sputum.

Table 9. Should molecular WHO-recommended rapid diagnostic tests (mWRDs) vs. WHO-recommended 4 symptom screen followed by mWRD be used to screen for TB disease in people living with HIV?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	• People with HIV were 19 times more at risk of TB incidence than those without HIV in 2018 and a third of AIDS deaths were due to TB in 2018. Ensuring early detection and timely treatment of TB among people living with HIV is critical for reducing the mortality. • An estimated 44% of PLHIV with TB were not notified to have reached care in 2018. A systematic review of autopsy studies found as high as 64% of TB prevalence among people who had died from AIDS, in half of whom TB had been undetected prior to death. • Since 2011 WHO has recommended "Adults and adolescents living with HIV should be screened for TB with a clinical algorithm; those who report any one of the symptoms of current cough, fever, weight loss or night sweats may have active TB and should be evaluated for TB and other diseases" Screening with the WHO 4 symptom screen (W4SS) has been recommended at every health visit. • Designed primarily for ruling out active TB prior to the initiation of TB preventive treatment, this screen has a relatively high negative predictive value. However, evidence from symptom screening in certain sub-populations of PLHIV, namely PLHIV not on ART, with low CD4 cell counts and inpatients, the screen has low specificity, and among PLHIV on ART and among pregnant women, the screen has low sensitivity. • With the advancement of new tools such as digital chest X-ray (CXR), molecular WHO-Approved rapid diagnostics (mWRD), and new evidence on TB screening among PLHIV, a systematic literature review and individual patient data meta-analysis was commissioned to determine the accuracy of screening tests and approaches that lead to better outcomes in comparison to the W4SS. • There has been some debate around screening for TB using mWRD among people living with HIV, particularly among sub-populations such as pregnant women living with HIV and attendees of ART clinics among whom the symptom screen has been shown to have reduced sensitivity.	

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very inaccurate ○ Inaccurate ● Accurate ○ Very accurate ○ Varies ○ Don't know	As part of the systematic review and individual patient data analysis, the accuracy of mWRD was assessed against the accuracy of the W4SS followed by an mWRD test in all people living with HIV, irrespective of ART status, compared to culture reference standard. **Test accuracy** molecular WHO-approved rapid diagnostics Sensitivity: 0.69 (95% CI: 0.60 to 0.76) Specificity: 0.98 (95% CI: 0.97 to 0.99) WHO-recommended 4 symptom screen (current cough, weight loss, night sweats, fever) followed by mWRD Sensitivity: 0.62 (95% CI: 0.56 to 0.69) Specificity: 0.99 (95% CI: 0.97 to 0.99)	mWRD was assessed first among medical ward inpatients living with HIV for whom a strong recommendation was made given the poor performance of the W4SS in this population. mWRD was then assessed among pregnant women living with HIV, compared to the W4SS followed by mWRD. A conditional recommendation was made for this group. The GDG then reviewed the accuracy of the mWRD across all PLHIV and among other HIV sub-populations and agreed that there were no significant

Population	Test	Test accuracy	W4SS accuracy	(persons)	evidence	Lower prevalence cutoff	Middle prevalence cutoff	higher prevalence cutoff
All PLHIV (prev: 5%, 10%, 20%)	mWRD	Se: 0.69	Se 0.62	14 (9209)	Moderate	TP: 34/ FN:16	TP: 69/ FN:31	TP: 138/FN: 62
All PLHIV (prev: 5%, 10%, 20%)	mWRD	Sp: 0.98	Sp 0.99	14 (9209)	High	TN: 931/ FP: 19	TN: 882/ FP: 18	TN: 784/FP: 16
Outpatients not on ART (prev: 5%, 10%,	mWRD	Se: 0.72	Se 0.65	10 (5796)	Moderate	TP: 36/ FN: 14	TP: 72/ FN: 28	TP: 144/ FN: 56
Outpatients not on ART (prev: 5%, 10%,	mWRD	Sp: 0.98	Sp 0.99	10 (5796)	High	TN: 931/ FP: 19	TN: 882/ FP: 18	TN: 784/ FP: 16
CD4 <200 (prev: 5%, 10%, 20%)	mWRD	Se: 0.76	Se 0.70	12 (3422)	Moderate	TP: 38/FN:12	TP: 76/ FN: 24	TP: 152/ FN: 48
CD4 <200 (prev: 5%, 10%, 20%)	mWRD	Sp: 0.97	Sp 0.97	12 (3422)	High	TN: 922/FP:28	TN: 873/ FP: 27	TN: 776/ FP: 24
Outpatients on ART (prev: 1%, 5%, 10%)	mWRD	Se: 0.54	Se 0.41	4 (2645)	Very low	TP: 5/ FN: 5	TP: 27/ FN: 23	TP: 54/ FN: 46
Outpatients on ART (prev: 1%, 5%, 10%)	mWRD	Sp:0.99	Sp 0.99	4 (2645)	High	TN: 980/ FP: 10	TP: 941/ FP: 9	TN: 891/ FP: 9
Pregnant women living (prev: 1%, 5%, 1	mWRD	Se: 0.55	Se 0.44	4 (473)	Moderate	TP: 6/ FN: 4	TP: 28/ FN: 22	TP: 55/ FN: 45
Pregnant women living (prev: 1%, 5%, 1	mWRD	Sp: 0.99	Sp 0.99	4 (473)	High	TN: 980/ FP: 10	TN: 941/ FP: 9	TN: 891/ FP: 9

Table: **All data for mWRD as a screening test in people living with HIV**

- Note that the GDG initially considered the data specific to medical inpatients with HIV and for pregnant women living with HIV, and subsequently compared accuracy, benefits, and harms for all subpopulations of PLHIV

differences in accuracy of the tool between the different populations, when compared with the W4SS, followed by mWRD.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ● Small ○ Moderate ○ Large ○ Varies ○ Don't know	The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not already started treatment (true negative). The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in early initiation of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality.	Although the GDG largely agreed that there is a small benefit, with a few more true positive detected - it was acknowledged that in the case of pregnant women living with HIV the higher number of true positives would have a beneficial impact both mother and child. Data on outpatients not on ART and on ART had similar findings to pregnant women and all PLHIV.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%		Prevalence 10%		Prevalence 20%			
	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD		
True positives patients with active TB	34 (30 to 38)	31 (28 to 34)	69 (60 to 76)	62 (56 to 69)	138 (120 to 152)	124 (112 to 138)	9209 (14)	⊕⊕⊕◯ MODERATE[a,b,c,d]
	3 more TP in molecular WHO-approved rapid diagnostics		7 more TP in molecular WHO-approved rapid diagnostics		14 more TP in molecular WHO-approved rapid diagnostics			
False negatives patients incorrectly classified as not having active TB	16 (12 to 20)	19 (16 to 22)	31 (24 to 40)	38 (31 to 44)	62 (48 to 80)	76 (62 to 88)		
	3 fewer FN in molecular WHO-approved rapid diagnostics		7 fewer FN in molecular WHO-approved rapid diagnostics		14 fewer FN in molecular WHO-approved rapid diagnostics			
True negatives patients without active TB	931 (922 to 941)	941 (922 to 941)	882 (873 to 891)	891 (873 to 891)	784 (776 to 792)	792 (776 to 792)	9209 (14)	⊕⊕⊕⊕ HIGH[a,b,e]
	10 fewer TN in molecular WHO-approved rapid diagnostics		9 fewer TN in molecular WHO-approved rapid diagnostics		8 fewer TN in molecular WHO-approved rapid diagnostics			

False positives patients incorrectly classified as having active TB	19 (9 to 28)	9 (9 to 28)	18 (9 to 27)	9 (9 to 27)	16 (8 to 24)	8 (8 to 24)		
	10 more FP in molecular WHO-approved rapid diagnostics		9 more FP in molecular WHO-approved rapid diagnostics		8 more FP in molecular WHO-approved rapid diagnostics			

a. Low risk of bias in all but one included studies. Flow and timing was at high risk of bias in that study. We did not downgrade.
b. Six studies were considered a concern for applicability. One study was in pregnant participants. Three studies evaluated only individuals with CD4 cell count ≤350 per µL; however, we recognize this is how patients may present in practice. Two studies evaluated only inpatients; however, sensitivity estimates were higher and specificity estimates were lower, but specificity was still high (90 and 95%) and may partly be because Xpert assay identifies patients with TB that the reference standard (culture) does not. We did not downgrade for indirectness.
c. The confidence intervals (CI) for sensitivity are sufficiently narrow (CI half width = 8) and the lower limit is not significantly lower than the lower limit and point estimate of WHO screen then Xpert strategy. The upper limit is significantly higher. Given that this may lead to small differences depending on which limits are assumed and that Xpert for all must have greater or equivalent sensitivity compared to WHO screen then Xpert, we did not downgrade for imprecision.
d. Sensitivity estimates ranged from 25% to 91% in all studies. Lower estimates were seen in pregnant and on ART populations and higher estimates were seen in inpatient studies; however, this was not always the case and we could not always explain the variability. We downgraded one level for inconsistency.
e. Specificity estimates ranged from 97% to 100% in all but two studies done in inpatients where the specificity was 90% and 95% and may explain the variability. CIs also overlapped. We did not downgrade for inconsistency.

Summary for pretest probability of 5%

Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD alone would correctly classify 30-38 people with TB as positive (true positive), 12-20 people with TB as negative (false negative), 9-28 people without TB as positive for TB (false positive), and 922 to 941 people without TB as negative (true negative).

Summary for pretest probability of 10%

Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD alone would correctly classify 60-76 people with TB as positive (true positive), 24-40 people with TB as negative (false negative), 9-27 people without TB as positive for TB (false positive), and 873-891 people without TB as negative (true negative).

Summary for pretest probability of 20%

Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD alone would correctly classify 120-152 people with TB as positive (true positive), 48-80 people with TB as negative (false negative), 8-24 people without TB as positive for TB (false positive), and 776-792 people without TB as negative (true negative).

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS

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

The anticipated desirable effect is the correct classification of people living with HIV with active TB as positive for TB (true positive), resulting in timely treatment and reducing further transmission of the mycobacterium and reducing mortality. Another anticipated desirable effect is the identification of people living with HIV who do not have active TB and might be eligible for TB preventive treatment or ART if they have not already started treatment (true negative).

The anticipated undesirable effect is the incorrect classification of people living with HIV without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences, and delay in early initiation of antiretrovirals. Another anticipated undesirable effect is the incorrect classification of a person living with HIV who have active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, potentially inappropriate preventive therapy treatment, or increased risk of mortality.

Outcome	Study design	Test accuracy CoE	Effect per 1000 patients/year for pre-test probability of 5%		Effect per 1000 patients/year for pre-test probability of 10%		Effect per 1000 patients/year for pre-test probability of 20%		Importance
			molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	
True positives	cross-sectional (cohort type accuracy study)	⊕⊕⊕◯ MODERATE[a,b,c,d]	34 (30 to 38)	31 (28 to 34)	69 (60 to 76)	62 (56 to 69)	138 (120 to 152)	124 (112 to 138)	
TP absolute difference			3 more TP in molecular WHO-approved rapid diagnostics		7 more TP in molecular WHO-approved rapid diagnostics		14 more TP in molecular WHO-approved rapid diagnostics		
False negatives			16 (12 to 20)	19 (16 to 22)	31 (24 to 40)	38 (31 to 44)	62 (48 to 80)	76 (62 to 88)	
FN absolute difference			3 fewer FN in molecular WHO-approved rapid diagnostics		7 fewer FN in molecular WHO-approved rapid diagnostics		14 fewer FN in molecular WHO-approved rapid diagnostics		
True negatives	cross-sectional (cohort type accuracy study)	⊕⊕⊕⊕ HIGH[a,b,e]	931 (922 to 941)	941 (922 to 941)	882 (873 to 891)	891 (873 to 891)	784 (776 to 792)	792 (776 to 792)	
TN absolute difference			10 fewer TN in molecular WHO-approved rapid diagnostics		9 fewer TN in molecular WHO-approved rapid diagnostics		8 fewer TN in molecular WHO-approved rapid diagnostics		
False positives			19 (9 to 28)	9 (9 to 28)	18 (9 to 27)	9 (9 to 27)	16 (8 to 24)	8 (8 to 24)	
FP absolute difference			10 more FP in molecular WHO-approved rapid diagnostics		9 more FP in molecular WHO-approved rapid diagnostics		8 more FP in molecular WHO-approved rapid diagnostics		

a. Low risk of bias in all but one included studies. Flow and timing was at high risk of bias in that study. We did not downgrade.
b. Six studies were considered a concern for applicability. One study was in pregnant participants. Three studies evaluated only individuals with CD4 cell count ≤350 per µL; however, we recognize this is how patients may present in practice. Two studies evaluated only inpatients; however, sensitivity estimates were higher and specificity estimates were

- For pregnant women false positives and false negatives are lower than with comparison therefore undesirable effects were considered trivial.
- For the entire cohort of outpatients false positives were similar for mWRD alone when compared with mWRD + W4SS and false negatives were lower. For outpatients not yet on ART false negatives were also lower but false positives were increased. At 5% prevalence, false positives were: 19 (9 to 19), compared with 9 (9 to 19) for the W4SS + mWRD.
- It was noted that the use of mWRDs

	lower, but specificity was still high (90 and 95%) and may partly be because Xpert assay identifies patients with TB that the reference standard (culture) does not. We did not downgrade for indirectness. c. The confidence intervals (CI) for sensitivity are sufficiently narrow (CI half width = 8) and the lower limit is not significantly lower than the lower limit and point estimate of WHO screen then Xpert strategy. The upper limit is significantly higher. Given that this may lead to small differences depending on which limits are assumed and that Xpert for all must have greater or equivalent sensitivity compared to WHO screen then Xpert, we did not downgrade for imprecision. d. Sensitivity estimates ranged from 25% to 91% in all studies. Lower estimates were seen in pregnant and on ART populations and higher estimates were seen in inpatient studies; however, this was not always the case and we could not always explain the variability. We downgraded one level for inconsistency. e. Specificity estimates ranged from 97% to 100% in all but two studies done in inpatients where the specificity was 90% and 95% and may explain the variability. CIs also overlapped. We did not downgrade for inconsistency. **Summary for pretest probability of 5%** Based on the findings of the review, in a hypothetical population of people living with HIV with 5% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 30-38 people with TB as positive (true positive), 12-20 people with TB as negative (false negative), 9-28 people without TB as positive for TB (false positive), and 922 to 941 people without TB as negative (true negative). **Summary for pretest probability of 10%** Based on the findings of the review, in a hypothetical population of people living with HIV with 10% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 60-76 people with TB as positive (true positive), 24-40 people with TB as negative (false negative), 9-27 people without TB as positive for TB (false positive), and 873-891 people without TB as negative (true negative). **Summary for pretest probability of 20%** Based on the findings of the review, in a hypothetical population of people living with HIV with 20% prevalence of TB (consistent with measured prevalences among all people living with HIV), in a hypothetical population of 1,000 people living with HIV being screened, mWRD would correctly classify 120-152 people with TB as positive (true positive), 48-80 people with TB as negative (false negative), 8-24 people without TB as positive for TB (false positive), and 776-792 people without TB as negative (true negative).	for screening vs diagnosis needs to be clearly conveyed - as there is potential harm if mWRD performed without follow up diagnostic assessment, particularly in lower prevalence settings when specificity becomes an issue.

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 systematic review of the performance of mWRD to screen for TB among people living with HIV, regardless of ART status included 14 studies with a total of 9209 participants. The prevalences of TB in PLHIV in the 14 included studies were 1%, 2%, 3% (x2), 10%, 12% (x3), 15% (x2), 16%, 17%, 25% and 26%. These results may not apply in lower-prevalence settings. Overall, the certainty is moderate reflecting the lowest level of certainty for the GRADE evidence. The reasons for downgrading evidence certainty included for inconsistency around the estimates for sensitivity.	• Average TB prevalence among all outpatients living with HIV was 8.6%. 51.2% were W4SS positive.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%		Prevalence 10%		Prevalence 20%			
	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD		
True positives patients with active TB	34 (30 to 38)	31 (28 to 34)	69 (60 to 76)	62 (56 to 69)	138 (120 to 152)	124 (112 to 138)	9209 (14)	⊕⊕⊕◯ MODERATE[a,b,c,d]
	3 more TP in molecular WHO-approved rapid diagnostics		7 more TP in molecular WHO-approved rapid diagnostics		14 more TP in molecular WHO-approved rapid diagnostics			
False negatives patients incorrectly classified as not having active TB	16 (12 to 20)	19 (16 to 22)	31 (24 to 40)	38 (31 to 44)	62 (48 to 80)	76 (62 to 88)		
	3 fewer FN in molecular WHO-approved rapid diagnostics		7 fewer FN in molecular WHO-approved rapid diagnostics		14 fewer FN in molecular WHO-approved rapid diagnostics			
True negatives patients without active TB	931 (922 to 941)	941 (922 to 941)	882 (873 to 891)	891 (873 to 891)	784 (776 to 792)	792 (776 to 792)	9209 (14)	⊕⊕⊕⊕ HIGH[a,b,e]
	10 fewer TN in molecular WHO-approved rapid diagnostics		9 fewer TN in molecular WHO-approved rapid diagnostics		8 fewer TN in molecular WHO-approved rapid diagnostics			
False positives patients incorrectly classified as having active TB	19 (9 to 28)	9 (9 to 28)	18 (9 to 27)	9 (9 to 27)	16 (8 to 24)	8 (8 to 24)		
	10 more FP in molecular WHO-approved rapid diagnostics		9 more FP in molecular WHO-approved rapid diagnostics		8 more FP in molecular WHO-approved rapid diagnostics			

a. Low risk of bias in all but one included studies. Flow and timing was at high risk of bias in that study. We did not downgrade.
b. Six studies were considered a concern for applicability. One study was in pregnant participants. Three studies evaluated only individuals with CD4 cell count ≤350 per µL; however, we recognize this is how patients may present in practice. Two studies evaluated only inpatients; however, sensitivity estimates were higher and specificity estimates were lower, but specificity was still high (90 and 95%) and may partly be because Xpert assay identifies patients with TB that the reference standard (culture) does not. We did not downgrade for indirectness.
c. The confidence intervals (CI) for sensitivity are sufficiently narrow (CI half width = 8) and the lower limit is not significantly lower than the lower limit and point estimate of WHO screen then Xpert strategy. The upper limit is significantly higher. Given that this may lead to small differences depending on which limits are assumed and that Xpert for all must have greater or equivalent sensitivity compared to WHO screen then Xpert, we did not downgrade for imprecision.
d. Sensitivity estimates ranged from 25% to 91% in all studies. Lower estimates were seen in pregnant and on ART populations and higher estimates were seen in inpatient studies;

however, this was not always the case and we could not always explain the variability. We downgraded one level for inconsistency.

e. Specificity estimates ranged from 97% to 100% in all but two studies done in inpatients where the specificity was 90% and 95% and may explain the variability. CIs also overlapped. We did not downgrade for inconsistency.

Test result	Number of results per 1000 patients tested (95% CI)						№ of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 5%		Prevalence 10%		Prevalence 20%			
	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD	molecular WHO-approved rapid diagnostics	WHO-recommended 4 symptom screen followed by mWRD		
True positives patients with active TB	34 (30 to 38)	31 (28 to 34)	69 (60 to 76)	62 (56 to 69)	138 (120 to 152)	124 (112 to 138)	9209 (14)	⊕⊕⊕◯ MODERATE[a,b,c,d]
	3 more TP in molecular WHO-approved rapid diagnostics		7 more TP in molecular WHO-approved rapid diagnostics		14 more TP in molecular WHO-approved rapid diagnostics			
False negatives patients incorrectly classified as not having active TB	16 (12 to 20)	19 (16 to 22)	31 (24 to 40)	38 (31 to 44)	62 (48 to 80)	76 (62 to 88)		
	3 fewer FN in molecular WHO-approved rapid diagnostics		7 fewer FN in molecular WHO-approved rapid diagnostics		14 fewer FN in molecular WHO-approved rapid diagnostics			
True negatives patients without active TB	931 (922 to 941)	941 (922 to 941)	882 (873 to 891)	891 (873 to 891)	784 (776 to 792)	792 (776 to 792)	9209 (14)	⊕⊕⊕⊕ HIGH[a,b,e]
	10 fewer TN in molecular WHO-approved rapid diagnostics		9 fewer TN in molecular WHO-approved rapid diagnostics		8 fewer TN in molecular WHO-approved rapid diagnostics			
False positives patients incorrectly classified as having active TB	19 (9 to 28)	9 (9 to 28)	18 (9 to 27)	9 (9 to 27)	16 (8 to 24)	8 (8 to 24)		
	10 more FP in molecular WHO-approved rapid diagnostics		9 more FP in molecular WHO-approved rapid diagnostics		8 more FP in molecular WHO-approved rapid diagnostics			
Inconclusive	undefined						(0)	-
Complications	undefined						(0)	-

a. Low risk of bias in all but one included studies. Flow and timing was at high risk of bias in that study. We did not downgrade.

b. Six studies were considered a concern for applicability. One study was in pregnant participants. Three studies evaluated only individuals with CD4 cell count ≤350 per µL; however, we recognize this is how patients may present in practice. Two studies evaluated only inpatients; however, sensitivity estimates were higher and specificity estimates were lower, but specificity was still high (90 and 95%) and may partly be because Xpert assay

	identifies patients with TB that the reference standard (culture) does not. We did not downgrade for indirectness. c. The confidence intervals (CI) for sensitivity are sufficiently narrow (CI half width = 8) and the lower limit is not significantly lower than the lower limit and point estimate of WHO screen then Xpert strategy. The upper limit is significantly higher. Given that this may lead to small differences depending on which limits are assumed and that Xpert for all must have greater or equivalent sensitivity compared to WHO screen then Xpert, we did not downgrade for imprecision. d. Sensitivity estimates ranged from 25% to 91% in all studies. Lower estimates were seen in pregnant and on ART populations and higher estimates were seen in inpatient studies; however, this was not always the case and we could not always explain the variability. We downgraded one level for inconsistency. e. Specificity estimates ranged from 97% to 100% in all but two studies done in inpatients where the specificity was 90% and 95% and may explain the variability. CIs also overlapped. 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	No direct evidence was considered here. No direct benefits, adverse effects or burden resulting directly from the test itself are anticipated.	

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	TB treatment is highly effective among PLHIV, and successful treatment outcomes comparable to those among people not living with HIV can be achieved (Owiti et al). Conversely, undetected TB and delays in diagnosis of TB and MDR-TB diagnosis among people living with HIV are associated with increased mortality as well as increased transmission. Findings from a systematic review (Gupta et al) of post mortem studies among people who had died from AIDS, highlighted close to 50% of post mortem diagnosis of TB had not been detected prior to death, and TB was the cause of death in more 91.4% (95% CI 85.8-97.0%) of all the TB cases. Should active TB not be identified through further investigations, programmatic management of latent TB infection and initiation of antiretroviral treatment would be the next course of action if eligible. A systematic review of 12 randomized controlled trials found that TB preventive treatment reduced the overall risk of TB by 33% (RR 0.67, 95% CI 0.51-0.87) as well as a reduction in mortality. Clinical trials confirm that early use of ART keeps people living with HIV alive and healthier and reduces the risk of transmitting the virus to others. Earlier treatment has the further advantage of simplifying the operational demands on programmes.	

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	mWRD tests are already recommended as the first TB diagnostic test for people living with HIV so its use as a screening test would be unlikely to be an additional burden in terms of linkage between test results and management decisions. The time from test to test result for mWRD tests is 90 minutes. If the instrument is not available at the clinic where the patient presents, a sample would need to be sent to the nearest facility where one is available. A randomized controlled trial (*Theron et al*) conducted in South Africa, Tanzania, Zambia and Zimbabwe found that Xpert MTB/RIF can be accurately administered as a diagnostic test by a nurse in primary-care clinics, resulting in more patients starting same-day treatment, more culture-positive patients starting therapy, and a shorter time to treatment. However, access to and linkage between the diagnostic test, confirmation and subsequent treatment initiation is dependent upon the	

	distribution, infrastructure, digital technology and colocation of respective services in relation to one another within the given health system, as well as location of where the person living with HIV first presents.	

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	No study of effectiveness of people living with HIV was sought.	

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	Screening aims to identify people with active TB earlier and therefore ensure earlier treatment and better health outcomes for individuals and lower TB transmission to community. No direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is no important uncertainty in what we know about how patients perceive and value the outcomes explored above.	

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	The panel determines that the balance between desirable and undesirable effects favours **using** mWRD as an alternative screening test to the W4SS followed by an mWRD test for detection of TB among people living with HIV, regardless of ART use.	

● 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 included studies directly addressed this subquestion. Data was available from three modeling studies that provided indirect evidence on the resources required for the use of mWRDs to screen PLHIV with Xpert, regardless of symptoms, prior to initiation of ART (Andrews 2012, Orlando 2018, Reddy 2019). No data was available on programmatic costs for the use of mWRDs to screen in PLHIV irrespective of signs and symptoms. Among all studies that reported the unit tests costs for Xpert (including equipment purchase, maintenance and consumable costs) the unit cost ranged from: US$24.80-$163.50. Note that from the studies reporting indirect evidence, Xpert unit costs represent a small proportion of the total implementation costs for an ACF program.	Large costs anticipated to detect just a few more cases, thus requiring a lot of mWRD tests. Opportunity costs are also high for the health system with a lot of people living with HIV so a lot of mWRD capacity would be required.

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	Three studies provide indirect evidence for the use of mWRDs for PTB in PLHIV, irrespective of symptoms, all three of which were of good quality.	

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	Indirect evidence was available for the cost-effectiveness of three studies on mWRDs for PTB in PLHIV, irrespective of symptoms, from three studies conducted in Malawi, Mozambique and South Africa. ICF interventions using mWRDs to screen for PTB among PLHIV were found to be cost-effective in all three studies with ICERs of US$410-$5,100 per year of life saved (Andrews 2012, Reddy 2019) or US$56 per DALY averted. Major drivers of cost identified were the prevalence of TB, costs of the Xpert and SSM, as well as the survival rate of HIV patients.	

comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ No included studies		

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	WHO-approved rapid molecular tests are currently recommended as the first diagnostic test for people living with HIV so its use as a screening test should have limited negative impact on equity. Increased sensitivity in a screen facilitates earlier detection of a debilitating impoverishing disease. WHO approved rapid molecular diagnostics are quantitative. Compared with the 4 symptom screen, the result and interpretation might be less influenced by personal judgement, response bias, discrimination or stigma. One literature review found that at-risk individuals report that fear of TB stigma and the social and economic impact of stigma affects their willingness to undergo TB screening - Tuberculosis and stigmatization: pathways and Interventions Public Health Rep. 2010; 125(Suppl 4): 34 42.doi: 10.1177/00333549101250S407PMCID: PMC2882973PMID: 20626191	● Varies - and dependent upon the supply and access to mWRDs. ● The intervention may detract from other vulnerable groups as regard to TB. ● It may impact upon use of mWRDs for diagnosis. ● Probably does not increase equity if compared with the four symptom screen which is available everywhere. ● Costs of mWRD are likely to be reduced with new tests in the pipeline so things might change. ● Scale-up, implementation and making it available free of charge would improve equity. ● If you find more cases and don't have adverse events from mWRD, equity

		would be increased. ● More resources would be required with mWRD as a screening tool - but if the resources are available equity would be increased as more people would be screened and diagnosed.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	To date the review team has found very little qualitative evidence relevant for this group, but drawing upon the theoretical findings from our wider body of evidence, we would suggest that perceived and enacted HIV and TB stigma will complicate the decision-making of persons living with HIV. (1 study, Tuot 2019 Cambodia, Total respondents N = 120; PLHIV, N = 6, Quality assessment not yet done).	● In HIV-positive people most people are aware that TB is a special risk. ● If mWRD is not acceptable then conducting the w4SS would not be meaningful, as diagnostic confirmation requires mWRD.

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ○ Yes ● Varies ○ Don't know	mWRD tests are already recommended by WHO as the primary mode of TB diagnosis. Use of screening would increase the workload for sample transportation and the facilities operating the diagnostic instruments. A randomized controlled trial (*Theron et al*) found that Xpert MTB/RIF can be accurately administered as a diagnostic test by a nurse in primary-care clinics, resulting in more patients starting same-day treatment, more culture-positive patients starting therapy, and a shorter time to treatment.	Depends on the context; operational challenges, access to machines and cartridges, and resources available to meet the increased needs and costs.

SUMMARY OF JUDGEMENTS

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	**Yes**		Varies	Don't know

	JUDGEMENT						
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**
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

WHO-approved rapid diagnostic tests may be used to screen for TB disease among adults and adolescents living with HIV **(conditional recommendation, moderate certainty of the evidence for test accuracy).**

Justification

- The W4SS followed by mWRD was found to have 0.62 (95% CI: 0.56 to 0.69) sensitivity and 0.99 (95% CI: 0.97 to 0.99) specificity in all PLHIV regardless of ART status, compared with 0.69 (95% CI: 0.60 to 0.76) sensitivity and 0.98 (95% CI: 0.97 to 0.99) specificity using mWRD alone.
- 9,046 study participants reviewed for mWRD as a TB screening tool were adults (>19) and 163 were adolescents (10-19).
- Due to the increased sensitivity but potential challenges for some countries relating to access, feasibility and costs, the GDG made a conditional recommendation in favour of mWRDs to be used as an option for screening for active TB among adults and adolescents living with HIV who are not medical inpatients in settings where TB prevalence exceeds 10% (for whom a strong recommendation exists to systematically test for TB using mWRD).
- TB prevalence according to the IPD reviewed in the 14 included studies ranged from 1% to 26%. The average TB prevalence among outpatients was 8.6%. In lower-prevalence settings there would be a higher number of false positives.
- Data on mWRD from the individual patient data meta-analysis were primarily based on studies that used Xpert MTB/RIF.

Subgroup considerations

- Depending on feasibility and resources available, countries may choose to prioritize TB screening using mWRD among certain subpopulations, such as medical inpatients living with HIV in settings where TB prevalence is less than 10%, among those who are acutely unwell, or pregnant women living with HIV.

Implementation considerations

- The GDG recommended mWRD as an initial TB screening test among people living with HIV, and that it should be subject to a diagnostic confirmation to inform a decision to treat for TB.
- Screening with mWRD in lower prevalence settings than those included in the individual patient data meta-analysis will result in higher numbers of false positives with overtreatment and related social and economic consequences including potential delay in start of ART should diagnosis not be confirmed. This recommendation may not be applicable in settings with a low pre-test probability of TB.
- Due consideration should be made to prioritization of provision of mWRD as a diagnostic test for all people with presumptive TB, before scaling up mWRD as a screening test. Use of mWRD as a screening tool would require significant resources, including increased capacity and expansion of diagnostic and sample transportation networks to allow implementation.
- Data were not available for determining optimal periodicity but the GDG suggested a pragmatic approach that is aligned with HIV service delivery and provided as part of annual check-ups, at the time of viral load monitoring, depending on the context, TB prevalence, and feasibility.
- The W4SS should always be conducted at all health visits, as part of a comprehensive clinical evaluation to inform the need for increased infection control and for other investigations such as LF-LAM, and to help inform any decision to treat for latent or active TB.
- Careful clinical assessment is recommended to ensure TB is the primary cause of illness and other conditions leading to the clinical presentation are also managed.
- A negative mWRD does not exclude TB. Patients in such settings who are mWRD-negative but are manifestly sick, may not be able to produce a quality sputum sample or may have extrapulmonary TB.
- Should a patient be unable to provide sputum, other biological specimens should be considered as indicated.
- For patients with a prior history of TB in the past 2 years, a positive result may be due to the presence of DNA detected from previously treated TB.

Monitoring and evaluation

- Countries are encouraged to monitor and evaluate the yield of TB screening among people living with HIV, disaggregated by screening tools to inform programming and resource planning.

Research priorities

- Well designed clinical trials to strengthen the evidence on the accuracy, effectiveness (including impact on patient-important outcomes e.g. mortality), feasibility and cost implications of using mwRD to screen for TB across all HIV sub-populations in low, medium and high HIV and TB burden settings with and without high ART coverage, compared with other screening strategies.
- Subpopulations of PLHIV for further investigation would include but not be limited to inpatients, acute care service attendees, patients failing ART, newly diagnosed HIV patients enrolling in ART clinics, stable patients established on ART, pregnant women, children and adolescents living with HIV and key populations.

- Frequency of screening - effectiveness, cost effectiveness, feasibility and acceptability and optimal periodicity of routine regular screening with mWRD among PLHIV in care.
- Research to evaluate the effectiveness and accuracy of combining mWRD and LF-LAM with repeat testing with mWRD and other tests for diagnostic confirmation.
- Screening with mWRD using specimens other than sputum.
- Placement of mWRD for screening in antenatal care settings versus screening in ART clinics.

Table 10. What tools should be used to screen for TB disease in children at high risk of TB (child contacts and children living with HIV)?

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	The WHO's End TB Strategy envisions a 90% reduction in TB incidence and 95% reduction in TB deaths by 2035, and the Resolution adopted by the United Nations General Assembly in September 2018 commits to diagnosing and treating 40 million people with TB - *including 3.5 million children* - and treating at least 30 million people for TB infection - *including 4 million children under 5 years of age* - by 2022. These targets will not be met by current practices in case detection. In order to achieve these ambitious targets there is an urgent need to deploy strategies to improve detection of TB among children. Approximately 1.2 million children less than 15 years developed active TB in 2019, 47% of which were under 5 years old, with an estimated 230,000 child deaths from TB, 80% of which were under 5 years of age and 96% were among children who did not access treatment. The case detection gap among children under 5 is estimated at 65%. Major factors lead to underdiagnosis of tuberculosis in children including: symptoms tend to be less specific in children and overlap with those of other common childhood diseases, diagnostic tests for children have sub-optimal sensitivity leading to reliance on clinical diagnosis, young children mostly present to primary health services where there is little or no awareness and capacity for TB diagnosis or treatment. Further guidance is needed on optimal screening strategies among children who are eligible for screening, in the healthcare setting and among contacts.	

Test accuracy

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very inaccurate ○ Inaccurate ○ Accurate ○ Very accurate ○ Varies ○ Don't know	<u>Individuals under 15 years accessing health care</u> **1. Test accuracy: mWRDs** Molecular WHO-approved Rapid Diagnostics (mWRDs) Sensitivity: 0.43 to 1.00 Specificity: 0.98 to 0.99 <u>Individuals under 15 years HIV outpatients:</u> **2. Test accuracy: Symptom screen** Symptom screen (current cough, fever, poor weight gain, or TB contact for children; current cough, weight loss, night sweats, or fever for adolescents) Sensitivity: 0.61 (95% CI: 0.58 to 0.64) Specificity: 0.94 (95% CI: 0.86 to 0.98) <u>Individuals under 15 years who are TB close contacts:</u> **3. Test accuracy: abnormal CXR** Chest radiography (suggestive of TB) Sensitivity: 0.84 (95% CI: 0.70 to 0.92) Specificity: 0.91 (95% CI: 0.90 to 0.92)	*GDG Judgements for each tool:* mWRDs in children accessing healthcare: Varies (from accurate to inaccurate) ICF symptom screen in children LHIV: Accurate CXR in child contacts: Accurate Symptoms in child contacts: Accurate The group noted concerns that the data presented in the review do not accurately portray the performance of the tests in practice, due to potential bias incorporating the reference standard (when composite reference standard is used). There is also concern about what

4. Test accuracy: cough, fever, poor weight gain

Symptom screening involving any one of cough, fever, or poor weight gain

Sensitivity: 0.89 (95% CI: 0.52 to 0.98) Specificity: 0.69 (95% CI: 0.51 to 0.83)

happens after the screening test is applied in this population, as children are more likely to go on to clinical evaluation to establish or rule out a diagnosis, including applying clinical criteria (typical signs and symptoms, history of contact, often combined with chest radiography, non-response to broad spectrum antibiotics) to make a clinical diagnosis of TB, with a higher risk of a false-positive diagnosis following a false-positive screening test in children compared to adults.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ○ Moderate ○ Large ● Varies ○ Don't know	**Desirable: Identifying true positive and negatives:** The anticipated desirable effect is the correct classification of children and adolescent as screening positive for TB (true positive), resulting in appropriate referral for further evaluation (as indicated by the screening algorithm in use), leading to timely diagnosis and treatment and reducing further transmission of the mycobacterium. Another anticipated desirable effect is the identification of children who do not have active TB and are eligible for TB preventive treatment (true negative). **Undesirable: Identifying false positive and negatives:** The anticipated undesirable effect is the incorrect classification of those without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences. Another anticipated undesirable effect is the incorrect classification of active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, and potentially inappropriate preventive therapy treatment with associated risk of development of drug resistance.	*Judgements for each tool:* mWRDs in children accessing healthcare : Varies ICF symptom screen in children LHIV: Moderate desirable effects CXR in child contacts: Moderate desirable effects Symptoms in child contacts: Moderate desirable effects

Population/ setting	Test	Test accuracy	Studies (persons)	Certainty in evidence	Lower prev (0.5%)	Middle prev (5%)	Higher prev (10%)
Accessing healthcare	MRS Xpert MTB/RIF	Sens: 43 to 100*	2(16)	Very low	TP: 2-5/ FN: 0-3	TP: 22-50/ FN: 0-28	TP: 43-100/ FN: 0-57
Accessing healthcare		Spec: 98 to 99	2 (771)	Moderate	TN: 975-985/ FP: 10-20	TN:931-941 / FP: 9-19	TN: 882-891/ FP: 9-18
Close TB contacts	CRS Suggestive CXR	Sens: 84 (70 to 92)	4 (113)	Low	TP: 3-5 / FN: 0-2	TP:35-46 / FN: 4-15	TP:70-92 / FN: 8-30
Close TB contacts		Spec: 91 (90 to 92)	4 (2437)	Moderate	TN:896-915/FP:80-99	TN:855-874/FP:76-95	TN:810-828/FP:72-90
Close TB contacts	CRS Cough, fever, or poor weight gain	Sens: 89 (52 to 98)	4 (113)	Low	TP: 3-5 / FN: 0-2	TP: 26-49 / FN: 1-24	TP: 52-98 / FN: 2-48
Close TB contacts		Spec: 69 (51 to 83)	4 (2582)	Low	TN:507-826/FP:169-488	TN:485-789/FP:161-465	TN:459-747/FP:153-441
Outpatients living with HIV	CRS ICF symptom screen	Sens: 61 (58 to 64)	2 (1219)^	Moderate	TP: 3-3 / FN: 2-2	TP: 29-32 / FN: 18-21	TP: 58-64 / FN: 36-42
Outpatients living with HIV		Spec: 94 (86 to 98)	2 (201916)	Low	TN:856-975/FP:20-139	TN:817-931/FP:19-133	TN:774-882/FP:18-126

*Reported as range from studies as meta-analysis did not converge and pooled estimates could not be obtained

^Reported as number of screens rather than number of persons

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS

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

Desirable: Identifying true positive and negatives:

The anticipated desirable effect is the correct classification of children and adolescent as screening positive for TB (true positive), resulting in appropriate referral for further evaluation (as indicated by the screening algorithm in use), leading to timely diagnosis and treatment and reducing further transmission of the mycobacterium. Another anticipated desirable effect is the identification of children who do not have active TB and are eligible for TB preventive treatment (true negative).

Undesirable: Identifying false positive and negatives: The anticipated undesirable effect is the incorrect classification of those without active TB as positive for TB (false-positive), with resulting inappropriate treatment and related social or economic consequences. Another anticipated undesirable effect is the incorrect classification of active TB as negative for TB (false-negative), resulting in a missed opportunity for appropriate treatment, further treatment delay, continued transmission, and potentially inappropriate preventive therapy treatment with associated risk of development of drug resistance.

Population/ setting	Test	Test accuracy	Studies (persons)	Certainty in evidence	Lower prev (0.5%)	Middle prev (5%)	Higher prev (10%)
Accessing healthcare	MRS Xpert MTB/RIF	Sens: 43 to 100*	2(16)	Very low	TP: 2-5/ FN: 0-3	TP: 22-50/ FN: 0-28	TP: 43-100/ FN: 0-57
Accessing healthcare		Spec: 98 to 99	2 (771)	Moderate	TN: 975-985/ FP: 10-20	TN:931-941 / FP: 9-19	TN: 882-891/ FP: 9-18
Close TB contacts	CRS Suggestive CXR	Sens: 84 (70 to 92)	4 (113)	Low	TP: 3-5 / FN: 0-2	TP:35-46 / FN: 4-15	TP:70-92 / FN: 8-30
Close TB contacts		Spec: 91 (90 to 92)	4 (2437)	Moderate	TN:896-915/FP:80-99	TN:855-874/FP:76-95	TN:810-828/FP:72-90
Close TB contacts	CRS Cough, fever, or poor weight gain	Sens: 89 (52 to 98)	4 (113)	Low	TP: 3-5 / FN: 0-2	TP: 26-49 / FN: 1-24	TP: 52-98 / FN: 2-48
Close TB contacts		Spec: 69 (51 to 83)	4 (2582)	Low	TN:507-826/FP:169-488	TN:485-789/FP:161-465	TN:459-747/FP:153-441
Outpatients living with HIV	CRS ICF symptom screen	Sens: 61 (58 to 64)	2 (1219)^	Moderate	TP: 3-3 / FN: 2-2	TP: 29-32 / FN: 18-21	TP: 58-64 / FN: 36-42
Outpatients living with HIV		Spec: 94 (86 to 98)	2 (201916)	Low	TN:856-975/FP:20-139	TN:817-931/FP:19-133	TN:774-882/FP:18-126

*Reported as range from studies as meta-analysis did not converge and pooled estimates could not be obtained

^Reported as number of screens rather than number of persons

Judgements for each tool:

mWRDs in children accessing healthcare : Varies

ICF symptom screen in children LHIV : Moderate undesirable effects (as half of those with TB are falsely negative, with a missed opportunity for diagnosis and treatment)

CXR in child contacts : Small undesirable effects

Symptoms in child contacts : Small undesirable effects

There was concern, specifically with the poor sensitivity children living with HIV, about the low sensitivity and thus a large proportion of false negative screening tests possibly leading to missed diagnosis - it was felt that, for children, a less sensitive test was more of a problem compared to adults as there is more urgency for timely diagnosis. There was also concern that the estimate of false negative and false positive results could be incorrect due to bias from the reference standard, as mentioned above. There was a thought that better strategies for screening children with HIV are needed.

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 is as follows: <u>Individuals under 15 years accessing health care</u> **1. Test accuracy: mWRDs - *Very low/moderate*** molecular WHO-approved Rapid Diagnostics Sensitivity: 0.43 to 1.00 Specificity: 0.98 to 0.99 <u>Individuals under 15 years HIV outpatients:</u>	

	2. Test accuracy: ICF symptom screen - *Moderate/low* Symptom screen (current cough, fever, poor weight gain, or TB contact for children; current cough, weight loss, night sweats, or fever for adolescents) Sensitivity: 0.61 (95% CI: 0.58 to 0.64) Specificity: 0.94 (95% CI: 0.86 to 0.98) <u>Individuals under 15 years who are TB close contacts:</u> **3. Test accuracy: abnormal CXR - *Low/low*** chest radiography (suggestive of TB) Sensitivity: 0.84 (95% CI: 0.70 to 0.92) Specificity: 0.91 (95% CI: 0.90 to 0.92) **4. Test accuracy: cough, fever,poor weight gain - *Moderate/moderate*** symptom screening involving any one of cough, fever, or poor weight gain Sensitivity: 0.89 (95% CI: 0.52 to 0.98) Specificity: 0.69 (95% CI: 0.51 to 0.83)	

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; the direct benefits and adverse effects of the screening tests are unknown. The possible adverse effects of any screening test include additional burden of anxiety and time for the patient. The possible benefits of symptom screening for TB including detection of other conditions. Direct CXR is a safe technology using a radiation dose of 0.1 mSv, which corresponds to 1/30 of the average annual radiation dose from the environment (3 mSv) and 1/10 of the annual accepted dose of ionizing radiation for the general public (1 mSv). Innovations in x-ray technology in recent years have substantially reduced the radiation exposure levels. A proportion of the X-rays used in radiography are absorbed by the body. The potential effects from ionizing radiation depend on the dose. The long term risks from ionizing radiation include an increased risk of cancer. Therefore, exposure to the low radiation doses delivered to patients during a CXR poses a small risk. Children and pregnant women are especially vulnerable to ionizing radiation. Also, children have a longer life expectancy, resulting in a larger window for developing long-term radiation-induced health effects. The direct benefit or adverse effect of molecular WRD testing are unknown. Producing sputum not easy or pleasant for patients.	The GDG noted that less is known about the direct effects of screening tests for children than for adults.

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	The primary objective of screening for active TB is to ensure that active TB is detected early and treatment is initiated promptly, with the ultimate aim of reducing the risk of poor treatment outcomes, health sequelae and the adverse social and economic consequences of TB, as well as helping to reduce TB transmission. Without appropriate treatment, children with TB have a very high risk of dying. WHO recommends that children aged < 5 years who are household contacts of people with bacteriologically confirmed pulmonary TB and who are found not to have active TB should be given TB preventive treatment (Strong recommendation, high certainty in the estimates of effect). Those ≥5 years who are household contacts of people with bacteriologically confirmed pulmonary TB who are found not to have active TB may be given TB preventive treatment. (Conditional recommendation, low certainty in the estimates of effect). People living with HIV who are unlikely to have active TB should receive TB preventive treatment as part of a comprehensive package of HIV care (Strong recommendation, high certainty in the estimates of effect). Treatment of drug sensitive TB is highly effective (approximately 90%). Treatment of MDR TB can be effective as well, if quality assured (approximately 70%) (WHO consolidated treatment guidelines 2020). Effectiveness of preventive therapy is between 60-90% (WHO preventive treatment guidelines 2020). See here for evidence - https://tuberculosis.evidenceprime.com/	

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	No included studies reported directly on linkage to care among children receiving screening.	Although there were no studies focused on this subject, most health care services attach importance to an evaluation suggestive of TB and will investigate further or consider starting treatment

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	No included studies.	

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 direct evidence of values and preferences regarding how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results outcomes. There is important uncertainty in what we know about how patients perceive and value the outcomes explored above.	The GDG noted that it can be assumed that children and their caregivers would value an accurate screening test.

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	Individuals under 15 years accessing health care **1. Test accuracy: mWRDs** molecular WHO-approved Rapid Diagnostics Sensitivity: 0.43 to 1.00 Specificity: 0.98 to 0.99 • The panel determines that the balance between desirable and undesirable effects **varies** for mWRDs to screen for active pulmonary TB in children and adolescents accessing healthcare. Individuals under 15 years HIV outpatients: **2. Test accuracy: Symptom screen** Symptom screen (current cough, fever, poor weight gain, or TB contact for children; current cough, weight loss, night sweats, or fever for adolescents) Sensitivity: 0.61 (95% CI: 0.58 to 0.64) Specificity: 0.94 (95% CI: 0.86 to 0.98) • The panel determines that the balance between desirable and undesirable effects **probably favors using** WHO symptom screen to screen for active pulmonary TB in children and adolescents living with HIV in an outptainet setting. Individuals under 15 years who are TB close contacts: **3. Test accuracy: suggestive CXR** Chest radiography (suggestive of TB) Sensitivity: 0.84 (95% CI: 0.70 to 0.92) Specificity: 0.91 (95% CI: 0.90 to 0.92) • The panel determines that the balance between desirable and undesirable effects **favors using** chest radiography (suggestive) to screen for pulmonary TB in children who are close contacts of someone with TB. **4. Test accuracy: Symptom screen (cough, fever, poor weight gain)** Symptom screening involving any one of cough, fever, or poor weight gain Sensitivity: 0.89 (95% CI: 0.52 to 0.98) Specificity: 0.69 (95% CI: 0.51 to 0.83) • The panel determines that the balance between desirable and undesirable effects **favors using** symptoms (cough, fever, poor weight gain) to screen for active pulmonary TB in children who are close contacts of someone with TB.	*Judgements for each tool:* mWRDs in children accessing healthcare: Varies (the lack of studies in this area highlight that this is a research gap) Symptom screen in children LHIV : Probably favors CXR in child contacts: Favors Symptoms in child contacts: Favors

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	Resources required: No studies directly compared the costs of approaches for screening among children. Two studies addressed the question of the resources required for systematic screening for TB among children (Htet 2017, Mupere 2013). The average cost per child screened ranged from US$2.88-$4.13 (Htet 2017, Mupere 2013).	The GDG felt that the resources required varied, depending on the screening and diagnostic algorithm used and the diagnostic test used. Some of the populations under consideration (contacts, children living with HIV) are relatively small so the costs for screening would not be large. Screening using CXR is more resource intensive. Screening much larger patient

		groups (such as all children accessing healthcare) would be more resource intensive.

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	The small amount of evidence is a limitation, with differences in methodological approaches and heterogeneity of results which limits comparability and generalizability to other settings. All three studies that contributed to this sub-question were of good quality.	

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 intervention ○ Varies ● No included studies	Indirect evidence of cost effectiveness of screening among children: The cost per child diagnosed with TB ranged from US$18-$42 (Malik 2019, Htet 2017). One study in Uganda found that door-to-door symptom screening in children was cost-effective with an ICER of US$538 per QALY (Mupere 2013).	The studies found do not directly address the populations under consideration.

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 included studies.	The group felt it is likely screening would increase equity in this population, given the accuracy data, but the lack of data indicates that this is an important research gap.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS

○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	In a previous TB screening acceptability study, a median of 86% (range 84-91%) of children approached for screening (their caretakers) accepted to participate (WHO TB screening guidelines 2013). A qualitative evidence synthesis (QES) was conducted to understand individual responses to ACF and systematic TB screening. Key themes specific to children were extracted from nine studies and include this synthesized statement: • *Parents may prioritize providing food, retaining employment and sustaining community networks over engaging TB services.* • *Children may have non-specific symptoms, and parents may avoid TB care until children are ill.* • *Parents and carers want to avoid TB and HIV stigma; they also know that the TB care pathway is inconvenient and time consuming and want to avoid TB medications.* Evidence relevant to decision-making about children echoes many themes across the wider body of evidence for the QES. Parents and caregivers prioritize providing food and nutrition, retaining their own employment and sustaining family and community networks over engagement with TB programmes. Parents may avoid TB services due to anticipated TB or HIV stigma. Children may have non-specific symptoms, and parents may engage care only when children are very ill. Parents and carers know that systems for follow up TB diagnosis and treatment are inconvenient and time consuming. A mother who was familiar with TB asked specifically for an X ray to ensure diagnosis of her child. A single study found confusion regarding IPT and breastfeeding amongst parents and community health workers; parents in the same study wanted to avoid TB medications.	

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 was identified. Feasibility of screening by different approaches: • Screening with symptoms is very feasible • Screening with chest radiography requires access to radiography, in a health facility or a mobile screening unit. Basic radiography is an essential technology in primary care, and in recent years mobile digital basic X-ray systems have been developed that can provide radiology services outside the clinic setting, including in mobile vans. However, access to high-quality radiography is limited in many settings, including access to trained personnel to read images. • mWRDs are recommended by WHO as the primary mode of TB diagnosis. mWRDs are able to be implemented at most levels of health care but resource and availability may prohibit use for screening at scale	This varies and depends on population (age and risk factors of the children being screened), the screening being considered, and the volume of patients to be screened. mWRDs in children accessing healthcare: Probably no, this would be difficult given access to the test platforms and difficulty in obtaining samples from this patient population ICF symptom screen in children LHIV: Yes, feasible CXR in child contacts: Probably yes, depending on access to radiography and the ability to read the images Symptoms in child contacts: Yes, feasible

SUMMARY OF JUDGEMENTS

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	**Yes**		Varies	Don't know

	JUDGEMENT						
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**
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

Children and adolescents living with HIV: screening using a symptom screen:

- *Existing recommendation*: Adolescents 10-19 years living with HIV should be screened for TB with a clinical algorithm; those who report any symptoms of TB including current cough, fever, weight loss, or night sweats, may have TB disease and should be evaluated for TB and other diseases **(strong recommendation, moderate quality of evidence)**.
- *New recommendation*: Among children under 10 years living with HIV, the GDG recommends using a clinical algorithm specific to children including: TB including current cough, fever, poor weight gain, or close contact with a TB patient as a screening test for detection of TB **(strong recommendation, low certainty of the evidence for test accuracy).**

Symptom and CXR screening among TB close contacts:

- Among individuals under 15 years who are close contacts with someone with TB, the GDG recommends using chest radiography as a screening test for detection of TB **(strong recommendation, low certainty of the evidence for test accuracy).**
- Among individuals under 15 years who are close contacts with someone with TB, the GDG recommends using symptoms (cough, fever, poor weight gain) as a screening test for detection of TB **(strong recommendation, moderate certainty of the evidence for test accuracy).**

Individuals under 15 years accessing health care - screening using mWRDs:

The GDG decided not to issue a recommendation on this question and highlighted this as a research gap: more data is needed in order to address this question.

Justification

- Children and adolescents living with HIV represent an important group for regular TB screening and provision of preventive therapy, given their high risk of TB and of poor outcomes if not diagnosed in a timely manner. Regular screening for TB symptoms at each visit to the health center is essential in this group as a minimum screening strategy.
- Child contacts likewise represent an important group for screening for TB disease and initiation of preventive therapy, given their high risk for TB, and it is essential that screening be conducted in this group using the most accurate screening tools available and feasible.

Subgroup considerations

- Child contacts and children living with HIV are essential groups for screening using the most accurate tools available and feasible. For these populations of children, while existing tools should be used, better screening tools are urgently needed.
- Children and adolescents under 15 years of age accessing healthcare represent a much larger potential population for screening with important resource implications for scaling up screening, particularly with more expensive screening and diagnostic tools. In addition, with this patient population, given the lower pre-test probability of disease and the diagnostic pathway that children follow after screening, there is concern that screening tests with low specificity could lead to high numbers of children receiving false positive diagnoses and inappropriate treatment, therefore caution is warranted.

Implementation considerations

Feasibility of implementation is dependent on the population being screened (including the age and risk factors of the children), the location where screening is being conducted, the screening and diagnostic tests being used, and the volume of patients being screened. See general recommendations for screening of people living with HIV and close contacts, and general recommendations for symptom screening and chest radiography, for further implementation considerations.

Linkage to care and TB preventive treatment upon negative/normal screen, or upon ruling out TB through diagnostic evaluation, is essential for both populations of children considered here.

Importantly, implementation of screening for children may be impacted by parental and caregiver concerns and need to be managed with sensitivity and care by health care providers, health care managers and health programmes. Several factors may affect likelihood and willingness to engage with screening programmes:

· Parents and caregivers may prioritize providing food and retaining employment rather than attending a clinic for screening

· Children's signs and symptoms are often non-specific and parents may only attend facilities when their child is severely ill.

· Issues of TB and HIV stigma, and concerns with engaging in an inconvenient and time-consuming screening process still pose a barrier to parents and caregivers attending clinics.

Monitoring and evaluation

Countries are encouraged to monitor and evaluate the yield of TB screening approaches among children to be screened, including child close contacts and children living with HIV, disaggregated by screening tools and approaches, to broaden the evidence base of the yield, costs, safety, and clinical outcomes of different strategies.

Research priorities

- Further research is needed from well-designed clinical trials to provide evidence on patient-important outcomes for TB screening in children and adolescents, across the spectrum of screening tools and sub-populations, and on the effectiveness of different screening algorithms in the subpopulations of children and adolescents at highest risk of TB, including close contacts, children living with HIV, and malnourished children.
- More data is needed on the frequency with which screening should be conducted among subpopulations of children at highest risk of TB.
- More data is needed to establish the potential role of MWRDs in screening for children, and for more possible screening tests for children living with HIV.
- The potential impact of different screening algorithms on issues of equity, costs, feasibility, and acceptability to all stakeholders is needed.
- More data on other screening approaches, specific to distinct age ranges, including: infants under 12 months, children under 5 years of age, children up to 10 years of age, and adolescents (ages 10-19), are needed.

For further information, please contact:

World Health Organization
20, Avenue Appia CH-1211 Geneva 27 Switzerland
Global TB Programme
Web site: www.who.int/tb

World Health Organization

9789240022713