WHO consolidated guidelines on tuberculosis

Module 2: Screening

Systematic screening for tuberculosis disease

Web Annex B.
GRADE Summary of Findings Tables
WHO consolidated guidelines on tuberculosis

Module 2: Screening

Systematic screening for tuberculosis disease

Web Annex B.
GRADE Summary of Findings Tables
### Table 1. Should systematic screening for TB disease, compared to passive case detection, be conducted in the general population? (individual-level outcomes)

<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Nº of patients</th>
<th>Effect</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nº of patients</td>
<td>Effect</td>
<td>Certainty</td>
</tr>
<tr>
<td></td>
<td>Nº of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
</tr>
<tr>
<td>Treatment outcome: treatment success (cured + treatment completed)</td>
<td>3</td>
<td>observational studies</td>
<td>serious</td>
</tr>
<tr>
<td>Treatment outcome: case fatality</td>
<td>4</td>
<td>observational studies</td>
<td>serious</td>
</tr>
<tr>
<td>Earlier case detection: severity at diagnosis – smear grade (proportion 2+ and 3+)</td>
<td>3</td>
<td>observational studies</td>
<td>serious</td>
</tr>
<tr>
<td>Linkage to care – initial default</td>
<td>2</td>
<td>observational studies</td>
<td>serious</td>
</tr>
</tbody>
</table>

CI: Confidence interval; ACF: Active case-finding; PCF: Passive case-finding

**Explanations**

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. 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.

f. 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.

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.

k. Population: the study population in both studies were smear-positive TB cases.

l. Sample sizes are relatively large.

m. 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).

References


Table 2. Should systematic screening for TB disease, compared to passive case detection, be conducted in the general population? (community-level outcomes)

<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>№ of patients</th>
<th>Effect</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>№ of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
</tr>
<tr>
<td>TB disease prevalence (ZAMSTAR) (follow up: 4.5 years)</td>
<td>1</td>
<td>randomised trials</td>
<td>serious (^a)</td>
</tr>
<tr>
<td>TB disease prevalence (ACT3) (follow up: 3 years)</td>
<td>1</td>
<td>randomised trials</td>
<td>serious (^a)</td>
</tr>
<tr>
<td>TB disease prevalence (DETECTB)</td>
<td>1</td>
<td>observational studies</td>
<td>serious (^c)</td>
</tr>
<tr>
<td>TB disease prevalence (other non-randomised studies)</td>
<td>1</td>
<td>observational studies</td>
<td>very serious (^h)</td>
</tr>
<tr>
<td>Case notification rate (DETECTB)</td>
<td>1</td>
<td>randomised trials</td>
<td>not serious (^k)</td>
</tr>
<tr>
<td>No. of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>4</td>
<td>observational studies</td>
<td>very serious (^a)</td>
<td>very serious (^b)</td>
</tr>
</tbody>
</table>

**Case notification rate (non-randomized studies)**

<table>
<thead>
<tr>
<th></th>
<th>No. of patients</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design</td>
<td>No. of patients</td>
<td>Relative (95% CI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TST positivity in children (ZAMSTAR)</td>
<td><strong>RR 1.36</strong> (0.59 to 3.14)</td>
<td><strong>24 more per 1,000</strong> (from 27 fewer to 142 more)</td>
</tr>
<tr>
<td>randomised trials</td>
<td>391/4934 (7.9%)</td>
<td>342/5169 (6.6%)</td>
</tr>
<tr>
<td>not serious</td>
<td>391/4934 (7.9%)</td>
<td>342/5169 (6.6%)</td>
</tr>
<tr>
<td>serious (^{s})</td>
<td>391/4934 (7.9%)</td>
<td>342/5169 (6.6%)</td>
</tr>
<tr>
<td>not serious</td>
<td>391/4934 (7.9%)</td>
<td>342/5169 (6.6%)</td>
</tr>
<tr>
<td>not serious</td>
<td>391/4934 (7.9%)</td>
<td>342/5169 (6.6%)</td>
</tr>
<tr>
<td>none</td>
<td>391/4934 (7.9%)</td>
<td>342/5169 (6.6%)</td>
</tr>
</tbody>
</table>

**IGRA positivity in children (ACT3)**

<table>
<thead>
<tr>
<th>Study design</th>
<th>No. of patients</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>randomised trials</td>
<td>18/705 (2.6%)</td>
<td>32/779 (4.1%)</td>
</tr>
<tr>
<td>not serious</td>
<td>18/705 (2.6%)</td>
<td>32/779 (4.1%)</td>
</tr>
<tr>
<td>serious (^{s})</td>
<td>18/705 (2.6%)</td>
<td>32/779 (4.1%)</td>
</tr>
<tr>
<td>not serious</td>
<td>18/705 (2.6%)</td>
<td>32/779 (4.1%)</td>
</tr>
<tr>
<td>not serious</td>
<td>18/705 (2.6%)</td>
<td>32/779 (4.1%)</td>
</tr>
<tr>
<td>none</td>
<td>18/705 (2.6%)</td>
<td>32/779 (4.1%)</td>
</tr>
</tbody>
</table>

CI: Confidence interval; RR: Risk ratio

**Explanations**

- a. 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.
- 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. ZAMSTAR used community mobilization, education and sputum drop off points (mobile sputum collection points and ‘fast track’ at permanent facilities). Importantly ZAMSTAR used smear microscopy as the primary diagnostic tool. 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. Denominator refers to number of adults who gave informed consent, completed questionnaire and provided a sputum sample that was evaluable.
- 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. 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.

g. Denominator is number of adults (selected at random from intervention areas) who were located, consented to be surveyed and provided sputum.

h. Assessed using ROBINS-i. Multiple issues identified, including no accounting for confounding or temporal trends in TB case notifications.

i. No confidence interval provided.

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). 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. Risk of bias assessed using ROBINS-i (slightly modified), 3 studies at moderate ROB, 5 at serious ROB and 5 at critical ROB.

n. Differences in effect size and direction of effect.

o. Different studies used different methods of ACF.

p. In general, no measures of uncertainty (confidence intervals) available.

q. We are aware of a body of unpublished literature around ACF interventions.

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.

References


Table 3. Should systematic screening for TB disease, compared to passive case detection, be conducted among household and close contacts of individuals with TB disease?

<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Nº of patients</th>
<th>Effect</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nº of patients</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>Death (follow up: 2 years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>TB prevalence ratio (follow up: 4.5 years; assessed with: culture confirmed TB among adults)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Case notification (follow up: 2 years; assessed with: Cases registered with NTP)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Case detection (assessed with: Microbiologically confirmed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>serious (^{k})</td>
<td>not serious</td>
</tr>
<tr>
<td>Co-prevalent TB cases detected among contacts of any bacteriologically-confirmed index patients (assessed with: Case detection)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>observational studies</td>
<td>very serious (^{m})</td>
<td>very serious (^{h})</td>
</tr>
<tr>
<td>Co-prevalent TB cases detected among contacts of MDR/XDR index patients (assessed with: Case detection)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>observational studies</td>
<td>very serious (^{m})</td>
<td>very serious (^{h})</td>
</tr>
<tr>
<td>Co-prevalent TB cases detected among contacts (All TB cases) (assessed with: Case detection)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>observational studies</td>
<td>very serious (^{m})</td>
<td>very serious (^{h})</td>
</tr>
<tr>
<td>Co-prevalent TB cases detected among contacts (&lt;5 years old) (assessed with: Case detection)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>observational studies</td>
<td>very serious (^{m})</td>
<td>very serious (^{h})</td>
</tr>
</tbody>
</table>
### Co-prevalent TB cases detected among contacts (5–14 years old) (assessed with: Case detection)

<table>
<thead>
<tr>
<th>№ of patients</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contacts with TB = 283</td>
</tr>
<tr>
<td></td>
<td>Contacts screened = 14,622</td>
</tr>
<tr>
<td></td>
<td>Weighted pooled prevalence = 2.5% (1.7–3.5%)</td>
</tr>
<tr>
<td></td>
<td>Median NNS = 36 (17–61) (n=16)</td>
</tr>
</tbody>
</table>

### Co-prevalent TB cases detected among HIV infected contacts (assessed with: Case detection)

<table>
<thead>
<tr>
<th>№ of patients</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contacts with TB = 149</td>
</tr>
<tr>
<td></td>
<td>Contacts screened = 1,696</td>
</tr>
<tr>
<td></td>
<td>Weighted pooled prevalence = 11.7% (7.0–17.2%)</td>
</tr>
<tr>
<td></td>
<td>Median NNS = 24 (17–28)</td>
</tr>
</tbody>
</table>

**Certainty assessment**

<table>
<thead>
<tr>
<th>Nº of studies</th>
<th>Study design</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
<th>№ of patients</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>observational studies</td>
<td>very serious</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>Contacts with TB = 283</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Contacts screened = 14,622</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weighted pooled prevalence = 2.5% (1.7–3.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Median NNS = 36 (17–61) (n=16)</td>
<td></td>
</tr>
</tbody>
</table>

**Certainty**

<table>
<thead>
<tr>
<th>Nº of studies</th>
<th>Study design</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
<th>№ of patients</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>observational studies</td>
<td>very serious</td>
<td>very serious</td>
<td>not serious</td>
<td>serious</td>
<td>none</td>
<td>Contacts with TB = 149</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Contacts screened = 1,696</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weighted pooled prevalence = 11.7% (7.0–17.2%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Median NNS = 24 (17–28)</td>
<td></td>
</tr>
</tbody>
</table>

**Certainty**

<table>
<thead>
<tr>
<th>Nº of studies</th>
<th>Study design</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
<th>№ of patients</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>observational studies</td>
<td>very serious</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>Contacts with TB = 283</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Contacts screened = 14,622</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weighted pooled prevalence = 2.5% (1.7–3.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Median NNS = 36 (17–61) (n=16)</td>
<td></td>
</tr>
</tbody>
</table>

**Certainty**

- **CI:** Confidence interval; **RR:** Risk ratio; **OR:** Odds ratio

### Explanations

- a. Mortality was evaluated as part of a post-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 generalisable to all other countries relevant for this recommendation.
- c. Downgraded by one level for serious indirectness. Ayles 2013 was a community-randomised 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.
- g. Downgraded by two levels for very serious risk of bias. 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.
- h. Downgraded by two levels for very serious inconsistency. 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).
- i. No significant concerns regarding indirectness were identified.
- j. Imprecision was not a major concern, given the large number of participants in most included studies.
- k. Downgraded by one level for serious imprecision. This was based on the small number of overall participants evaluated.

### References


43. Azit NA, Ismail A, Ahmad N, et al. Factors associated with tuberculosis disease among children who are household contacts of tuberculosis cases in an urban setting in Malaysia. BMC Public Health 2019;19(1)


Table 4. Should systematic screening for TB disease, compared to passive case detection, be conducted in prison settings?

<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Impact</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Certainty assessment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>№ of studies</strong></td>
<td>Study design</td>
<td>Risk of bias</td>
</tr>
<tr>
<td>1</td>
<td>observational studies</td>
<td>serious&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Piaiao 2016:</td>
<td>n/N (%; 95%CI) vs PCF n/N (%; 95%CI)</td>
</tr>
<tr>
<td></td>
<td>4/40 (10%; 3–24%)</td>
<td>27/53 (51%; 37–65%)</td>
</tr>
<tr>
<td></td>
<td>Earlier case detection: severity at diagnosis – smear positivity (smear positive among culture positive cases)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>observational studies</td>
<td>very serious&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>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).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TB disease prevalence (non-randomized studies) (Sanchez et al 2013 in Brazil and Tsegaye Sahle et al 2019 in Ethiopia)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>randomized trials</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td>Mean case detection rate, defined as &quot;the number of new smear positive cases detected divided by the estimated number of incident smear positive cases, expressed as a percentage&quot;, mean difference in case detection rate +52.9 percentage points (95% CI 17.5–88.3). CNR ratio= 1.78 (no uncertainty estimate available).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TB case notification rates (randomized studies) (Adane et al 2019)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>observational studies</td>
<td>very serious&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Four observational studies in Zambia (Maggard et al 2015), India (Mallick et al 2017), Uganda (Karamaggi et al 2018) and USA (Degner et al 2016). All uncontrolled before-after design. Variety of ACF interventions evaluated, one study compared two types of ACF rather than to standard case detection. Three had co-interventions in addition to ACF. Point estimate favoured ACF in all four (ratio of CNR ratios 2.96 (Maggard), 1.30 (Mallick), 1.24 (Maggard), 3.96 (Degner). Measures of uncertainty not available.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TB case notification rates (non-randomized studies)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>randomized trials</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td>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).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Knowledge, attitudes and practices (Adane 2019)</td>
<td></td>
</tr>
</tbody>
</table>

CI: Confidence interval; ACF: Active case-finding, PCF: Passive case-finding; CNR: Case notification ratio; aOR: Adjusted odds ratio

Explanations

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).
WHO consolidated guidelines on tuberculosis: Web Annex B. GRADE Summary of Findings Tables

15

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).
i. As assessed using ROBINS-i (downgraded by 2 levels for very serious risk of bias).
j. Measures of uncertainty not generally available. Small numbers of studies. In some studies, ACF applied to small subset of population but outcome is measured in wider population (not all of whom were exposed to ACF) (downgraded by 2 levels for very serious imprecision).
k. Measured by survey rather than observation (downgraded by 2 levels for very serious indirectness).

References
### Table 5. Should prolonged cough (2 weeks or more) be used to screen for TB disease in the general population?

| Sensitivity | 0.42 (95% CI: 0.36 to 0.48) |
| Specificity | 0.94 (95% CI: 0.92 to 0.96) |
| Prevalences | 0.5% | 1% | 2% |

<table>
<thead>
<tr>
<th>Outcome</th>
<th>№ of studies (№ of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (patients with active TB)</td>
<td>40 studies 6,737 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>very serious a</td>
<td>not serious</td>
<td>serious b</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td>3 (3 to 3)</td>
<td>6 (5 to 6)</td>
<td>12 (10 to 13)</td>
<td>☐ ☐ ☐ VERY LOW</td>
<td></td>
</tr>
<tr>
<td>True negatives (patients without active TB)</td>
<td>40 studies 1284181 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious d</td>
<td>not serious</td>
<td>not serious e</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active TB)</td>
<td>57 (42 to 75)</td>
<td>56 (42 to 75)</td>
<td>56 (42 to 74)</td>
<td>☐ ☐ ☐</td>
<td></td>
</tr>
</tbody>
</table>

**Explanations**

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.
References

1. Morishita 2017–2
2. Morishita 2017–1
3. Pelissari 2018
4. Morishita 2017–3
5. Seri 2017
6. Telsingehe 2014
7. Lewis 2009a
8. Morishita 2017–4
9. Claassens 2017–2a
10. Claassens 2017–1a
11. Corbett 2010a
12. Wei 2014
13. MoPH DPRK 2017
15. Qadeer 2016
16. MoPH Thailand 2017
17. Chadha 2018
18. FRoNigeria 2014
19. Republic of Uganda 2018
20. Ghana NTP 2015
21. MoH Cambodia 2012
22. Kebede 2014
23. Adetifa 2016
24. Kenya MoH 2018
25. Law 2015
26. Republic of Zimbabwe 2015
27. Federal MoH Sudan 2018
28. Cheng 2015
29. Mongolia MoH 2016
30. MoH Indonesia 2015
31. Van’t Hoog 2012
32. NTP Philippines 2018
33. Den Boon 2006
34. Ho 2016
35. Nair 2016
36. Koesoemadinata 2018
37. Fox 2012
38. Ntinginya 2012
39. Muyoyeta 2017
40. Moosazadeh 2015
41. Morishita 2017–5
### Table 6. Should any cough be used to screen for TB disease in the general population?

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>0.5%</th>
<th>1%</th>
<th>2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.51 (95% CI: 0.43 to 0.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>0.88 (95% CI: 0.82 to 0.92)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>№ of studies (№ of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (patients with active TB)</td>
<td>21 studies 2,734 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>very serious</td>
<td>not serious</td>
<td>serious</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives (patients without active TB)</td>
<td>21 studies 768,291 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious</td>
<td>not serious</td>
<td>serious</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active TB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Explanations

- **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 (specificity 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.

#### References

1. Pelissari 2018
2. Lewis 2009a
3. Kimerling 1999
4. Corbett 2010a
5. MoPH Thailand 2017
6. FRoNigeria 2014
7. Rwanda MoH 2014
8. Republic of Uganda 2018
9. MoH Cambodia 2012
10. Kenya MoH 2018
11. Republic of Zimbabwe 2015
12. Cheng 2015
14. MoH Myanmar 2012
15. Van’t Hoog 2012
16. Ho 2016
17. Wood 2007
18. Ayles 2009a
20. Singh 2013
21. Little 2018
### Table 7. Should any TB symptom be used to screen for TB disease in the general population?

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.71 (95% CI: 0.62 to 0.79)</td>
<td>0.64 (95% CI: 0.52 to 0.74)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies ( Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (patients with active TB)</td>
<td>28 studies 3915 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>very serious a</td>
<td>not serious</td>
<td>serious b</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td>1 (1 to 2)</td>
<td>3 (2 to 4)</td>
<td>6 (4 to 8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives (patients without active TB)</td>
<td>28 studies 460,878 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious d</td>
<td>not serious</td>
<td>serious e</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active TB)</td>
<td>361 (256 to 480)</td>
<td>359 (235 to 478)</td>
<td>355 (252 to 473)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Explanations**

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.

**References**

1. Morishita 2017–5
2. Ntinginya 2012
3. Little 2018
4. Muyoyeta 2017
5. Nair 2016
6. FRoNigeria 2014
7. Malawi MoH 2016
8. Kenya MoH 2018
9. MoH Cambodia 2005
10. Republic of Zimbabwe 2015
11. Mongolia MoH 2016
12. MoH Myanmar 2012
13. Van’t Hoog 2012
15. Ho 2016
16. Ayles 2009a
17. Claassens 2017–2a
18. Claassens 2017–1a
19. Corbett 2010a
20. Morishita 2017–4
21. Lewis 2009a
22. Mabuto 2015
23. Morishita 2017–3
24. Seri 2017
25. Telisinghe 2014
26. Morishita 2017–1
27. Morishita 2017–2
28. Cheng 2008a
### Table 8. Should chest X-ray (any abnormality) be used to screen for TB disease in the general population?

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>0.94 (95% CI: 0.92 to 0.96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity</td>
<td>0.89 (95% CI: 0.85 to 0.92)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalences</th>
<th>0.5%</th>
<th>1%</th>
<th>2%</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies (Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives</td>
<td>22 studies 4243 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>very serious a</td>
<td>not serious</td>
<td>serious b</td>
</tr>
<tr>
<td>False negatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives</td>
<td>22 studies 1012752 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious d</td>
<td></td>
<td>serious e</td>
</tr>
<tr>
<td>False positives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Explanations**

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 TB 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%).

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

**References**

1. Telisinghe 2014
2. Morasert 2018
3. Den Boon 2006
4. NTP Philippines 2018
5. Van’t Hoog 2012
6. MoH Myanmar 2012
7. MoH Indonesia 2015
8. NTP Bangladesh 2017
10. Federal MoH Sudan 2018
11. Republic of Zimbabwe 2015
12. Law 2015
13. Melendez 2017–1
14. Kenya MoH 2018
15. Kebede 2014
16. MoH Cambodia 2012
17. Ghana NTP 2015
18. Republic of Uganda 2018
19. Rwanda MoH 2014
20. MoPH Thailand 2017
22. MoPH DPRK 2017
23. Fox 2012
Table 9. Should chest X-ray (suggestive for TB) be used to screen for TB disease in the general population?

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>0.85 (95% CI: 0.77 to 0.90)</th>
<th>Specificity</th>
<th>0.96 (95% CI: 0.93 to 0.97)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalences</td>
<td>0.5%</td>
<td>1%</td>
<td>2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies ( Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (patients with active TB)</td>
<td>19 studies 2,152 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>serious a</td>
<td>not serious</td>
<td>serious b</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives (patients without active TB)</td>
<td>19 studies 464,818 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious d</td>
<td>not serious</td>
<td>not serious e</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active TB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Explanations

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.

References

1. Morasert 2018
2. Pelissari 2018
3. Seri 2017
4. Telisinghe 2014
5. Mor 2012
6. Wei 2014
7. Hoa 2012
8. Malawi MoH 2016
9. FRoNigeria 2014
10. MoH Cambodia 2012
11. Adetifa 2016
12. Kenya MoH 2018
13. Melendez 2017–2
14. MOH Myanmar 2012
15. Van’t Hoog 2011b
17. Nair 2016
18. Koesoemadinata 2018
19. Lu 2016
**Table 10. Should molecular WHO-recommended rapid diagnostic tests be used to screen for TB disease in the general population?**

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.69 (95% CI: 0.48 to 0.86)</td>
<td>0.99 (95% CI: 0.97 to 0.99)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies ( Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (patients with pulmonary tuberculosis)</td>
<td>5 studies 337 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious</td>
<td>Risk of bias</td>
<td>Indirectness</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having pulmonary tuberculosis)</td>
<td>5 studies 8619 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>very serious</td>
<td>not serious</td>
<td>serious</td>
</tr>
<tr>
<td>True negatives (patients without pulmonary tuberculosis)</td>
<td>5 studies 8619 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious</td>
<td>very serious</td>
<td>not serious</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having pulmonary tuberculosis)</td>
<td>5 studies 8619 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
</tbody>
</table>

### Explanations

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.

### References

Table 11. 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?

<table>
<thead>
<tr>
<th></th>
<th>Chest X-ray with CAD software</th>
<th>Chest X-ray with human reader interpretation (any TB abnormality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.90 to 0.92</td>
<td>0.82 to 0.98</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.23 to 0.66</td>
<td>0.14 to 0.63</td>
</tr>
<tr>
<td>Prevalences</td>
<td>0.5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies ( Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Risk of bias</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>3 studies 1325 patients ( cohort &amp; case- control type studies)</td>
<td>not serious</td>
<td>serious * not serious not serious none</td>
<td>5 to 5</td>
<td>4 to 5</td>
<td>45 to 46</td>
<td>41 to 49</td>
<td>none</td>
</tr>
<tr>
<td>False negatives</td>
<td>0 to 0 ( patients incorrectly classified as not having active TB )</td>
<td>4 to 5</td>
<td>1 to 9</td>
<td>8 to 10</td>
<td>2 to 18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives</td>
<td>3 studies 8391 patients ( cohort &amp; case-control type studies)</td>
<td>not serious</td>
<td>serious * not serious serious b none</td>
<td>229 to 658</td>
<td>136 to 622</td>
<td>219 to 628</td>
<td>130 to 594</td>
<td>none</td>
</tr>
<tr>
<td>False positives</td>
<td>337 to 766 ( patients incorrectly classified as having active TB )</td>
<td>322 to 731</td>
<td>356 to 820</td>
<td>305 to 693</td>
<td>337 to 777</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect per 1,000 patients tested</th>
<th>pre-test probability of 0.5%</th>
<th>pre-test probability of 5%</th>
<th>pre-test probability of 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXR with CAD software</td>
<td>human reader (any TB abnormality)</td>
<td>CXR with CAD software</td>
<td>human reader (any TB abnormality)</td>
</tr>
<tr>
<td>1 more to 0 fewer TP in CXR with CAD software</td>
<td>4 more to 3 fewer TP in CXR with CAD software</td>
<td>8 more to 6 fewer TP in CXR with CAD software</td>
<td></td>
</tr>
<tr>
<td>0 to 0</td>
<td>4 to 5</td>
<td>1 to 9</td>
<td>8 to 10</td>
</tr>
<tr>
<td>1 more to 0 fewer FN in CXR with CAD software</td>
<td>4 fewer to 3 more FN in CXR with CAD software</td>
<td>8 fewer to 6 more FN in CXR with CAD software</td>
<td></td>
</tr>
<tr>
<td>229 to 658</td>
<td>136 to 622</td>
<td>219 to 628</td>
<td>130 to 594</td>
</tr>
<tr>
<td>93 more to 36 more TN in CXR with CAD software</td>
<td>89 more to 34 more TN in CXR with CAD software</td>
<td>84 more to 32 more TN in CXR with CAD software</td>
<td></td>
</tr>
<tr>
<td>337 to 766</td>
<td>373 to 859</td>
<td>322 to 731</td>
<td>356 to 820</td>
</tr>
<tr>
<td>93 fewer to 36 fewer FP in CXR with CAD software</td>
<td>89 fewer to 34 fewer FP in CXR with CAD software</td>
<td>84 fewer to 32 fewer FP in CXR with CAD software</td>
<td></td>
</tr>
<tr>
<td>207 to 595</td>
<td>123 to 563</td>
<td>305 to 693</td>
<td>337 to 777</td>
</tr>
</tbody>
</table>

Explanations

a. The population here was pre-screened with this analysis focusing on bacteriological testing. Only people who got tested by a microbiological test could be included in this. We downgrade one level for indirectness as this is not representative of the entire screening population.

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.
References


Table 12. Should chest X-ray with CAD software, compared to human reader interpretation, be used to triage for TB disease in people eligible for TB triage, using a bacteriologic reference standard?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies (Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Risk of bias</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Effect per 1,000 patients tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pre-test probability of 10%</td>
<td>Pre-test probability of 20%</td>
<td>Pre-test probability of 30%</td>
<td>Test accuracy CoE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True positives (patients with active TB)</td>
<td>3 studies 4911 patients</td>
<td>cohort &amp; case-control type studies</td>
<td>not serious</td>
<td>serious *</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>CXR with CAD software</td>
<td>human reader (any TB abnormality)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 more to 5 fewer TP in CXR with CAD software</td>
<td>3 more to 10 fewer TP in CXR with CAD software</td>
<td>4 more to 15 fewer TP in CXR with CAD software</td>
<td>106 to 148 more TN in CXR with CAD software</td>
<td>94 to 132 more TN in CXR with CAD software</td>
<td>82 to 115 more TN in CXR with CAD software</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td>3 studies 23801 patients</td>
<td>cohort &amp; case-control type studies</td>
<td>not serious</td>
<td>serious *</td>
<td>not serious</td>
<td>serious †</td>
<td>none</td>
<td>CXR with CAD software</td>
<td>human reader (any TB abnormality)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>106 more to 148 fewer FP in CXR with CAD software</td>
<td>94 more to 132 fewer FP in CXR with CAD software</td>
<td>82 more to 115 fewer FP in CXR with CAD software</td>
<td>106 more to 148 more FP in CXR with CAD software</td>
<td>94 more to 132 fewer FP in CXR with CAD software</td>
<td>82 more to 115 fewer FP in CXR with CAD software</td>
</tr>
</tbody>
</table>

WHO consolidated guidelines on tuberculosis:
Web Annex B. GRADE Summary of Findings Tables
Explanations

a. Downgraded by one level for serious indirectness: The FIND study had data on 59% of the patients presenting with signs and symptoms at the referral hospitals. One site in this study (Japan) had patients who were going to the health care center as part of their regular check-up for active TB. This site also included healthy individuals. Another site in this study included data from Germany that contributed majority of the data and data on signs and symptoms was available only for 54% of the included participants. Across all three included studies, there may be important differences in sub-groups such as HIV status, smear-negative status amongst others. This data was unavailable in two of the three studies to investigate further.

b. Of the three studies, one study by FIND had high risk of concern for flow and timing domain as 46% of the participants did not have MRS performed on the specimens. However, as that dataset contribute only 3% to the entire dataset, we did not downgrade for risk of bias.

c. Across all three included studies, there may be important differences in sub-groups such as HIV status, smear-negative status amongst others. This data was unavailable in two of the three studies to investigate further. As we had downgraded one level for imprecision, we decided to not downgrade for inconsistency.

d. 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.

References


Table 13. 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?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>NP of studies (NP of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (patients with active TB)</td>
<td>6 studies 3971 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious a, not serious, not serious b, serious c, none</td>
<td>a C-Reactive Protein (CRP) cutoff of 5mg per litre</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td>6 studies 3971 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious a, not serious, serious d, serious e, none</td>
<td>a C-Reactive Protein (CRP) cutoff of 5mg per litre</td>
</tr>
<tr>
<td>True negatives (patients without active TB)</td>
<td>6 studies 3971 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious a, not serious, serious d, serious e, none</td>
<td>a C-Reactive Protein (CRP) cutoff of 5mg per litre</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active TB)</td>
<td>6 studies 3971 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious a, not serious, serious d, serious e, none</td>
<td>a C-Reactive Protein (CRP) cutoff of 5mg per litre</td>
</tr>
</tbody>
</table>

### Prevalences

<table>
<thead>
<tr>
<th>Prevalences</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>0.90 (95% CI: 0.78 to 0.96)</td>
<td>0.50 (95% CI: 0.29 to 0.71)</td>
</tr>
<tr>
<td>10%</td>
<td>0.83 (95% CI: 0.74 to 0.89)</td>
<td>0.38 (95% CI: 0.25 to 0.53)</td>
</tr>
<tr>
<td>20%</td>
<td>0.78 (95% CI: 0.70 to 0.86)</td>
<td>0.29 (95% CI: 0.18 to 0.41)</td>
</tr>
</tbody>
</table>
Explanations

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 positives may lead to different decisions depending on which limits are assumed.

References


**Table 14. 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?**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>NP of studies (NP of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Prevalences</th>
<th>Effect per 1,000 patients tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (patients with active TB)</td>
<td>8 studies 6238 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.93 (95% CI: 0.88 to 0.96)</td>
<td>Sensitivity</td>
<td>0.83 (95% CI: 0.74 to 0.89)</td>
<td>Pre-test probability of 5%</td>
<td>47 (44 to 48)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.20 (95% CI: 0.10 to 0.38)</td>
<td>Specificity</td>
<td>0.38 (95% CI: 0.25 to 0.53)</td>
<td>Pre-test probability of 10%</td>
<td>5 more TP in chest X-ray (any abnormality) or WHO-4 symptom screen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pre-test probability of 20%</td>
<td>3 fewer FN in chest X-ray (any abnormality) or WHO-4 symptom screen</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td>8 studies 6238 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True negatives (patients without active TB)</td>
<td>8 studies 6238 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HIGH</td>
<td>171 fewer TN in chest X-ray (any abnormality) or WHO-4 symptom screen</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active TB)</td>
<td>8 studies 6238 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>Prevalences</td>
<td>5%</td>
<td>10%</td>
<td>20%</td>
<td>Test accuracy CoE</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>10%</td>
<td>20%</td>
<td>171 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen</td>
<td>162 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen</td>
<td>144 more FP in chest X-ray (any abnormality) or WHO-4 symptom screen</td>
</tr>
</tbody>
</table>
Explanations

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.

References


Table 15. 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?

<table>
<thead>
<tr>
<th>Molecular WHO-recommended rapid diagnostic test (mWRD)</th>
<th>WHO-recommended 4 symptom screen followed by mWRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>0.77 (95% CI: 0.69 to 0.84)</td>
<td>0.76 (95% CI: 0.68 to 0.83)</td>
</tr>
<tr>
<td>Specificity</td>
<td>Specificity</td>
</tr>
<tr>
<td>0.93 (95% CI: 0.89 to 0.96)</td>
<td>0.93 (95% CI: 0.89 to 0.96)</td>
</tr>
</tbody>
</table>

Prevalences 10% 20% 30%

<table>
<thead>
<tr>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of bias</td>
<td>WHO-recommended molecular rapid diagnostics</td>
</tr>
<tr>
<td>Indirectness</td>
<td>WHO-recommended 4 symptom screen followed by an mWRD diagnostic test</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>WHO-recommended molecular rapid diagnostics</td>
</tr>
<tr>
<td>Imprecision</td>
<td>WHO-recommended 4 symptom screen followed by an mWRD diagnostic test</td>
</tr>
<tr>
<td>Publication bias</td>
<td>WHO-recommended molecular rapid diagnostics</td>
</tr>
<tr>
<td>CoE</td>
<td>WHO-recommended 4 symptom screen followed by an mWRD diagnostic test</td>
</tr>
</tbody>
</table>

### True positives (patients with active TB)
- **4 studies** 639 patients cross-sectional (cohort type accuracy study)
- Sensitivity: 0.77 (95% CI: 0.69 to 0.84)
- Specificity: 0.93 (95% CI: 0.89 to 0.96)
- Pre-test probability of 10%:
  - 77 (69 to 84)
  - 1 more TP in molecular WHO-approved rapid diagnostics
- Pre-test probability of 20%:
  - 76 (68 to 83)
  - 2 more TP in molecular WHO-approved rapid diagnostics
- Pre-test probability of 30%:
  - 152 (136 to 166)
  - 3 more TP in molecular WHO-approved rapid diagnostics

### False negatives (patients incorrectly classified as not having active TB)
- **23 (16 to 31)**
- Pre-test probability of 10%:
  - 69 (48 to 93)
  - 1 fewer FN in molecular WHO-approved rapid diagnostics
- Pre-test probability of 20%:
  - 46 (32 to 62)
  - 2 fewer FN in molecular WHO-approved rapid diagnostics
- Pre-test probability of 30%:
  - 72 (51 to 96)
  - 3 fewer FN in molecular WHO-approved rapid diagnostics

### True negatives (patients without active TB)
- **4 studies** 639 patients cross-sectional (cohort type accuracy study)
- Sensitivity: 0.76 (95% CI: 0.68 to 0.83)
- Specificity: 0.93 (95% CI: 0.89 to 0.96)
- Pre-test probability of 10%:
  - 837 (801 to 864)
  - 0 fewer TN in molecular WHO-approved rapid diagnostics
- Pre-test probability of 20%:
  - 837 (801 to 864)
  - 0 fewer TN in molecular WHO-approved rapid diagnostics
- Pre-test probability of 30%:
  - 651 (623 to 672)
  - 0 fewer TN in molecular WHO-approved rapid diagnostics

### False positives (patients incorrectly classified as having active TB)
- **63 (36 to 99)**
- Pre-test probability of 10%:
  - 63 (36 to 99)
  - 0 fewer FP in molecular WHO-approved rapid diagnostics
- Pre-test probability of 20%:
  - 63 (36 to 99)
  - 0 fewer FP in molecular WHO-approved rapid diagnostics
- Pre-test probability of 30%:
  - 49 (28 to 77)
  - 0 fewer FP in molecular WHO-approved rapid diagnostics
Explanations

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. 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.

c. Sensitivity estimates ranged from 25% to 83% with overlapping CIs. We did not downgrade for inconsistency.

d. Specificity estimates ranged from 90% to 96%. We did not downgrade for inconsistency.

References


Table 16. 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?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies ( Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (patients with active TB)</td>
<td>14 studies 9209 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.69 (95% CI: 0.60 to 0.76)</td>
<td>Sensitivity</td>
<td>0.62 (95% CI: 0.56 to 0.69)</td>
<td>Sensitivity</td>
<td>0.98 (95% CI: 0.97 to 0.99)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.69 (95% CI: 0.60 to 0.76)</td>
<td>Sensitivity</td>
<td>0.62 (95% CI: 0.56 to 0.69)</td>
<td>Sensitivity</td>
<td>0.98 (95% CI: 0.97 to 0.99)</td>
</tr>
<tr>
<td>Prevalences</td>
<td>5%</td>
<td>10%</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active TB)</td>
<td>14 studies 9209 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious</td>
<td>not serious a</td>
<td>serious b</td>
</tr>
<tr>
<td>True negatives (patients without active TB)</td>
<td>14 studies 9209 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>not serious</td>
<td>not serious a</td>
<td>not serious b</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active TB)</td>
<td>10 fewer TN in molecular WHO-approved rapid diagnostics</td>
<td>9 fewer TN in molecular WHO-approved rapid diagnostics</td>
<td>8 fewer TN in molecular WHO-approved rapid diagnostics</td>
<td>16 (9 to 28)</td>
<td>9 (9 to 28)</td>
</tr>
<tr>
<td>3 more TP in molecular WHO-approved rapid diagnostics</td>
<td>7 more TP in molecular WHO-approved rapid diagnostics</td>
<td>14 more TP in molecular WHO-approved rapid diagnostics</td>
<td>3 fewer FN in molecular WHO-approved rapid diagnostics</td>
<td>7 fewer FN in molecular WHO-approved rapid diagnostics</td>
<td>14 fewer FN in molecular WHO-approved rapid diagnostics</td>
</tr>
</tbody>
</table>

WHO consolidated guidelines on tuberculosis: Web Annex B. GRADE Summary of Findings Tables
Explanations

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. 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.

d. 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.

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.

References


Table 17. Should symptom screening involving any one of cough, fever, or poor weight gain be used to screen for TB disease in child and adolescent close contacts (under 15 years, composite reference standard)?

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.89 (95% CI: 0.52 to 0.98)</td>
<td>0.69 (95% CI: 0.51 to 0.83)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Specificity</td>
<td>CI: 0.52 to 0.98</td>
<td>CI: 0.51 to 0.83</td>
</tr>
<tr>
<td>True positives</td>
<td>(patients with active pulmonary TB)</td>
<td>4 studies 1113 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
</tr>
<tr>
<td>False negatives</td>
<td>(patients incorrectly classified as not having active pulmonary TB)</td>
<td>4 studies 2582 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
</tr>
<tr>
<td>True negatives</td>
<td>(patients without active pulmonary TB)</td>
<td>4 studies 2582 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
</tr>
<tr>
<td>False positives</td>
<td>(patients incorrectly classified as having active pulmonary TB)</td>
<td>4 studies 1113 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
</tr>
</tbody>
</table>

Explanations

a. The two studies with relatively lower sensitivity estimates only included patients <5 years of age, this may explain in part differences in sensitivity. We downgraded one level for inconsistency.

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

c. The single study with notably lower specificity used a symptom screen that assessed the presence of symptoms over the past month, while the symptom screens of other studies were composed of more recent symptoms. This may explain differences in specificity. We downgraded one level for inconsistency.

d. We thought the 95% CI around false positives and true negatives would likely lead to different decisions depending on which confidence limits are assumed. We downgraded one level for imprecision.

References


Table 18. Should chest X-ray (suggestive of TB) be used to screen for TB disease in child and adolescent close contacts of individuals with TB?

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84 (95% CI: 0.70 to 0.92)</td>
<td>0.91 (95% CI: 0.90 to 0.92)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies ( Nº of patients)</th>
<th>Study design</th>
<th>Factors that may decrease certainty of evidence</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (patients with active pulmonary TB)</td>
<td>4 studies 113 patients</td>
<td>cohort &amp; case-control type studies</td>
<td>serious[^a] not serious[^b] not serious[^c] serious[^d] none</td>
<td>4 (3 to 5) 42 (35 to 46) 84 (70 to 92)</td>
<td>LOW</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having active pulmonary TB)</td>
<td>4 studies 2437 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>serious not serious[^b] not serious[^c] not serious none</td>
<td>1 (0 to 2) 8 (4 to 15) 16 (8 to 30)</td>
<td></td>
</tr>
<tr>
<td>True negatives (patients without active pulmonary TB)</td>
<td>4 studies 2437 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>serious not serious[^b] not serious[^c] not serious none</td>
<td>905 (896 to 915) 864 (855 to 874) 819 (810 to 828)</td>
<td>MODERATE</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having active pulmonary TB)</td>
<td>4 studies 2437 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>serious not serious[^b] not serious[^c] not serious none</td>
<td>90 (80 to 99) 86 (76 to 95) 81 (72 to 90)</td>
<td></td>
</tr>
</tbody>
</table>

Explanations

a. Chest radiography was a component of the composite reference standard in all four studies. We downgraded one level for risk of bias.

b. The one study contributing >70% of these data was conducted in four different countries, one of which is a high TB burden country. One of the other studies was conducted in a high TB burden country. The main contributing study had a TB prevalence of 2.3%, and the range of prevalences was 1.9 to 13.1%. All studies were conducted in outpatient settings.

c. For individual studies, sensitivity estimates ranged from 78% to 100%, with the later only based upon analysis of four cases of active TB. We did not downgrade for inconsistency.

d. There were few patients contributing to the analysis for sensitivity. We downgraded one level for imprecision.

e. For individual studies, specificity estimates ranged from 87% to 100%. All three of the smaller studies had estimated specificity of 100%. We did not downgrade for inconsistency.

References


Table 19. Should symptom screening (current cough, fever, poor weight gain, or TB contact) be used to screen for TB disease in children living with HIV in outpatient settings (composite reference standard)?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Nº of studies ( Nº of patients)</th>
<th>Study design</th>
<th>Risk of bias</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Effect per 1,000 patients tested</th>
<th>Test accuracy CoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>2 studies 1219 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>serious *</td>
<td>not serious</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>3 (3 to 3)</td>
<td>MODERATE</td>
</tr>
<tr>
<td>False negatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 (2 to 2)</td>
<td></td>
</tr>
<tr>
<td>False positives</td>
<td>2 studies 201916 patients</td>
<td>cross-sectional (cohort type accuracy study)</td>
<td>serious *</td>
<td>not serious</td>
<td>serious *</td>
<td>not serious</td>
<td>none</td>
<td>935 (856 to 975)</td>
<td>LOW</td>
</tr>
<tr>
<td>True negatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>893 (817 to 931)</td>
<td></td>
</tr>
<tr>
<td>False positives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>846 (774 to 882)</td>
<td></td>
</tr>
</tbody>
</table>

Explanations:

a. As assessed by QUADAS-2, both studies had high risk of bias in the Flow and Timing domain. We downgraded one level for risk of bias.

b. For individual studies, specificity estimates ranged from 89% to 97%. We thought that differences in threshold for clinical diagnosis could explain in part the heterogeneity. We downgraded one level for inconsistency.

References:


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