THE USE OF THE XPERT MTB/RIF ASSAY FOR THE DETECTION OF PULMONARY AND EXTRAPULMONARY TUBERCULOSIS AND RIFAMPICIN RESISTANCE IN ADULTS AND CHILDREN

EXPERT GROUP MEETING REPORT
October 2013

This report contains the collective views of an international group of experts, and does not necessarily represent the decisions or the stated policy of the World Health Organization. Mention of a technology does not imply endorsement of any specific commercial product.
1. Executive summary

Background

Earlier and improved tuberculosis (TB) case detection, including smear-negative disease often associated with HIV-TB co-infection and young age, together with enhanced capacity to diagnose multidrug-resistant tuberculosis (MDR-TB) are global priorities for TB care and control. In September 2010, the World Health Organization (WHO) convened an Expert Group to review the available evidence on the Xpert MTB/RIF® assay (Cepheid, Sunnyvale, USA) for the purpose of formulating recommendations to guide the use of the test. Policy recommendations on the Xpert MTB/RIF® assay (Xpert MTB/RIF) were issued by WHO early in 2011, supported by an operational ‘How-to’ document and a Checklist for country implementation.

Current WHO policy guidance recommends that Xpert MTB/RIF be used as an initial diagnostic test in individuals suspected of MDR or HIV-associated TB (strong recommendation, moderate quality of evidence). The guidance also provides a conditional recommendation that Xpert MTB/RIF be used as a follow-on test to smear microscopy in settings where MDR or HIV are of lesser concern, especially in further testing of smear-negative specimens. Generalizing from adult data, the recommendation includes the use of Xpert MTB/RIF in children, acknowledging the difficulties in the microbiological diagnosis of childhood TB.

Since 2010, more than 85 peer-reviewed research papers on Xpert MTB/RIF for pulmonary, extrapulmonary and pediatric TB have been published and ongoing studies are being performed. Extrapulmonary TB accounts for about 25% of all TB and an even higher percentage in children and immunocompromised patients. Diagnosis of extrapulmonary TB is often challenging, requiring the clinician to obtain specimens for microscopy, culture and histopathological examination from the suspected sites of involvement. However, the availability of these techniques is limited and the need for alternative diagnostics for TB in non-respiratory samples is great. In 2011, the global burden of TB in children was estimated at 500,000 cases representing approximately 6% of all TB cases. However, this burden is in all likelihood an underestimate due to the difficulties in obtaining microbiological confirmation of the diagnosis of childhood TB.

The Xpert MTB/RIF assay remains the only fully automated cartridge based real-time DNA based test which can detect both TB and resistance to rifampicin in less than two hours, and the only mature technology representing a new generation of automated molecular diagnostic platforms.

Given the amount of additional data on Xpert MTB/RIF having emerged since 2010, an update of the current WHO policy guidance was warranted. The WHO Global TB Programme therefore commissioned three systematic reviews to update and revise current policy guidance, including the utility of Xpert MTB/RIF for the diagnosis of tuberculosis and rifampicin resistance in pulmonary, extra-pulmonary and

---

1 WHO Policy statement “Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF MTB/RIF system”

2 WHO Rapid Implementation document “Technical and operational ‘How-to’, Practical considerations”

3 WHO Checklist of prerequisites to country implementation of Xpert MTB/RIF MTB/RIF and key action points at country level
paediatric TB. An updated review of published studies on the affordability and cost-effectiveness of Xpert MTB/RIF was also done. WHO convened an Expert Group to review the evidence at Les Pensieriès, Veyrier-du-Lac, France on 20-21 May 2013. The major findings and Expert Group recommendations are summarized below and a detailed meeting report will be available at http://www.who.int/tb/dots/laboratory/policy/en.

Summary of results

Xpert MTB/RIF for the diagnosis of pulmonary TB in adults

27 unique studies involving 9558 participants were included in the review. Two of the 27 studies were multicentre international studies (one with five and the other with six distinct study centres). Two of the 27 studies evaluated Xpert MTB/RIF in primary care clinics where Xpert MTB/RIF results could be used to begin treatment on the same day. Sixteen studies (59%) were performed in low-income or middle-income countries. For pulmonary TB detection, the reference standards were solid or liquid culture. For rifampicin resistance, the reference standard was phenotypic culture-based drug susceptibility testing.

As an initial diagnostic test replacing smear microscopy, Xpert MTB/RIF achieved an overall pooled sensitivity of 88% (95% CrI 84% - 92%) and pooled specificity of 99% (95% CrI 98% - 99%), (22 studies, 9008 participants). As an add-on test following a negative smear microscopy result, Xpert MTB/RIF yielded a pooled sensitivity of 68% (95% CrI 61% - 74%) and pooled specificity of 99% (95% CrI 98% - 99%), (23 studies, 7151 participants). For smear-positive, culture-positive TB, Xpert MTB/RIF pooled sensitivity was 98% (95% CrI 97% - 99%), (23 studies, 1952 participants), while for smear-negative, culture-positive TB, it was 68% (95% CrI 61%-74%), (23 studies, 7151 participants).

For people living with HIV, Xpert MTB/RIF pooled sensitivity was 79% (95% CrI 70% - 86%), (7 studies, 1789 participants), while for people without HIV infection, pooled sensitivity was 86% (95% CrI 76% - 92%), (7 studies, 1470 participants).

For rifampicin resistance detection, Xpert MTB/RIF achieved a pooled sensitivity of 95% (95% CrI 90% - 97%), (17 studies, 555/2624 total specimens) and pooled specificity of 98% (95% CrI 97% - 99%), (24 studies, 2414 specimens, true negatives and false positives).

Expert Group consensus

The evidence synthesis process confirmed a high-quality evidence base to support widespread use of Xpert MTB/RIF for the detection of adult pulmonary TB and rifampicin resistance. Therefore:

- Xpert MTB/RIF should be used rather than conventional microscopy, culture and DST as the initial diagnostic test in adults presumed to have MDR-TB or HIV-associated TB (strong recommendation, high-quality evidence).

- Xpert MTB/RIF may be used as a follow-on test to microscopy in adults where MDR-TB and HIV is of lesser concern, especially in further testing of smear-negative specimens. (Conditional recommendation acknowledging resource implications, high-quality evidence).

- Xpert MTB/RIF may be used rather than conventional microscopy and culture as the initial diagnostic test in all adults presumed to have TB. (Conditional recommendation acknowledging resource implications, high-quality evidence).
Xpert MTB/RIF for the diagnosis of extrapulmonary TB in adults and children

15 published and 7 unpublished studies, involving 5922 samples, were included in the review. The majority of studies (59%) were performed in high-burden settings. Due to the heterogeneity of sample types included in the studies, pre-specified subgroups of sample types (pleural fluid, lymph node samples (biopsy and aspirate combined), other tissues and CSF) with a comparison against culture and against a composite reference standard (CRS) were included for meta-analysis. The CRS might have included a nucleic acid amplification test (NAAT other than Xpert MTB/RIF), histology, smear, culture, biochemical testing results, presenting signs or a response to treatment with anti-TB therapy.

Using culture as the reference standard, Xpert MTB/RIF pooled sensitivity was 84.9% (95% CI 72.1% - 92.4%) and pooled specificity was 92.5% (95% CI 80.3%-97.4%) in lymph node tissues or aspirates (14 studies, 849 samples). Five studies (one unpublished) assessed Xpert MTB/RIF on lymph node samples against an author-defined composite reference standard (CRS). The pooled sensitivity was estimated to be 83.7% (95% CI 73.8%-90.3) and the pooled specificity to be 99.2% (95% CI 88.4%-100%).

In cerebrospinal fluid (CSF), Xpert MTB/RIF pooled sensitivity and pooled specificity was 79.5% (95% CI 62.0% - 90.2%) and 98.6% (95% CI 95.8%-99.6%) respectively against culture reference standard (16 studies, 709 samples). Xpert MTB/RIF in CSF against a CRS yielded a pooled sensitivity of 55.5% (95% CI 44.2% - 66.3%) and pooled specificity 98.8% (95% CI 94.5%-99.8%) (6 studies, 512 samples).

Using culture as the reference standard, Xpert MTB/RIF pooled sensitivity was 83.8% (95% CI 65.9% - 93.2%) and pooled specificity was 98.1% (95%CI 92.3%-99.5%) in gastric fluid (12 studies, 1258 samples); and in other tissue samples (12 studies, 699 samples) Xpert MTB/RIF pooled sensitivity was 81.2% (95% CI 67.7% - 89.9%) and pooled specificity was 98.1% (95% CI 87.0%-99.8%).

In pleural fluid, Xpert MTB/RIF pooled sensitivity was 43.7% (95% CI 24.8% - 64.7%) against culture (17 studies, 1385 samples) and 17.0% (95% CI 7.5% - 34.2%) against a composite reference standard (7 studies, 698 samples). Pooled specificity was high against both the culture reference standard and the CRS. The data for additional sample types (i.e. ascitic fluid, pericardial fluid, urine, blood and stool) were limited and therefore not considered for analysis.

Expert Group consensus

CSF: The Expert Group recommended that Xpert MTB/RIF should be used in preference to conventional microscopy and culture as the initial diagnostic test in testing CSF from patients presumed to have TB meningitis (strong recommendation given the urgency for rapid diagnosis, very low quality of evidence). The Expert Group noted that a negative Xpert MTB/RIF result on CSF should be followed up by other tests. The Expert Group also noted that concentration methods should be used to enhance yield when sufficient volumes of CSF are available. These recommendations apply to both children and adults.

Lymph node and tissue: The Expert Group recommended that Xpert MTB/RIF may be used as a replacement test for usual practice (including conventional microscopy, culture, and histology) in testing lymph nodes and tissues from patients presumed to have extrapulmonary TB (conditional recommendation, very low quality of evidence). The Expert Group noted that a negative Xpert MTB/RIF result should be followed by other tests. The Expert Group also noted that sample processing methods for lymph nodes and tissues need to be standardised to optimise yield. These recommendations apply to both adults and children.

Pleural fluid: The Expert Group noted that pleural fluid is a suboptimal sample for the diagnosis of pleural TB and that a pleural biopsy is the preferred sample for bacteriological confirmation (including by Xpert RIF
MTB/RIF. The Expert Group noted that if the Xpert MTB/RIF was used in the diagnostic workup in patients presumed to have pleural TB a positive Xpert MTB/RIF result is considered confirmatory, (conditional recommendation, very low quality of evidence) and that a negative Xpert MTB/RIF result should be followed by other tests. These recommendations apply to both adults and children.

**Xpert MTB/RIF for the diagnosis of pulmonary TB and rifampicin resistance in children**

16 studies, 12 published and 4 unpublished, were included in the review. All studies were performed at higher levels of care, and the children included were mainly inpatients. Thirteen studies were performed in low or middle-income countries.

Pulmonary TB was evaluated in 13 studies including 2603 participants. The overall pooled sensitivity of Xpert MTB/RIF against culture (10 studies) in children presumed to have TB was 66% in 10 studies where expectorated sputum (ES) or induced sputum (IS) was used (pooled 95% CrI 52% - 77%), and 66% in seven studies where gastric lavage aspirates (GLA) were used (pooled 95% CrI 51% - 81%). Pooled specificity of Xpert MTB/RIF against culture as the reference standard was ±98% with narrow confidence intervals.

The pooled sensitivity of Xpert MTB/RIF in culture-negative paediatric specimens against clinical TB as the reference standard was very low at 4% for ES/IS (8 studies) and 15% for GLA (3 studies) with wide confidence intervals. It is likely that the apparent poor performance of Xpert MTB/RIF was the result of a clinical TB reference standard that lacked specificity. The sensitivity of Xpert MTB/RIF to detect rifampicin resistance in paediatric specimens was 86% (95%Crl 53% - 98%)

**Expert Group consensus**

The Expert Group recommended that Xpert MTB/RIF should be used rather than conventional microscopy, culture and DST as the initial diagnostic test in children presumed to have MDR-TB or HIV-associated TB (strong recommendation given the difficulties in diagnosing paediatric TB, very low quality of evidence).

The Expert Group also recommended that Xpert MTB/RIF may be used rather than conventional microscopy and culture in all other children presumed to have pulmonary TB. (conditional recommendation acknowledging resource implications, very low quality of evidence).

The Expert Group noted that Xpert MTB/RIF should not be used as the only test in the diagnostic pathway of children with presumed TB, and that a child with high clinical suspicion for TB should be treated even if the Xpert MTB/RIF result is negative or if the test is not available.

**Affordability and cost-effectiveness of Xpert MTB/RIF for the diagnosis of tuberculosis**

Twelve (12) published papers were identified comparing the costs of using Xpert MTB/RIF as the initial diagnostic test or as a follow-on test to microscopy for the diagnosis of TB and MDR-TB versus current diagnostic algorithms for diagnosing TB and MDR-TB. The setting for the majority of analyses was South Africa; two studies included other countries of Sub-Saharan Africa (Botswana, Lesotho, Namibia, Swaziland and Uganda); one study covered the countries of the Former Soviet Union; and one global analysis included all countries. Seven (7/12) of these studies were analysis of cost and 5 were cost-effectiveness analyses. Wide variation in methodology, underlying assumptions, and intended use of Xpert MTB/RIF made systematic review impossible.

**Expert Group Consensus**

Although the use of Xpert MTB/RIF was overall found to be cost-effective, more directly measured costing evidence in more countries is needed for improving cost-effectiveness analyses.
Contents

1. EXECUTIVE SUMMARY ........................................................................................................................................... 2
2. BACKGROUND ................................................................................................................................................................ 11
3. EVIDENCE BASE ............................................................................................................................................................ 12
   3.1. EVIDENCE SYNTHESIS ........................................................................................................................................... 12
   3.2. MEETING OBJECTIVES ........................................................................................................................................ 12
   3.3. GRADE EVALUATION ............................................................................................................................................ 13
      3.3.1. PICO Questions for each review .................................................................................................................... 14
      3.3.2. Determination of relative importance of patient outcomes ................................................................... 15
      3.3.3. Assessment of study quality ...................................................................................................................... 16
   3.4. MEETING PROCEDURAL ISSUES .......................................................................................................................... 16
4. RESULTS ............................................................................................................................................................................ 17
   4.1. XPERT MTB/RIF FOR THE DIAGNOSIS OF PULMONARY TB ............................................................................ 17
      4.1.1. Study characteristics .................................................................................................................................. 17
      4.1.2. Study Quality .............................................................................................................................................. 17
      4.1.3. Xpert MTB/RIF used as an initial test replacing smear microscopy .............................................................. 19
      4.1.4. Investigations of heterogeneity -TB detection in smear-positive and smear-negative individuals suspected of having TB ............................................................................................................................................................ 21
      4.1.5. TB detection in HIV-negative and HIV-positive individuals suspected of having TB ......................................................... 23
      4.1.6. TB detection among HIV-positive individuals by smear status ........................................................................ 23
      4.1.7. Effect of condition of the specimen .............................................................................................................. 24
      4.1.8. Effect of specimen preparation ................................................................................................................... 25
      4.1.9. Effect of the proportion of culture-confirmed TB cases in the study .............................................................. 25
      4.1.10. Effect of country income status .................................................................................................................. 25
      4.1.11. Rifampicin resistance detection ................................................................................................................ 25
      4.1.12. Investigations of heterogeneity, rifampicin resistance detection .......................................................................... 26
      4.1.13. Summary of findings and GRADE evidence profiles .................................................................................. 28
      4.1.15. GRADE Evaluation and Recommendations ............................................................................................... 30
   4.2. XPERT MTB/RIF FOR THE DIAGNOSIS OF EXTRAPULMONARY TB AND RIFAMPICIN RESISTANCE IN ADULTS AND CHILDREN .......................................................................................................................... 49
      4.2.1. Study characteristics .................................................................................................................................. 49
      4.2.2. Study Quality .............................................................................................................................................. 52
      4.2.3. Xpert MTB/RIF for TB detection .................................................................................................................. 53
      4.2.4. Rifampicin resistance detection .................................................................................................................. 60
      4.2.5. Summary of findings .................................................................................................................................. 60
      4.2.6. Strengths and limitations of the evidence base .......................................................................................... 62
      4.2.7. GRADE Evaluations and Recommendations ............................................................................................... 63
      4.2.8. Further research needs ................................................................................................................................ 79
   4.3. XPERT MTB/RIF FOR THE DIAGNOSIS OF PULMONARY TB, PERIPHERAL LYMPH NODE TB AND TB MENINGITIS IN CHILDREN ............................................................................................................. 80
      4.3.1. Study characteristics .................................................................................................................................. 80
      4.3.2. Study Quality .............................................................................................................................................. 80
      4.3.3. TB Detection- Xpert MTB/RIF as the initial test ............................................................................................ 81
      4.3.4. Xpert MTB/RIF compared with smear microscopy ...................................................................................... 83
      4.3.5. Investigations of heterogeneity ................................................................................................................... 85
      4.3.6. Xpert MTB/RIF for the detection of peripheral lymph node TB in children ..................................................... 89
      4.3.7. Xpert MTB/RIF for the detection of TB meningitis in children ........................................................................ 89
      4.3.8. Xpert MTB/RIF for the detection of rifampicin resistance in children ................................................................. 90

Page | 6
4.3.9. Summary of findings ............................................................................................................................. 90
4.3.10. Strengths and limitations of the evidence base ................................................................................... 91
4.3.11. GRADE Evaluations and Recommendations ......................................................................................... 91

4.4. Affordability and cost-effectiveness of Xpert MTB/RIF for the diagnosis of tuberculosis ...................... 114

4.4.1. Summary of Cost Analysis studies: ......................................................................................................... 114
4.4.2. Summary of Cost-effectiveness analysis studies: .................................................................................. 114
4.4.3. Summary of findings: ............................................................................................................................. 115
4.4.4. Recommendations ............................................................................................................................... 115

Annex 1: Meeting participants .......................................................................................................................... 116
Annex 2: Meeting agenda ................................................................................................................................... 119
Annex 3: Declarations of interest ...................................................................................................................... 125
Annex 4: Selection of studies evaluating Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults ...................................................................................................................... 126
Annex 5: Selection of studies evaluating Xpert MTB/RIF for extrapulmonary tuberculosis in adults and children ........................................................................................................................................ 138
Annex 6: Selection of studies evaluating Xpert MTB/RIF for pulmonary and extrapulmonary tuberculosis and rifampicin resistance in children ...................................................................................................................... 144
Annex 7: Literature review of affordability and cost-effectiveness of Xpert MTB/RIF for the diagnosis of tuberculosis ........................................................................................................................................ 147

List of Figures

Figure 1: Risk of bias and applicability judgements regarding each domain presented as percentages across the 36 included study centres (27 studies) ................................................................................................. 18
Figure 2: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection ........................................ 19
Figure 3: Forest plot of Xpert MTB/RIF sensitivity for TB detection in smear-negative individuals suspected of having TB. ........................................................................................................................................... 21
Figure 4: Forest plots of Xpert MTB/RIF sensitivity and specificity for TB detection in HIV-negative and HIV-positive individuals suspected of having TB .................................................................................. 24
Figure 5: Forest plots of Xpert MTB/RIF sensitivity and specificity for detection of rifampicin resistance, Xpert MTB/RIF used as an initial test replacing phenotypic culture-based drug susceptibility testing as the initial test ...................................................................................... 25
Figure 6: Risk of bias and applicability concerns summary for studies using a culture reference standard for TB detection ........................................................................................................................................... 52
Figure 7: Risk of bias and applicability concerns summary for studies using an author-defined composite reference standard for TB detection ........................................................................................................... 52
Figure 8: Forest plot of Xpert MTB/RIF sensitivity for TB detection in smear-positive subgroup ........................................ 54
Figure 9: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in lymph node samples (tissue or aspirate) with culture reference standard ........................................................................................... 55
Figure 10: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in lymph node samples (tissue or aspirate) with composite reference standard ......................................................................................... 56
Figure 11: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in pleural fluid with culture reference standard ...................................................................................................................... 57
Figure 12: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in cerebrospinal fluid with culture reference standard ...................................................................................................................... 58
Figure 13: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in cerebrospinal fluid with composite reference standard ........................................................................................................... 58
Figure 14: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in gastric fluid with culture reference standard ...................................................................................................................... 59
Figure 15: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in tissue (other than lymph node) with culture reference standard ........................................................................................................... 60
Figure 16: Summary estimates across sample types for (A) sensitivity and (B) specificity ........................................................................................................................................................................... 61

Page | 7
Figure 17: Risk of bias and applicability concerns graph: Review authors’ judgments about each domain presented as percentages across included studies ................................................................. 81
Figure 18: Forest plot: Sensitivity and specificity of Xpert MTB/RIF against a reference standard ‘Culture’ by study and specimen type .............................................................................................................. 82
Figure 19: Forest plot. Sensitivity and specificity of smear microscopy against a reference standard ‘Culture’ by study and specimen type .............................................................................................................. 84
Figure 20: Forest plot. Xpert MTB/RIF sensitivities in smear positive and smear negative children ......................................................... 85
Figure 21: Forest plot. Xpert MTB/RIF sensitivities in children aged 0-4 and 5-15 years ........................................................................ 87
Figure 22: Forest plot. Xpert MTB/RIF sensitivities and specificities in HIV infected and uninfected children ......................... 88
Figure 23: Forest plot. Xpert MTB/RIF for the diagnosis of RIF resistance, peripheral lymph node TB and TB meningitis. .... 90
List of Tables

TABLE 1: Xpert MTB/RIF assay for detection of TB and rifampicin resistance ................................................................. 20
TABLE 2: Impact of covariates on heterogeneity of Xpert MTB/RIF sensitivity and specificity, TB detection ......................... 22
TABLE 3: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB........................................ 31
TABLE 4: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in sputum smear-positive individuals ........................................................................................................ 33
TABLE 5: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in sputum smear-negative individuals ........................................................................................................ 35
TABLE 6: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in persons living with HIV ........................................................................................................................................ 37
TABLE 7: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in persons without HIV infection ........................................................................................................ 39
TABLE 8: GRADE evidence profile: The incremental yield of Xpert MTB/RIF compared with microscopy in patients with culture-confirmed TB .............................................................. 41
TABLE 9: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB, as an add-on test following a negative sputum smear microscopy.................................................. 43
TABLE 10: Xpert MTB/RIF sensitivity in smear-negative (culture-confirmed) individuals by HIV status ................................. 45
TABLE 11: GRADE evidence profile: Additional yield of Xpert MTB/RIF over microscopy in smear-negative TB .................. 46
TABLE 12: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing ........................................ 47
TABLE 13: Sample types and processing methods for included studies .................................................................................. 50
TABLE 14: Sensitivity analysis by sample type, TB detection ............................................................................................... 55
TABLE 15: Xpert MTB/RIF assay for tuberculosis detection in lymph node fluid and tissue: A. Evidence profile B. Summary of findings ........................................................................................................ 64
TABLE 16: Xpert MTB/RIF assay for tuberculosis detection in pleural fluid: A. Evidence profile B. Summary of findings ...... 67
TABLE 17: Xpert MTB/RIF assay for tuberculosis detection in cerebrospinal fluid: A. Evidence profile B. Summary of findings ........................................................................................................... 70
TABLE 18: Xpert MTB/RIF assay for tuberculosis detection gastric fluid. ............................................................................. 73
TABLE 19: Xpert MTB/RIF assay for tuberculosis detection in tissue samples ........................................................................ 75
TABLE 20: Xpert MTB/RIF assay for rifampicin resistance detection in non-respiratory specimens. A. Evidence profile........ 77
TABLE 21: Meta-analysis. Estimated Xpert MTB/RIF Sensitivity and Specificity against the reference standard ‘Culture’ (published and unpublished studies) as well as against the reference standard ‘Clinical TB’ ................................................................. 83
TABLE 22: Meta-analysis: Estimated sensitivity and specificity of smear microscopy against a reference standard ‘Culture’. 84
TABLE 23: Meta-analysis. Xpert MTB/RIF against a reference standard ‘Culture’ in smear negative and smear positive children ........................................................................................................ 86
TABLE 24: Meta-analysis. Xpert MTB/RIF against a reference standard ‘Culture’ in HIV-infected and -uninfected children, subdivided by smear status ............................................................... 88
TABLE 25: Meta-regression model for Xpert MTB/RIF on ES/IS, controlling for smear and HIV status ............................. 89
TABLE 26: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for the detection TB in children against culture, where Xpert MTB/RIF is used as a replacement test for usual practice. A. ES/IS specimens B. GLA................................. 93
TABLE 27: Diagnostic accuracy of Xpert MTB/RIF for detection TB in children compared with a clinical reference standard, where Xpert MTB/RIF is used as a replacement test for usual practice .................................................. 98
TABLE 28: Diagnostic accuracy of Xpert MTB/RIF for detection of TB in children, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result .................................. 102
TABLE 29: Incremental yield of Xpert MTB/RIF compared with smear microscopy in children with culture confirmed TB.. 106
TABLE 30: Diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance in respiratory specimens in children ........................................................................................................... 108
2. BACKGROUND

Earlier and improved tuberculosis (TB) case detection, including smear-negative disease often associated with HIV-TB co-infection and young age, together with enhanced capacity to diagnose multidrug-resistant tuberculosis (MDR-TB) are global priorities for TB care and control. In September 2010, the WHO convened an Expert MTB/RIF Group to review the available evidence on the Xpert MTB/RIF® assay (Cepheid, Sunnyvale, USA) for the purpose of formulating recommendations to guide the use of the test. Policy recommendations on the Xpert MTB/RIF assay® (Xpert MTB/RIF) were issued by WHO early in 2011, supported by an operational ‘How-to’ document and a Checklist for country implementation.

In accordance with current WHO standards for evidence assessment in the formulation of policy recommendations, WHO engages in a systematic, transparent process using the GRADE approach (www.gradeworkinggroup.org). GRADE provides a structured framework for evaluating diagnostic test accuracy and the patient and public health impact of new diagnostic tests. An updated systematic review on the diagnostic accuracy of Xpert MTB/RIF for pulmonary TB and rifampicin resistance in adults was considered during the WHO Expert Group Meeting. Two additional systematic reviews were performed and were presented as separate reports: Xpert MTB/RIF assay for tuberculosis in non-respiratory specimens (extrapulmonary TB) and Xpert MTB/RIF assay for tuberculosis and rifampicin resistance in children.

**Xpert MTB/RIF**

Xpert MTB/RIF is an automated polymerase chain reaction (PCR) test (molecular test) utilizing the GeneXpert platform. Xpert MTB/RIF is a single test that can detect both *M. tuberculosis* complex and rifampicin resistance within two hours after starting the test, with minimal hands-on technical time. Unlike conventional nucleic acid amplification tests (NAATs), Xpert MTB/RIF is unique because sample processing and PCR amplification and detection are integrated into a single self-contained test unit, the Xpert MTB/RIF cartridge. Following sample loading, all steps in the assay are completely automated and self-contained. In addition, the assay’s sample reagent, used to liquefy sputum, has potent tuberculocidal (the ability to kill TB bacteria) properties and this largely eliminates biosafety concerns during the test procedure. These features allow the technology to be taken out of a reference laboratory and used nearer to the patient. Xpert MTB/RIF requires an uninterrupted and stable electrical power supply, temperature control, and yearly calibration of the instrument modules. The test procedure may be used directly on clinical specimens, either raw sputum samples or sputum pellets (also called sputum sediment), samples created after decontaminating and concentrating the sputum. In both cases, the test material is combined with the assay sample reagent, mixed by hand or vortex, and incubated at room temperature for 15 minutes. After the incubation step, 2 ml of the treated sample are transferred to the cartridge and the run initiated.

---


According to the manufacturer, Xpert MTB/RIF may be used with fresh sputum samples and fresh prepared sediments.

Xpert MTB/RIF uses molecular beacon technology to detect rifampicin resistance. Molecular beacons are nucleic acid probes that recognize and report the presence or absence of the normal, rifampicin-susceptible, ‘wild type’ sequence of the \textit{rpoB} gene of TB. Five different coloured beacons are used, each covering a separate nucleic acid sequence within the amplified \textit{rpoB} gene. When a beacon binds to the matching sequence, it fluoresces or ‘lights up’, which indicates the presence of one of the gene sequences that is characteristic of rifampicin-susceptible TB. Failure of the beacon to bind or delayed binding to the matching sequence indicates potential rifampicin resistance. The number and timing of detection of positive beacons (when the fluorescent signal rises above a pre-determined baseline cycle threshold) as well as results of sample processing controls allow the test to distinguish among the following results: ‘No TB’; ‘TB detected, rifampicin resistance detected’; ‘TB detected, no rifampicin resistance detected’; and an ‘invalid result’.

3. EVIDENCE BASE

3.1. Evidence synthesis
In order to facilitate policy guidance on the use of new diagnostic tools, new methods, and/or novel approaches using existing tools, WHO has developed a systematic, structured, evidence-based process. The first step involves a systematic review of available data, using standard methods appropriate for diagnostic accuracy studies. The second step involves the convening of an Expert Group to evaluate the strength of the evidence base and recommend operational and logistical considerations for mainstreaming such tools/approaches into national TB control programmes, and/or identify gaps to be addressed in future research. The third step involves WHO policy guidance on the use of these tools/approaches, presented to the WHO Strategic and Technical Advisory Group for TB (STAG-TB) for endorsement, and subsequent dissemination to Member States for implementation.

This document presents the findings and recommendations from the Expert Group meeting on Xpert MTB/RIF convened by WHO at Les Pensieres, Veyrier-du-Lac, France on 20-21st May 2013. The Expert Group (Annex 1) consisted of researchers, clinicians, epidemiologists, end-users (programme and laboratory representatives), a community representative and an evidence synthesis expert. The Expert Group meeting followed a structured agenda (Annex 2) and was chaired by a clinical epidemiologist with expertise and extensive experience in evidence synthesis and guideline development.

3.2. Meeting objectives
- To review the evidence base and evaluate data from an updated systematic review on the accuracy of Xpert MTB/RIF assay for the diagnosis of pulmonary TB and rifampicin resistance in adults;
- To review the evidence base and evaluate data from a systematic review on the accuracy of the Xpert MTB/RIF assay for the diagnosis of TB on non-respiratory samples;
- To review the evidence base and evaluate data from a systematic review on the accuracy of the Xpert MTB/RIF assay for the diagnosis of TB and rifampicin resistance in children;
- To review the evidence on the cost-effectiveness and affordability of the Xpert MTB/RIF in different epidemiological and resource settings;
- To outline issues to be addressed by WHO in subsequent policy recommendations.
3.3. GRADE evaluation

To comply with current standards for evidence assessment in formulation of policy recommendations, the GRADE system ([www.gradeworkinggroup.org](http://www.gradeworkinggroup.org)), adopted by WHO for all policy and guidelines development, was used. The GRADE approach, assessing both the quality of evidence and strength of recommendations, aims to provide a comprehensive and transparent approach for developing policy guidance.

Evaluation followed the GRADE system for grading quality of evidence and strength of recommendations for diagnostic tests. The quality of evidence was evaluated according to six criteria:

- **Overall study design**
  - Cross-sectional: Random or consecutive selection of patients/specimens at risk (preferred); Case-control: Selection of patients/specimens according to reference standard.

- **Risk of bias or limitations in study design and execution (as reflected by the QUADAS 2 tool)**

- **Directness**
  - Presence of direct evidence of impact on patient-important outcomes and generalisability in relation to the population, the diagnostic test used, the comparator of the test and whether tests were directly or indirectly compared.

- **Inconsistency**
  - Unexplained inconsistency in sensitivity or specificity estimates.

- **Imprecision**
  - Wide confidence intervals for pooled sensitivity or specificity estimates.

- **Publication bias**
  - Publications of research based on their nature and outcome, e.g. studies showing poor performance not being published, language bias, etc.

As called for by GRADE, the Expert Group also considered the strength of the recommendation (strong or conditional), based on a balance of effects (advantages weighed against disadvantages), patient values and preferences, and costs.

Using the GRADE framework, sensitivity and specificity results were interpreted as proxy measures for patient-important outcomes based on the relative importance/impact of false-positive and false-negative results. Poor sensitivity would result in false-negative results where TB and MDR-TB patients would be missed with negative consequences for morbidity, mortality and transmission of disease. Poor specificity would result in false-positive results where patients without TB or MDR-TB would be prescribed unnecessary treatment, with negative consequences such as serious adverse events related to second-line anti-TB drugs.

Rates for true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were calculated based on pre-test probabilities, i.e. an assumed prevalence of 2.5%, 5% and 10% of TB among suspects being screened for TB, and an assumed prevalence of 5% and 15% for rifampicin resistance (as a proxy for MDR) among patients with confirmed TB. Patient impact was based on a balance between the following values:

**True positives**: Benefit to patients from rapid diagnosis and treatment;

**True negatives**: Patients spared unnecessary treatment; benefit of reassurance and alternative diagnosis;

---

False positives: Likely patient anxiety and morbidity from additional testing, unnecessary treatment; may halt further diagnostic evaluation;

False negatives: Increased risk of patient morbidity and mortality, and continued risk of community transmission of TB.

3.3.1. PICO Questions for each review
Evaluation of the available evidence followed the GRADE system for grading quality of evidence and providing strength of recommendations, based on the formulation of priori agreed questions (the PICO question) by the Expert Group. PICO refers to the following four elements that should be in a question governing a systematic search of the evidence, and was defined as follows: Population targeted by the action/intervention; Intervention; Comparator; and Outcome. PICO questions for each review are given below:

Review 1: Updated systematic review: Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults
1. What is the diagnostic accuracy of Xpert MTB/RIF) for detection of pulmonary TB in adults, where Xpert MTB/RIF is used as a replacement test for smear microscopy?
2. What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result?
3. What is the diagnostic accuracy of Xpert MTB/RIF for detection of smear-positive pulmonary TB in adults?
4. What is the diagnostic accuracy of Xpert MTB/RIF for detection of smear-negative (culture-positive) pulmonary TB in adults?
5. What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in people living with HIV (adults)?
6. What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults without HIV infection?
7. What is the diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing?

Review 2. Systematic review: Xpert MTB/RIF for diagnosis of tuberculosis and detection of rifampicin resistance on non-respiratory samples (extra-pulmonary TB)
1. What is the diagnostic accuracy of Xpert MTB/RIF overall compared with culture for non-respiratory specimens, where Xpert MTB/RIF is used as a replacement test for usual practice*?
2. What is the diagnostic accuracy of Xpert MTB/RIF overall compared with a combined clinical and laboratory reference standard for non-respiratory specimens, where Xpert MTB/RIF is used as a replacement test for usual practice?
2a. What is the diagnostic accuracy of Xpert MTB/RIF for lymph node fluid and tissue, where Xpert MTB/RIF is used as a replacement test for usual practice?
2b. What is the diagnostic accuracy of Xpert MTB/RIF for pleural fluid, where Xpert MTB/RIF is used as a replacement test for usual practice?
2c. What is the diagnostic accuracy of Xpert MTB/RIF for cerebrospinal fluid (CSF), where Xpert MTB/RIF is used as a replacement test for usual practice?

* Most analyses will be performed using two reference standards: culture (the current reference standard) and a combined clinical and laboratory reference standard chosen by the study authors (given technical limitations of culture diagnosis).
2d. What is the diagnostic accuracy of Xpert MTB/RIF for gastric fluid, where Xpert MTB/RIF is used as a replacement test for usual practice?
2e. What is the diagnostic accuracy of Xpert MTB/RIF for tissue samples, where Xpert MTB/RIF is used as a replacement test for usual practice?
3. What is the diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance in non-respiratory specimens, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing?


1. What is the diagnostic accuracy of Xpert MTB/RIF for detection TB in children compared with culture, where Xpert MTB/RIF is used as a replacement test for usual practice? 
2. What is the diagnostic accuracy of Xpert MTB/RIF for detection TB in children compared with a combined clinical and laboratory reference standard, where Xpert MTB/RIF is used as a replacement test for usual practice? 
3. What is the diagnostic accuracy of Xpert MTB/RIF for detection of TB in children, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result? 
4. What is the diagnostic accuracy of Xpert MTB/RIF compared with smear microscopy for detection of TB in children? 
5. What is the diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance in children, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing? 
6. What is the diagnostic accuracy of Xpert MTB/RIF for detection of peripheral lymph node TB in children, where Xpert MTB/RIF is used as a replacement test for usual practice? 
7. What is the diagnostic accuracy of Xpert MTB/RIF for detection of TB meningitis in children, where Xpert MTB/RIF is used as a replacement test for usual practice? 


1. For which diagnostic and screening algorithms is the Xpert MTB/RIF assay an affordable and a cost-effective intervention?

---

**3.3.2. Determination of relative importance of patient outcomes:**

PICO questions were drafted by the WHO Steering Group and were presented to the Expert Group for discussion and modification. The WHO Steering Group also prepared an initial list of relevant outcomes, including desirable and undesirable effects and requested the Expert Group to identify any other outcomes. A webinar was conducted with the Expert Group prior to the meeting to refine and finalise the proposed patient outcomes and to rate their relative importance. The following outcomes for each PICO question were determined and rating importance were unanimously agreed by the Expert Group:

---

[Given the difficulties of TB diagnosis in children, *usual practice* refers to customary practice in the field, which may vary from setting to setting. Usual practice for children (aged 0-15 years) suspected of having intrathoracic (i.e., pulmonary, pleural, and mediastinal or hilar lymph node) tuberculosis normally requires bacteriological confirmation through examination of sputum (by expectoration, gastric washings, or induced sputum) for smear microscopy and culture. In the event of negative bacteriological results, a diagnosis of tuberculosis may be based on the presence of abnormalities consistent with tuberculosis on chest radiography, a history of exposure to an infectious case, evidence of tuberculosis infection (positive tuberculin skin test or interferon-gamma release assay), and clinical findings suggestive of tuberculosis. For children suspected of having extrapulmonary tuberculosis, appropriate specimens from the suspected sites of involvement may be obtained for microscopy and for culture and histopathological examination.]
• Critical outcomes: Diagnostic accuracy as reflected by true positive (TP), true negative (TN), false positive (FP) and false negative (FN) results; Time to diagnosis.
• Important outcome: Cost

3.3.3. Assessment of study quality
An appraisal of the included studies used the Quality Assessment of diagnostic Accuracy Studies (QUADAS - 2) tool which consists of four domains: patient selection, index test, reference standard and flow and timing.

3.4. Meeting procedural issues

A meeting of the Expert Group was convened by WHO in Veyrier-du-Lac, France, May 20-21 2013. The meeting was chaired by an evidence synthesis expert. Decisions were based on consensus. Concerns and opinions by members were noted and included in the final meeting report. The detailed meeting report was prepared by the WHO Steering Group, with several iterations before final sign-off by Expert Group members.

The systematic reviews and reports were made available to the Expert Group for scrutiny before the meeting and all full copies of the reviews were made available to the Expert Group during the meeting.

As agreed, interchange by Expert Group meeting participants was restricted to those who attended the Expert Group meeting in person, both for the discussion and follow-up dialogue. It was explained that individuals were selected to be members of the Expert Group to carefully represent and balance important perspectives for the process of formulating recommendations. Therefore the Expert Group included technical experts, end-users, patient representatives and evidence synthesis methodologists.

The Expert Group members were asked to submit completed Declaration of Interest (DOI) form prior to the meeting. The completed DOI forms were evaluated by the WHO Steering Group before any formal invitation to the meetings was issued. A review of each DOI was conducted by the WHO Steering Group to determine if an interest had been declared and if so, whether it was insignificant or whether it was potentially significant. If the WHO Steering Group determined that no relevant interest had been declared or such interest was insignificant or minimal a letter of invitation to participate in the Expert Group was subsequently issued.

When the review by the WHO Steering Group indicated that a declared interest was significant or potentially significant the WHO Legal Department was consulted and their advice was followed for the meeting procedures. DOI statements were summarized by the Chair at the start of the meeting. Individuals determined to have significant relevant interests were invited as observers to provide technical input and answer technical questions. However, these individuals were not permitted to participate in the formulation of recommendations. In addition, they were excluded from the development of the final Guideline Development Group meeting report and for the preparation of the final WHO policy guidance. A summary is attached in Annex 3.

Selected individuals with intellectual and/or research involvement in Xpert MTB/RIF were invited as observers to provide technical input and answer technical questions. These individuals did not participate in the GRADE evaluation process at the Expert Group meeting nor during the final discussions when recommendations were developed. They were also not involved in the development of the Expert Group meeting report, nor in preparation of the STAG-TB documentation or preparation of the final WHO policy statement.
4. Results

4.1. Xpert MTB/RIF for the diagnosis of pulmonary TB

4.1.1. Study characteristics

Studies that assessed the diagnostic accuracy of Xpert MTB/RIF for pulmonary TB and/or rifampicin resistance, typically are cross-sectional in design, comparing Xpert MTB/RIF to an the reference standard (defined below) where true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values could be determined were included. Participants in the included studies were adult or predominantly adult patients, 15 years of age or older, presumed to have pulmonary TB or MDR-TB, with or without HIV infection. Studies which included other respiratory samples such as samples obtained by bronchoalveolar lavage were included. Data on specimens obtained by gastric aspiration were excluded.

The reference standard for pulmonary TB was culture on solid media or a commercial liquid culture system: such as BACTEC™ MGIT™ (mycobacterial growth indicator tube) 960 Mycobacterial Detection System, (BD, USA). The reference for rifampicin resistance was WHO recommended conventional phenotypic drug susceptibility testing (DST) on solid or liquid media (WHO Policy DST 2008).

Two literature searches were performed on 25 September 2011 and 15 December 2011 from which 18 studies were identified. These 18 studies were included in the Cochrane review published 31 January 2013\(^\text{10}\). A third literature search was performed on 7 February 2013 which identified nine additional studies. Annex 4 shows the PRISMA diagram with the flow of the studies. From the three literature searches 27 relevant studies (26 published studies and one unpublished study) involving 36 study centres were evaluated. Annex 4 lists the included and excluded studies along with the reason for exclusion of any studies.

For TB detection, the 27 studies included 9558 participants. Of the total 27 studies, 24 studies (33 study centres) including 2969 participants provided data for rifampicin resistance detection. Of the three studies that were not included, one study presented combined results for pulmonary and extrapulmonary specimens, one study did not report information on rifampicin resistance, and one study did not use the defined reference standard. Seven of the 27 studies detected no rifampicin resistance with the reference standard.

4.1.2. Study Quality

The quality of included studies was assessed with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. As recommended, all domains (patient selection, index test, reference standard, and flow and timing) were assessed in terms of risk of bias and the first domains were also assessed in terms of concerns about applicability.

In the patient selection domain, 28 of 36 study centres (78%) were considered to be at low risk of bias because participants were enrolled consecutively and these study centres avoided inappropriate exclusions. The remaining study centres were considered to be at high risk of bias, because either 1) the manner of patient selection was by convenience (five study centres) or not stated (one study) or 2) the study pre-selected smear-positive patients (two study centres). With regard to applicability (patient characteristics and setting), 26 of the 36 study centres (72%) (corresponding to 18 of the 27 studies, 67%) were judged to be of low concern because these study centres performed Xpert MTB/RIF in intermediate or peripheral-level laboratories associated with primary care clinics. The remaining study centres were judged

as unclear concern. These studies either did not provide any clinical information (one study) or ran Xpert MTB/RIF in central-level laboratories where culture (the reference standard) could also be performed (nine studies).

In the index test (i.e., Xpert MTB/RIF) domain, all study centres were considered to be at low concern for both risk of bias and applicability. In the reference standard domain, 33 study centres (92%) were deemed to be at low risk of bias for TB and 34 study centres (94%) to be at low risk of bias for rifampicin resistance because the reference standard results were interpreted without knowledge of the results of the Xpert MTB/RIF assay.

Applicability was considered to be of low concern for all studies in the reference standard domain. In the flow and timing domain, 32 study centres (89%) were considered to be of low concern for risk of bias because all patients were accounted for in the analysis and information about uninterpretable results was provided. Inconclusive (errors, invalid, no results) MTB/RIF test results were excluded from the analyses for determination of sensitivity and specificity for both TB detection and rifampicin resistance detection. Although the pooled rate of inconclusive results reported in the studies in the meta-analysis was considered to be low, there were some concerns regarding a potential bias, given observations from the field that rates might be higher (up to 8%). Figure 1 shows the overall quality of the 36 study centres.

**Figure 1**: Risk of bias and applicability judgements regarding each domain presented as percentages across the 36 included study centres (27 studies).
4.1.3. Xpert MTB/RIF used as an initial test replacing smear microscopy

Forest plots of Xpert MTB/RIF sensitivity and specificity for TB detection for the total 27 studies (36 study centres) are presented in Figure 2. Sensitivity estimates varied from 58% to 100% and specificity estimates, from 86% to 100%.

**Figure 2:** Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection.

TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets the 95% CI of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line).

22 of the total 27 studies, involving 9008 participants were included in this meta-analysis. Five studies that enrolled primarily smear-positive or smear-negative patients were excluded. For TB detection, Xpert MTB/RIF pooled sensitivity and specificity were 88% (95% CrI 84% - 92%) and 99% (95% CrI 98% - 99%), respectively (Table 1).

Twenty-one studies (8880 participants) provided data from which to compare the sensitivity of Xpert MTB/RIF and smear microscopy. For smear microscopy, the pooled sensitivity was 65% (95% CrI 57% - 72%) For Xpert MTB/RIF, the pooled sensitivity was 88% (95% CrI 84% - 92%). Therefore, in comparison with smear microscopy, Xpert MTB/RIF increased TB detection among culture-confirmed cases by 23% (95% CrI 15% - 32%). When Xpert MTB/RIF was used as an add-on test following a negative smear microscopy result, the pooled sensitivity was 68% (95% CrI 61% - 74%) and the pooled specificity was 99% (95% CrI 98% - 99%), (23 studies, 7151 participants) (Table 1). In other words, 68% of smear-negative culture-confirmed TB patients would have been missed.
cases were detected using Xpert MTB/RIF following smear microscopy, increasing case detection by 68% (95% CrI, 61% - 74%) in this group.

Table 1: Xpert MTB/RIF assay for detection of TB and rifampicin resistance

<table>
<thead>
<tr>
<th>Type of Analysis</th>
<th>Pooled sensitivity Median (95% CrI)</th>
<th>Pooled specificity Median (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xpert MTB/RIF used as an initial test for TB detection replacing microscopy (22, 9008)</td>
<td>88% (84, 92)</td>
<td>99% (98, 99)</td>
</tr>
<tr>
<td>Xpert MTB/RIF used as an add-on test for TB detection following a negative smear microscopy result (23, 7151)</td>
<td>68% (61, 74)</td>
<td>99% (98, 99)</td>
</tr>
<tr>
<td>Xpert MTB/RIF used as an initial test for rifampicin resistance detection replacing conventional drug susceptibility testing as the initial test *</td>
<td>95% (90, 97)</td>
<td>98% (97, 99)</td>
</tr>
</tbody>
</table>

CrI, credible interval, is the Bayesian equivalent of the confidence interval, CI

* For rifampicin resistance detection, pooled sensitivity and specificity estimates were determined separately by univariate analyses. Pooled sensitivity included 17 studies (555 participants) and pooled specificity, 24 studies (2414 participants).
4.1.4. Investigations of heterogeneity - TB detection in smear-positive and smear-negative individuals suspected of having TB

4.1.4.1. Smear-positive TB

There was little heterogeneity in sensitivity estimates (95-100%) for studies reporting smear-positive data (24 studies, 33 study centres, 2071 participants). In the meta-analysis, the pooled sensitivity for smear-positive, culture-positive TB was very high, 98% (95% CrI 97% - 99%) (23 studies, 1952 participants). Estimates of Xpert MTB/RIF pooled specificity were not performed as studies of the smear-positive subgroup as almost all participants were considered to be true TB positive.

4.1.4.2. Smear-negative TB

Figure 3 displays the forest plots for studies reporting smear-negative data (24 studies, 33 study centres, 7247 participants). There was considerable variability in sensitivity estimates, range 43% to 100%. Specificity estimates showed less variation, range 86% to 100%. The meta-analysis included 23 studies with
a direct comparison between smear-positive and smear-negative subgroups. The pooled sensitivity estimate for smear-negative, culture-positive TB was 68% (95% CrI 61% - 74%), considerably lower than the pooled sensitivity estimate for smear-positive, culture-positive TB, 98% (95% CrI 97% - 99%) (Table 2). The 95% CrI for the difference in Xpert MTB/RIF sensitivity in smear-positive and smear-negative subgroups did not cross 0 suggesting this finding is statistically significant (Table 2).

Table 2: Impact of covariates on heterogeneity of Xpert MTB/RIF sensitivity and specificity, TB detection

<table>
<thead>
<tr>
<th>Covariate (No. of Studies)</th>
<th>Pooled Sensitivity Median (95% Credible Interval)</th>
<th>Pooled Specificity Median (95% Credible Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smear status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smear + (n=23)</td>
<td>98% (97, 99)</td>
<td>***</td>
</tr>
<tr>
<td>Smear - (n=23)</td>
<td>68% (61, 74)</td>
<td>99% (98, 99)</td>
</tr>
<tr>
<td>Difference (Smear+ minus Smear-)</td>
<td>31% (24, 37)</td>
<td>***</td>
</tr>
<tr>
<td>P (Smear+ &gt; Smear-)</td>
<td>1.00</td>
<td>***</td>
</tr>
<tr>
<td><strong>HIV status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV- (n=7)</td>
<td>86% (76, 92)</td>
<td>99% (98, 100)</td>
</tr>
<tr>
<td>HIV+ (n=7)</td>
<td>79% (70, 86)</td>
<td>98% (96, 99)</td>
</tr>
<tr>
<td>Difference (HIV- minus HIV+)</td>
<td>7% (-5, 18)</td>
<td>1% (-1, 3)</td>
</tr>
<tr>
<td>P (HIV- &gt; HIV+)</td>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Condition of specimen</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresh (n=13)</td>
<td>89% (83, 93)</td>
<td>99% (98, 100)</td>
</tr>
<tr>
<td>Frozen (n=6)</td>
<td>85% (78, 92)</td>
<td>98% (95, 99)</td>
</tr>
<tr>
<td>Difference (Fresh minus Frozen)</td>
<td>3% (-5, 14)</td>
<td>1% (-0.4, 4)</td>
</tr>
<tr>
<td>P (Fresh &gt; Frozen)</td>
<td>0.78</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Specimen preparation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unprocessed (n=11)</td>
<td>90% (84, 93)</td>
<td>98% (97, 99)</td>
</tr>
<tr>
<td>Processed (n=11)</td>
<td>88% (81, 93)</td>
<td>99% (98, 100)</td>
</tr>
<tr>
<td>Difference (Unprocessed minus Processed)</td>
<td>2% (-5, 10)</td>
<td>-1% (-2, 1)</td>
</tr>
<tr>
<td>P (Unprocessed &gt; Processed)</td>
<td>0.71</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>Proportion TB cases in the study</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 30% (n=12)*</td>
<td>90% (85, 93)</td>
<td>98% (96, 99)</td>
</tr>
<tr>
<td>≤ 30% (n=10)</td>
<td>86% (79, 92)</td>
<td>99% (98, 100)</td>
</tr>
<tr>
<td>Difference (&gt;30% minus ≤ 30%)</td>
<td>3% (-4, 11)</td>
<td>-1% (-3, 0.3)</td>
</tr>
<tr>
<td>P (&gt;30% minus ≤ 30%)</td>
<td>0.81</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Country income level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-income (n=8)</td>
<td>92% (87, 96)</td>
<td>99% (97, 99)</td>
</tr>
<tr>
<td>Low- and middle-income (n=14)</td>
<td>86% (81, 91)</td>
<td>99% (97, 99)</td>
</tr>
<tr>
<td>Difference (High-income minus Low- and middle-income)</td>
<td>6% (-1, 12)</td>
<td>0.1% (-2, 2)</td>
</tr>
<tr>
<td>P (High-income &gt; Low- and middle-income)</td>
<td>0.96</td>
<td>0.56</td>
</tr>
</tbody>
</table>

* 30% was selected as a cut-off based on the median proportion in the included studies

***Values could not be determined
4.1.5. TB detection in HIV-negative and HIV-positive individuals suspected of having TB

Figure 4 displays the forest plots for studies reporting data for HIV-negative individuals (nine studies, 18 study centres, 2555 participants) and HIV-positive individuals (10 studies, 16 study centres, 2378 participants). There was variability in sensitivity in both the HIV-negative subgroup (56% to 100%) and HIV-positive subgroup (0% to 100%). The small number of participants in several studies may contribute to some of the variability. Specificity varied less than sensitivity in both subgroups: 96% to 100% in HIV-negative subgroup and 92% to 100% in HIV-positive subgroup.

The meta-analysis included seven studies that provided data for both HIV-negative (1470 participants) and HIV-positive (1789 participants) individuals. The pooled sensitivity was 86% (95% Crl 76% - 92%) in the HIV-negative subgroup and 79% (95% Crl 70% - 86%) in the HIV-positive subgroup (Table 2). Corresponding pooled specificities were similar: 99%, (95% Crl 98% - 100%) and 98%, (95% Crl 96% - 99%), respectively (Table 3). When adjusting for the percentage of smear-positive patients in each study, the impact of the HIV covariate decreased suggesting that some of the differences between the HIV-positive and HIV-negative subgroups could be attributed to differences in smear status.

4.1.6. TB detection among HIV-positive individuals by smear status

Five studies reported data from which to assess the accuracy of Xpert MTB/RIF in HIV-positive individuals with culture-positive, smear-negative TB. Xpert MTB/RIF sensitivity ranged from 43% to 93% for smear-negative, culture positive TB and from 91% to 100% for smear-positive, culture positive TB. Data were sufficient to perform a univariate meta-analysis for Xpert MTB/RIF sensitivity. Among people living with HIV, Xpert MTB/RIF pooled sensitivity was 61% (95% Crl 42% - 79%) for smear-negative, culture-positive TB compared with 97% (95% Crl 91% - 99%) for smear-positive, culture-positive TB, a statistically significant result (data not shown). Hence, among people with HIV-TB co-infection, people with smear-positive disease were more likely to be diagnosed with TB using Xpert MTB/RIF than those with smear-negative disease.
Figure 4: Forest plots of Xpert MTB/RIF sensitivity and specificity for TB detection in HIV-negative and HIV-positive individuals suspected of having TB.

The squares represent the sensitivity and specificity of one study and the black line represent its CI. TP = true positive; FP = false positive; FN = false negative; TN = true negative.

### HIV negative

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Artaeh 2012</td>
<td>42</td>
<td>0</td>
<td>2</td>
<td>127</td>
<td>0.95 [0.85, 0.99]</td>
<td>1.00 [0.97, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010a</td>
<td>90</td>
<td>0</td>
<td>18</td>
<td>46</td>
<td>0.83 [0.75, 0.90]</td>
<td>1.00 [0.92, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010b</td>
<td>142</td>
<td>0</td>
<td>5</td>
<td>24</td>
<td>0.97 [0.92, 0.99]</td>
<td>1.00 [0.86, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010c</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>1.00 [0.85, 1.00]</td>
<td>1.00 [0.87, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010d</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>69</td>
<td>0.83 [0.36, 1.00]</td>
<td>0.99 [0.92, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010e</td>
<td>75</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>0.97 [0.91, 1.00]</td>
<td>1.00 [0.63, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011a</td>
<td>161</td>
<td>3</td>
<td>20</td>
<td>252</td>
<td>0.89 [0.83, 0.93]</td>
<td>0.99 [0.87, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011b</td>
<td>36</td>
<td>1</td>
<td>2</td>
<td>202</td>
<td>0.95 [0.82, 0.99]</td>
<td>1.00 [0.97, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011c</td>
<td>62</td>
<td>1</td>
<td>3</td>
<td>232</td>
<td>0.95 [0.87, 0.99]</td>
<td>1.00 [0.98, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011d</td>
<td>41</td>
<td>0</td>
<td>5</td>
<td>56</td>
<td>0.89 [0.76, 0.96]</td>
<td>1.00 [0.94, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011e</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1.00 [0.16, 1.00]</td>
<td>1.00 [0.16, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011f</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0.87 [0.92, 0.99]</td>
<td>1.00 [1.00, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanrahan 2013</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>182</td>
<td>0.56 [0.21, 0.86]</td>
<td>1.00 [0.88, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rachow 2011</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>53</td>
<td>0.89 [0.87, 0.93]</td>
<td>1.00 [0.93, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safenworsa 2012</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>127</td>
<td>0.88 [0.64, 0.99]</td>
<td>0.99 [0.86, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott 2011</td>
<td>12</td>
<td>0</td>
<td>2</td>
<td>17</td>
<td>0.86 [0.57, 0.98]</td>
<td>1.00 [0.80, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theron 2011</td>
<td>87</td>
<td>9</td>
<td>14</td>
<td>195</td>
<td>0.83 [0.73, 0.90]</td>
<td>0.96 [0.92, 0.98]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Rie 2013</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>33</td>
<td>0.67 [0.92, 0.99]</td>
<td>1.00 [0.89, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### HIV positive

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bailey 2012</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>47</td>
<td>0.92 [0.62, 1.00]</td>
<td>0.98 [0.89, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010a</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1.00 [0.59, 1.00]</td>
<td>1.00 [0.16, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010b</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.00 [0.00, 0.97]</td>
<td>1.00 [0.03, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010c</td>
<td>60</td>
<td>0</td>
<td>6</td>
<td>81</td>
<td>0.91 [0.81, 0.97]</td>
<td>1.00 [0.96, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010d</td>
<td>27</td>
<td>2</td>
<td>6</td>
<td>141</td>
<td>0.82 [0.65, 0.93]</td>
<td>0.99 [0.95, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010e</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.75 [0.19, 0.99]</td>
<td>Not estimable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2010f</td>
<td>90</td>
<td>1</td>
<td>18</td>
<td>263</td>
<td>0.83 [0.75, 0.90]</td>
<td>1.00 [0.98, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011d</td>
<td>80</td>
<td>0</td>
<td>19</td>
<td>88</td>
<td>0.81 [0.72, 0.88]</td>
<td>1.00 [0.96, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoehe 2011e</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>31</td>
<td>1.00 [0.29, 1.00]</td>
<td>0.94 [0.80, 0.99]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curiosity 2012</td>
<td>44</td>
<td>2</td>
<td>1</td>
<td>84</td>
<td>0.98 [0.88, 1.00]</td>
<td>0.98 [0.92, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanrahan 2013</td>
<td>36</td>
<td>2</td>
<td>16</td>
<td>325</td>
<td>0.69 [0.55, 0.81]</td>
<td>0.99 [0.98, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawn 2011</td>
<td>42</td>
<td>2</td>
<td>30</td>
<td>320</td>
<td>0.58 [0.46, 0.70]</td>
<td>0.99 [0.96, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rachow 2011</td>
<td>41</td>
<td>1</td>
<td>9</td>
<td>49</td>
<td>0.82 [0.69, 0.91]</td>
<td>0.99 [0.89, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott 2011</td>
<td>43</td>
<td>3</td>
<td>7</td>
<td>84</td>
<td>0.87 [0.74, 0.94]</td>
<td>0.97 [0.90, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theron 2011</td>
<td>32</td>
<td>7</td>
<td>14</td>
<td>77</td>
<td>0.70 [0.54, 0.82]</td>
<td>0.92 [0.84, 0.97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Rie 2013</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>99</td>
<td>0.67 [0.35, 0.90]</td>
<td>0.99 [0.95, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4.1.7. Effect of condition of the specimen

Although the manufacturer recommends the use of fresh specimens, some studies had been conducted using frozen specimens. The effect of the condition of the specimen on Xpert MTB/RIF performance was subsequently explored. The pooled sensitivity was 89% (95% CI 83% - 93%) in studies using fresh specimens (13 studies), slightly higher than the pooled sensitivity of 85% (95% CI 75% - 92%) in studies using frozen specimens (six studies) (Table 2). The pooled specificity was 99% (95% CI 98% - 100%) for fresh specimens and 98% (95% CI 95% - 99%) for frozen specimens. The difference in Xpert MTB/RIF sensitivity and specificity for fresh and frozen specimens was not significantly different from 0 (Table 2).
Meta-regression modelling suggested that, once smear-status was taken into account, there was no conclusive evidence supporting the impact of the condition of the specimen on Xpert MTB/RIF sensitivity.

4.1.8. Effect of specimen preparation
The pooled sensitivity estimate was 90% (95% CrI 84% - 93%) in studies using unprocessed specimens (eleven studies), higher than the pooled sensitivity estimate of 88% (95% CrI 81% - 93%) in studies using processed specimens (eleven studies) (Table 2). The pooled specificity was 98% (95% CrI 97% - 99%) for unprocessed specimens, similar to the pooled specificity of 99% (95% CrI 98% - 100%) for processed specimens. The difference in Xpert MTB/RIF sensitivity and specificity for processed and unprocessed specimens was not significantly different from 0.

4.1.9. Effect of the proportion of culture-confirmed TB cases in the study
For this analysis, a cut-off of 30% TB cases in the study was used because 30% was around the median proportion in the included studies. 12 studies were found to have proportion of TB cases ≥ 30% and 10 studies to have proportion of TB cases ≤ 30%. The pooled sensitivity was 90% (95% CrI 85% -93%) for studies with >30% and 86% (95% CrI 79% -92%) for studies with ≤30% (Table 2). The corresponding pooled specificity estimates were similar, 98% (95% CrI 96% -99%) and 99% (95% CrI 98% -100%), respectively (Table 2). After adjustment for smear status, the probability of any differences in accuracy was further decreased suggesting there was no conclusive evidence of the impact of TB prevalence on Xpert MTB/RIF performance.

4.1.10. Effect of country income status
The pooled sensitivity for studies in high-income countries (eight studies) was 92% (95% CrI 87% -96%), higher than the pooled sensitivity for studies in low/middle-income countries (14 studies), 86% (95% CrI 81% -91%) (Table 2). However, after further adjustment for smear status, there was no conclusive evidence supporting the impact of country income status on Xpert MTB/RIF sensitivity.

4.1.11. Rifampicin resistance detection
4.1.11.1. Xpert MTB/RIF used as an initial test replacing conventional drug susceptibility testing
The 24 studies (33 study centres) in the analysis included 555 rifampicin-resistant specimens. Figure 5 shows the forest plots of sensitivity and specificity for this analysis. The individual study centres in the plots are ordered by decreasing sensitivity and decreasing number of true positive results. Although, there was heterogeneity in sensitivity estimates (ranging from 33% to 100%), in general there was less variability among study centres with a higher number of rifampicin-resistant specimens. Specificity showed less variability than sensitivity, ranging from 83% to 100%. The pooled sensitivity and specificity by univariate analysis were 95% (95% CrI 90% - 97%) and 98% (95% CrI 97% - 99%), respectively (Table 1). The pooled sensitivity and specificity were the same by bivariate analysis of the subset of studies that provided data for both sensitivity and specificity (17 studies, 2624 participants).

Figure 5: Forest plots of Xpert MTB/RIF sensitivity and specificity for detection of rifampicin resistance, Xpert MTB/RIF used as an initial test replacing phenotypic culture-based drug susceptibility testing as the initial test.

The individual studies are ordered by decreasing sensitivity and decreasing number of true positives. The squares represent the sensitivity and specificity of one study, the black line its CI. TP = true positive; FP = false positive; FN = false negative; TN = true negative.
4.1.12. Investigations of heterogeneity, rifampicin resistance detection

4.1.12.1. Effect of the Xpert MTB/RIF assay version

The basis in the Xpert MTB/RIF system for rifampicin resistance detection is the difference between the first (early cycle threshold) and the last (late cycle threshold) of the \textit{M. tuberculosis}-specific beacon (probe). This difference is referred to as the delta cycle threshold. The original Xpert MTB/RIF system configuration reported rifampicin resistance when the delta cycle threshold was $>3.5$ cycles and rifampicin sensitive when the delta cycle threshold was $\leq 3.5$ cycles (Xpert MTB/RIF G1 assay). After May 2010, the manufacturer modified the delta cycle threshold cut-off to improve Xpert MTB/RIF specificity for rifampicin resistance detection (Xpert MTB/RIF G2 and G3 assays). Another modification was implemented in late 2011 (Xpert MTB/RIF G4 assay) which modified the molecular beacon sequence of Probe B for improved rifampicin detection in the event of annealing temperature fluctuations. Fluidic and software changes virtually eliminated signal loss detection error (5011 errors) and allowed for high sensitivity and specificity for TB and rifampicin resistance detection to be retained. These assay enhancements\textsuperscript{11} were considered to be part of a routine process in the product improvement cycle. Cepheid, FIND and the University of Medicine and Dentistry of New Jersey (UMDNJ) will continue to monitor clinical performance. The G4 cartridges are now the only type of cartridge available for use.

\textsuperscript{11}The FIND report on the enhanced assay is available at: http://www.stoptb.org/wg/gli/assets/documents/map/findg4cartridge.pdf
The effect of the Xpert MTB/RIF version on the sensitivity and specificity estimates for rifampicin resistance detection was subsequently investigated. The pooled sensitivity was 93% (95% CrI 87% - 97%) for studies using Xpert MTB/RIF G2, G3, or G4 assays (13 studies) and 97% (95% CrI 91% - 99%) for studies using Xpert MTB/RIF G1 assay (4 studies). The corresponding pooled specificities were 98% (95% CrI 96% - 99%) (15 studies) and 99% (95% CrI 98% - 100%) (four studies). The overlapping confidence intervals indicate no statistically significant difference in either the sensitivity or specificity estimates for the Xpert MTB/RIF G1 assay compared with later assays versions.

4.1.12.2. Accuracy of Xpert MTB/RIF G4 cartridge

Two studies used the Xpert MTB/RIF G4 assay and provided data for specificity determinations. One study observed a specificity of 100% (10/10) (95% CI 69% - 100%) while the second study reported a specificity of 95% (42/44) (95% CI 85% - 99%) (Figure 5).

FIND evaluated the diagnostic accuracy of Xpert MTB/RIF G4\textsuperscript{14} in a study that involved testing 233 archived sputum specimens from individuals with presumed TB stored in Borstel Germany, 184 frozen AFB positive sediments from Lima, Peru, as well as frozen sputum specimens from a further 231 patients consecutively enrolled from Baku, Azerbaijan all of which were shipped and tested in Borstel, Germany using the G4 assay. Fresh sputum from 30 patients were tested on both the G4 and G3 assay in Uganda and a further 218 specimens were evaluated using both the G4 and G3 assay version in Cape Town, South Africa.

The reference standard across all sites included at least one Löwenstein-Jensen (LJ) culture and at least one MGIT culture with confirmation of \textit{M. tuberculosis} species by Capilia (Tauns Laboratories, Shizuoka, Japan), GenoType MTBDR\textit{plus} (Hain Lifescience, Nehren, Germany) or GenoType Mycobacterium CM/AS (Hain Lifescience, Nehren, Germany). Conventional rifampicin resistance testing was performed by either the LJ proportion method or MGIT and, in a few cases, the Genotype MTBDR\textit{plus} assay only. Genetic sequencing was performed on discordant results between Xpert MTB/RIF and conventional DST. Six patients (smear and culture negative) were started on anti-TB treatment and excluded from the analysis. Genetic sequencing was used to resolve discordant results for determination of sensitivity and specificity. The overall sensitivity (rifampicin resistant) was 98.9% (87/88) (95% CI 93.8% - 99.8%) and the overall specificity (rifampicin sensitive) was 99.8% (433/434) (95% CI 98.7% - 100.0%). For Xpert MTB/RIF rifampicin-sensitive/DST-resistant discordants, sequencing of the \textit{rpoB} region was performed in four cases and discordant results resolved in three of these cases in favour of Xpert MTB/RIF; for Xpert MTB/RIF rifampicin-resistant/DST sensitive discordants, sequencing of the \textit{rpoB} region was performed in nine cases and discordant results resolved in eight of these cases in favour of Xpert MTB/RIF.

4.1.12.3. Accuracy of the reference standards used

Culture is regarded as the best available reference standard for active TB and was the reference standard used for TB in this review. Phenotypic culture-based DST methods using WHO recommended critical concentrations were the reference standards for rifampicin resistance (WHO Policy DST 2008)\textsuperscript{12}. Concerning the latter, three recent studies have raised concerns about phenotypic DST methods, in particular automated MGIT 960, for rifampicin using the recommended critical concentrations. Van Deun 2009 reported that BACTEC 460 and MGIT 960 missed certain strains associated with low-level rifampicin

resistance. Furthermore, using Xpert MTB/RIF and gene sequencing, Williamson 2012 identified four patients (three with clinical information available) whose TB isolates contained mutations to the \textit{rpoB} gene but appeared to be rifampicin susceptible using MGIT 960. In this study, 2/49 (4.1%) patients whose isolates did not have apparent \textit{rpoB} gene mutations, experienced treatment failure compared with 3/3 (100%) patients whose isolates did have \textit{rpoB} gene mutations and were deemed rifampicin susceptible with phenotypic methods. Recently, in a study involving retreatment patients, Van Deun and colleagues found that disputed \textit{rpoB} mutations conferring low-grade resistance were often missed by rapid phenotypic DST, particularly with the MGIT 960 system, but to a lesser extent also by conventional slow DST. The authors suggested this may be the reason for the perceived insufficient specificity of molecular DST for rifampicin. Although the study involved retreatment patients, the results appear to hold for individuals newly diagnosed with TB as well (personal communication, A. Van Deun, 6 May 2013). Specifically, the determination of the specificity of a molecular DST method based on phenotypic DST alone may underestimate the specificity of a molecular DST. In light of these findings, it is currently unclear whether and to what extent Xpert MTB/RIF might out-perform phenotypic DST methods for rifampicin resistance.

4.1.12.4. Effect of proportion rifampicin resistant samples in the study
For this analysis, we used a cutoff of 15% for the proportion of rifampicin resistant samples in the study. The pooled sensitivity was 96% (95% CI 91% - 98%) for studies with proportion rifampicin resistance > 15% (four studies), higher than the pooled sensitivity of 91% (95% CI 79% - 97%) for studies with proportion rifampicin resistance ≤ 15% (seven studies). The corresponding pooled specificities were 97% (95% CI 94% - 99%) and 99% (95% CI 98% - 99%). The difference in Xpert MTB/RIF sensitivity and specificity were not significantly different from 0.

4.1.12.5. Sensitivity analyses
For TB detection, sensitivity analyses were undertaken by limiting inclusion in the meta-analysis to: 1) studies that provided age data that met our inclusion criterion for adults; 2) studies that used consecutive sampling; 3) studies where a single specimen yielded a single Xpert MTB/RIF result for a given individual; and 4) studies that explicitly tested individuals with presumed TB. A sensitivity analysis was also performed by excluding from the meta-analysis the two large multicentre studies. These sensitivity analyses made no difference to any of the findings.

4.1.12.6. Nontuberculous mycobacteria NTM
Fourteen studies (2626 participants) provided data on a variety of nontuberculous mycobacteria (NTM) that grew from the specimens tested to look for evidence of cross-reactivity. Among these 14 studies comprising 180 NTM, Xpert MTB/RIF was positive in only one (0.6%) specimen that grew NTMs.

4.1.13. Summary of findings and GRADE evidence profiles
In adults with presumptive TB, with or without HIV infection, Xpert MTB/RIF is sensitive and specific. In comparison with smear microscopy, Xpert MTB/RIF substantially increases TB detection among culture-confirmed cases. Xpert MTB/RIF has higher sensitivity for TB detection in smear-positive patients than smear-negative patients. Nonetheless, Xpert MTB/RIF may also be valuable as an add-on test following smear microscopy in patients who have previously been found to be smear-negative. For detection of rifampicin resistance, in adults with presumptive TB or MDR-TB, Xpert MTB/RIF achieves high sensitivity and high specificity and can allow rapid initiation of MDR-TB treatment.

---

• When used as an initial test replacing smear microscopy, Xpert MTB/RIF detected 88% of TB cases with high specificity (99%).
• When used as an add-on test following smear microscopy, Xpert MTB/RIF detected 68% of TB cases with high specificity (99%).
• Xpert MTB/RIF sensitivity for smear-positive, culture-positive TB was 98%.
• Xpert MTB/RIF sensitivity for smear-negative, culture-positive TB was 68%.
• Xpert MTB/RIF detected 79% of pulmonary TB cases in people living with HIV and 86% of pulmonary TB cases in people without HIV infection. However, after adjustment for smear status, there was no evidence of a difference between the HIV-positive and HIV-negative subgroups.
• When used as an initial test replacing phenotypic culture-based drug susceptibility testing, Xpert MTB/RIF detected 95% of rifampicin-resistant TB cases with a specificity of 98%.
• Phenotypic DST is an imperfect reference standard. Hence, the determination of the specificity of a molecular DST method based on phenotypic DST alone may underestimate the specificity of a molecular DST.
• When genetic sequencing was used to resolve discordant Xpert MTB/RIF results for determination of sensitivity and specificity. The specificity (rifampicin sensitive) was 99.8% (433/434) (95% CI 98.7% to 100.0%)

4.1.14. Strengths and limitations of the evidence base

Review findings are based on comprehensive searching, strict inclusion criteria, and standardized data extraction. The strength of this review is that it allows an assessment to be made of the diagnostic accuracy of Xpert MTB/RIF for detection of TB when Xpert MTB/RIF is used as a replacement test for smear microscopy or as an add-on test following smear microscopy. In addition, the review allows for a determination of the accuracy of Xpert MTB/RIF for detection of rifampicin resistance when Xpert MTB/RIF is used as an initial test replacing conventional drug susceptibility testing.

This data set involved comprehensive searching and correspondence with experts in the field and the test manufacturer to identify additional studies, as well as repeated correspondence with study authors to obtain additional data and information that was missing in the papers. The search strategy included studies published in all languages.

The majority of studies used consecutive selection of participants and interpreted the reference standard results without knowledge of Xpert MTB/RIF results. Xpert MTB/RIF results are generated automatically, without requiring subjective interpretation. The majority of studies performed Xpert MTB/RIF in intermediate-level and peripheral-level laboratories, settings that matched the review question. In general, studies were fairly well reported, though we corresponded with almost all authors for additional data and missing information.

A major limitation for the determination of the specificity of Xpert MTB/RIF for the detection of rifampicin is the lack of a perfect reference standard. Phenotypic DST fails to detect some strains of *M. tuberculosis* with certain mutations in the *rpoB* gene which means that the specificity of Xpert MTB/RIF for rifampicin resistance detection is underestimated incorrectly inflating the number of expected false positive results.
4.1.15. GRADE Evaluation and Recommendations

GRADE evidence profiles are provided in Tables 3 to 12. The GRADE process confirmed a solid evidence base to support widespread and decentralised use of Xpert MTB/RIF for detection of TB and rifampicin resistance. The Expert Group therefore concluded that:

- Xpert MTB/RIF should be used rather than conventional microscopy, culture and DST as the initial diagnostic test in individuals presumed to have MDR-TB or HIV-associated TB (strong recommendation, high-quality evidence).

- Xpert MTB/RIF may be used as a follow-on test to microscopy in adults where MDR-TB and HIV is of lesser concern, especially in further testing of smear-negative specimens. (Conditional recommendation acknowledging resource implications, high-quality evidence).

- Xpert MTB/RIF may be used rather than conventional microscopy and culture as the initial diagnostic test in all adults presumed to have TB. (Conditional recommendation acknowledging resource implications, high-quality evidence).
Table 3: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB

**PICO Question A:** What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults, where Xpert MTB/RIF is used as a replacement test for smear microscopy?

**Participants:** Adults with presumed pulmonary TB

**Setting:** Mainly intermediate level laboratories and primary health care facilities

**Target condition:** Pulmonary TB

**Reference standard:** Solid or liquid culture

**Number of studies (number of participants):** 22 (9008)

**Pooled sensitivity:** 88% (95% CI 84, 92); **Pooled specificity:** 99% (95% CI 98, 99)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>None ⁴</td>
<td>None</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>None ⁴</td>
<td>None</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
Footnotes

1 Assumed numbers with Xpert MTB/RIF evaluated against culture.
2 Estimates of TB prevalence provided by WHO.
3 QUADAS-2 tool used to assess risk of bias. The majority of studies enrolled individuals consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and considered blinded in all studies.
4 The quality of evidence may be lowered if there are important differences in the tests studied and the expertise of those applying them in the studies compared with the settings for which the recommendations are intended. The majority of studies (15/22; 68%) evaluated Xpert MTB/RIF in settings of intended use. Although diagnostic accuracy studies may not provide direct evidence about patient-important outcomes, two studies provided information about time to treatment initiation. In Boehme 2011, for smear-negative, culture-positive TB, the median delay in beginning treatment was 56 days (IQR, 39, 81) before Xpert MTB/RIF was introduced, compared with five days (IQR, 2, 8) after Xpert MTB/RIF was introduced. In Van Rie 2013, for smear-negative, culture-positive TB patients with Xpert MTB/RIF positive results, treatment was begun on the same day compared with 13 days for patients diagnosed by other methods.
5 One unpublished study was included in the analysis. No formal assessment of publication bias was conducted using methods such as funnel plots or regression tests as such techniques have not been found to be helpful for diagnostic accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
Table 4: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in sputum smear-positive individuals

**PICO Question A.1.** What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in smear-positive individuals?

**Participants:** Adults who are smear positive and culture positive  
**Setting:** Mainly intermediate level laboratories and primary health care facilities  
**Target condition:** Pulmonary TB  
**Reference standard:** Solid or liquid culture  
**Number of studies (number of participants):** 23 (1952)  
**Pooled sensitivity:** 98% (95% CrI 97, 99); **Pooled specificity:** Not estimated

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
</tbody>
</table>
| True positives  
(individuals with TB) | Cross-sectional | None ¹ | None ² | None ³ | None | Undetected | High ²²²² | 25 (24, 25) | 49 (49, 50) | 98 (97, 99) |
| False negatives  
(individuals incorrectly classified as not having TB) | Cross-sectional | None ¹ | None ² | None | None | Undetected ⁴ | High ²²²² | 1 (0, 1) | 1 (1, 2) | 2 (1, 3) |
| False positives  
(individuals incorrectly classified as having TB) | *** | *** | *** | *** | *** | *** | *** | *** | *** | *** |
| True negatives  
(individuals without TB) | *** | *** | *** | *** | *** | *** | *** | *** | *** | *** |
*** Xpert MTB/RIF pooled specificity was not estimated in these studies because almost all participants were considered to be true TB positive.

Footnotes
1. QUADAS-2 tool used to assess risk of bias. The majority of studies enrolled individuals consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and was considered blinded in all studies.
2. The majority of studies (16/23; 70%) evaluated Xpert MTB/RIF in settings of intended use. Although diagnostic accuracy studies may not provide direct evidence about patient-important outcomes, the quality of evidence was not downgraded.
3. Sensitivity estimates were highly consistent.
4. One unpublished study was included in the analysis. No formal assessment of publication bias was conducted using methods such as funnel plots or regression tests as such techniques have not been found to be helpful for diagnostic accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
### Table 5: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in sputum smear-negative individuals

**PICO Question A.2.** What is the diagnostic accuracy of Xpert MTB/RIF for detection of (culture-confirmed) pulmonary TB in smear-negative individuals?

- **Participants:** Adults who are smear-negative with presumed pulmonary TB
- **Setting:** Mainly intermediate level laboratories and primary health care facilities
- **Target condition:** Pulmonary TB
- **Reference standard:** Solid or liquid culture
- **Number of studies (number of participants):** 23 (7151)
- **Pooled sensitivity:** 68% (95% CrI 61, 74); **Pooled specificity:** 99% (95% CrI 98, 99)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 smear-negative individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td><strong>True positives</strong> (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>None ²</td>
<td>Serious (-1)³</td>
</tr>
<tr>
<td><strong>False negatives</strong> (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>None ²</td>
<td>Serious (-1)³</td>
</tr>
<tr>
<td><strong>False positives</strong> (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>None ²</td>
<td>None</td>
</tr>
<tr>
<td><strong>True negatives</strong> (individuals without TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>None ²</td>
<td>None</td>
</tr>
</tbody>
</table>
Footnotes
1 QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled individuals consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and was considered blinded in all studies.
2 The majority of studies (16/23; 70%) evaluated Xpert MTB/RIF in settings of intended use. Although diagnostic accuracy studies may not provide direct evidence about patient-important outcomes, the quality of evidence was not downgraded.
3 There was some variability in sensitivity estimates across studies. This heterogeneity could not be explained by study quality or by removing the study by Lawn from the analysis. This study, which showed the lowest sensitivity, evaluated the use of Xpert MTB/RIF to screen HIV-infected patients with advanced immunodeficiency, regardless of symptoms, enrolling in antiretroviral therapy services. The observed sensitivity may have varied between studies according to characteristics associated with the participants, but there was insufficient data to investigate this possibility further. Quality of evidence was downgraded by one point.
4 At a pre-test probability of 10%, the CIs for TP and FN were relatively narrow.
5 One unpublished study was included in the analysis. No formal assessment of publication bias was conducted using methods such as funnel plots or regression tests as such techniques have not been found to be helpful for diagnostic accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
Table 6: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in persons living with HIV

**PICO Question A.3.** What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in people living with HIV?

**Participants:** Adults living with HIV with presumed pulmonary TB  
**Setting:** Mainly intermediate level laboratories and primary health care facilities  
**Target condition:** Pulmonary TB  
**Reference standard:** Solid or liquid culture  
**Number of studies (number of participants):** 7 (1789)  
**Pooled sensitivity:** 79% (70, 86); **Pooled specificity:** 98% (96, 99)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>None ²</td>
<td>Serious (-1) ³</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>None ²</td>
<td>Serious (-1) ³</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>None ²</td>
<td>None</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>None ²</td>
<td>None</td>
</tr>
</tbody>
</table>
Footnotes

1 QUADAS-2 tool was used to assess risk of bias. All studies enrolled individuals consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and was considered blinded in all studies.

2 The quality of evidence may be lowered if there are important differences in tests studied and the expertise of those applying them in the studies compared to the settings for which the recommendations are intended. The majority studies (6/7; 86%) evaluated Xpert MTB/RIF in settings of intended use. Although diagnostic accuracy studies may not provide direct evidence about patient-important outcomes, the quality of evidence was not downgraded.

3 There was some variability in sensitivity estimates across studies but this heterogeneity could not be explained by study quality. The observed sensitivity may have varied between studies according to characteristics associated with the participants, but there was insufficient data to investigate this possibility further. Quality of evidence was downgraded by one point.

4 At a pre-test probability of 10%, the CIs for TP and FN were relatively narrow.

5 No formal assessment of publication bias was conducted using methods such as funnel plots or regression tests as such techniques have not been found to be helpful for diagnostic accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
Table 7: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB in persons without HIV infection

PICO Question A.4 What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults without HIV infection?
Number of studies (number of participants): 7 (1470)
Participants: Adults who are HIV-negative with presumed pulmonary TB
Setting: Mainly intermediate level laboratories and primary health care facilities
Target condition: Pulmonary TB
Reference standard: Solid or liquid culture
Pooled sensitivity: 86% (76, 92); Pooled specificity: 99% (98, 100)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>Serious (-1)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>Serious (-1)</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
Footnotes

1 QUADAS-2 tool was used to assess risk of bias. All studies enrolled individuals consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and was considered blinded in all studies.

2 The quality of evidence may be lowered if there are important differences in tests studied and the expertise of those applying them in the studies compared to the settings for which the recommendations are intended. The majority studies (6/7; 86%) evaluated Xpert MTB/RIF in settings of intended use. Although diagnostic accuracy studies may not provide direct evidence about patient-important outcomes, the quality of evidence was not downgraded.

3 There was some variability in sensitivity estimates across studies but this heterogeneity could not be explained by study quality. The observed sensitivity may have varied between studies according to characteristics associated with the participants, but there was insufficient data to investigate this possibility further. Quality of evidence was downgraded by one point.

4 At a pre-test probability of 10%, the CIs for TP and FN were relatively narrow.

5 No formal assessment of publication bias was conducted using methods such as funnel plots or regression tests as such techniques have not been found to be helpful for diagnostic accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
Table 8: GRADE evidence profile: The incremental yield of Xpert MTB/RIF compared with microscopy in patients with culture-confirmed TB

PICO Question A.5. What is the incremental yield of Xpert MTB/RIF compared with microscopy in patients with culture-confirmed TB?  
Number of studies (number of participants): 21 (8880)  
Participants: Adults with culture-confirmed TB  
Setting: Mainly intermediate level laboratories and primary health care facilities  
Target condition: Pulmonary TB  
Reference standard: Solid or liquid culture  
Smear microscopy pooled sensitivity: 65% (95% CrI 57% to 72%); Xpert MTB/RIF pooled sensitivity: 88% (84, 92)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of results per 1000 culture-positive individuals tested (95% CrI)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 25 per 1000</td>
<td>Prevalence 50 per 1000</td>
</tr>
<tr>
<td></td>
<td>Smear Microscopy</td>
<td>Xpert MTB/RIF</td>
</tr>
<tr>
<td>True positives (patients with TB)</td>
<td>16 (14, 18)</td>
<td>22 (21, 23)</td>
</tr>
<tr>
<td>TP absolute difference</td>
<td>6 more</td>
<td>11 more</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having TB)</td>
<td>9 (7, 11)</td>
<td>3 (2, 4)</td>
</tr>
<tr>
<td>FN absolute difference</td>
<td>6 less</td>
<td>12 less</td>
</tr>
</tbody>
</table>
Footnotes

1 The sensitivity results were taken from bivariate analyses (including both sensitivity and specificity) to obtain the values for true positives and false negatives.

2 The GRADE framework was used for assessing quality of evidence. For microscopy, the sensitivity estimates for individual studies were variable (29% to 83%) and the pooled sensitivity estimate was imprecise. The main reason for heterogeneity/imprecision in the sensitivity estimate was considered to be the variability in smear-positive status across studies. Several additional factors may have contributed to this heterogeneity, including type of microscopy, specimen processing and HIV status. The quality of evidence was not downgraded.
Table 9: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for adult pulmonary TB, as an add-on test following a negative sputum smear microscopy

**PICO Question B:** What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result?

**Participants:** Adults who are smear negative with presumed pulmonary TB

**Setting:** Mainly intermediate level laboratories and primary health care facilities

**Target condition:** Pulmonary TB

**Reference standard:** Solid or liquid culture

**Number of studies (number of participants):** 23 (7151)

**Pooled sensitivity:** 68% (95% CI 61, 74); **Pooled specificity:** 99% (95% CI 98, 99)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 smear-negative individuals tested (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None $^1$</td>
<td>None $^2$</td>
<td>Serious (-1) $^3$</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None $^1$</td>
<td>None $^2$</td>
<td>Serious (-1) $^3$</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None $^1$</td>
<td>None $^2$</td>
<td>None $^3$</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>Cross-sectional</td>
<td>None $^1$</td>
<td>None $^2$</td>
<td>None $^3$</td>
</tr>
</tbody>
</table>
Footnotes

1 QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled individuals consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and was considered blinded in all studies.

2 The majority of studies (16/23; 70%) evaluated Xpert MTB/RIF in settings of intended use. Although diagnostic accuracy studies may not provide direct evidence about patient-important outcomes, the quality of evidence was not downgraded.

3 There was some variability in sensitivity estimates across studies but this heterogeneity could not be explained by study quality or by removing the study by Lawn from the analysis. This study, which showed the lowest sensitivity, evaluated the use of Xpert MTB/RIF to screen HIV-infected patients with advanced immunodeficiency, regardless of symptoms, enrolling in antiretroviral therapy services. The observed sensitivity may have varied between studies according to characteristics associated with the participants, but there was insufficient data to investigate this possibility further. Quality of evidence was downgraded by one point.

4 At a pre-test probability of 10%, the CIs for TP and FN were relatively narrow.

5 One unpublished study was included in the analysis. No formal assessment of publication bias was conducted using methods such as funnel plots or regression tests as such techniques have not been found to be helpful for diagnostic accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
**Table 10: Xpert MTB/RIF sensitivity in smear-negative (culture-confirmed) individuals by HIV status**

**PICO Question B.1** What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result, stratified by HIV status?

**Participants:** Adults with smear-negative (culture-confirmed) pulmonary TB

**Setting:** One intermediate level laboratory and one primary health care clinic

**Reference standard:** Phenotypic culture using solid or liquid media

**Number of studies (number of participants):** 2 (91)

<table>
<thead>
<tr>
<th>Study</th>
<th>HIV-positive (n=33)</th>
<th>HIV-negative (n=58)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (95% CI)</td>
<td>Sensitivity (95% CI)</td>
</tr>
<tr>
<td>Theron 2011</td>
<td>48% (27, 69)</td>
<td>45% (25, 67)</td>
</tr>
<tr>
<td>Van Rie 2013</td>
<td>60% (27, 86)</td>
<td>67% (13, 98)</td>
</tr>
</tbody>
</table>
**Table 11: GRADE evidence profile: Additional yield of Xpert MTB/RIF over microscopy in smear-negative TB**

**PICO Question B.2** What is the additional yield of Xpert MTB/RIF over microscopy in smear-negative TB?

**Number of studies (number of participants):** 23 (7151)

**Participants:** Adults who are smear negative and culture positive

**Setting:** Mainly intermediate level laboratories and primary health care facilities

**Target condition:** Pulmonary TB

**Reference standard:** Solid or liquid culture

**Smear microscopy pooled sensitivity:** 0%; **Xpert MTB/RIF pooled sensitivity:** 68% (95% CrI 61, 74)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of results per 1000 smear-negative (culture-positive) individuals tested (95% CrI)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 25 per 1000</td>
<td>Prevalence 50 per 1000</td>
</tr>
<tr>
<td><strong>True positives</strong> (patients with TB)</td>
<td>Smear Microscopy</td>
<td>Xpert MTB/RIF</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>17 (15, 19)</td>
</tr>
<tr>
<td><strong>TP absolute difference</strong></td>
<td>17 more</td>
<td>34 more</td>
</tr>
<tr>
<td><strong>False negatives</strong> (patients incorrectly classified as not having TB)</td>
<td>25 (7, 10)</td>
<td>50 (13, 20)</td>
</tr>
<tr>
<td><strong>FN absolute difference</strong></td>
<td>17 less</td>
<td>34 less</td>
</tr>
</tbody>
</table>

Footnotes

1 Sensitivity results were taken from the bivariate analyses (including both sensitivity and specificity) to obtain the values for true positives and false negatives.

2 The GRADE framework was used for assessing the quality of evidence. The quality of evidence was downgraded one point for inconsistency/imprecision.
**Table 12**: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing

**PICO Question C.** What is the diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing?

**Participants:** Adults with confirmed TB

**Setting:** Mainly intermediate level laboratories and primary health care facilities

**Target condition:** Rifampicin resistance

**Reference standard:** Phenotypic culture-based drug susceptibility testing

**Number of studies (number of participants) pooled sensitivity:** 17 (555)

**Number of studies (number of participants) pooled specificity:** 24 (2414)

**Pooled sensitivity:** 95% (90, 97); **Pooled specificity:** 98% (97, 99) [Sensitivity and specificity estimates were determined separately by univariate analyses]

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (individuals with rifampicin resistance)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Undetected</td>
<td>High</td>
<td>84 (45, 49)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as rifampicin susceptible)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Undetected</td>
<td>High</td>
<td>3 (2, 5)</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having rifampicin)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Undetected</td>
<td>High</td>
<td>19 (10, 29)</td>
</tr>
<tr>
<td>True negatives (individuals who are rifampicin susceptible)</td>
<td>Cross-sectional</td>
<td>None ²</td>
<td>None ³</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>High ⁶</td>
<td>931 (922, 941)</td>
<td>833 (825, 842)</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------------</td>
<td>-------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
</tbody>
</table>
4.2. Xpert MTB/RIF for the diagnosis of extrapulmonary TB and rifampicin resistance in adults and children

4.2.1. Study characteristics

The 22 studies including 5922 samples had a TB prevalence (based on culture) that ranged from 7% to 81%. Only one study (unpublished) was written in a language other than English (Portuguese). Thirteen studies (59%) were conducted in low/middle-income countries. All studies were performed in tertiary care centers or reference laboratories. For 19 studies HIV status of study participants was known, and only five studies did not include any HIV-positive patients. Fourteen studies included HIV-positive patients, with percentages ranging from 1% to 87% of the study population. Of the three studies with unknown HIV status, two were done in low HIV incidence settings (Germany and France) and one in a high HIV incidence setting (South Africa). Two studies (studies by Bates and by Walters) included children only, while nine studies included no children at all. In the remaining 11 studies the percentages of children in the study population ranged from 2% to 34%. The median number of samples per study was 145 (IQR 67-342). Three published and four unpublished studies included only one type of sample (e.g. pleural fluid only). The remainder of the studies included different sample types in varying percentages. Twelve studies reported only one sample per patient, while the other studies had multiple samples per patient or did not report the number of samples per patient. Six studies used archived samples (in frozen condition), while 15 used fresh samples, and one study used both fresh and frozen samples.

With respect to specimen processing, the studies varied widely (Table 13). Only four studies used the protocol recommended by the manufacturer for unprocessed respiratory samples. Thirteen studies (59%) reported a mechanical homogenization step for non-liquid samples. Twelve studies (55%) reported using N-acetyl-L-cysteine and sodium hydroxide (NALC-NaOH) solution for specimen digestion and decontamination and, one study used NaOH only. Some studies did not consistently decontaminate all specimens but only when bacterial contamination was identified. Most of the studies that included a mechanical homogenization step also performed a decontamination procedure. Fourteen studies reported a concentration step and ten had a re-suspension step with varying volumes that went into the two steps. We extracted data on sample processing on a study level although some steps (e.g. homogenization) might apply to only certain sample types.

The ratio of the sample reagent volume to sample volume also varied. Seven studies used a sample reagent to sample ratio of 3:1 while 15 studies used 2:1. Of the 12 studies that used digestion and decontamination and a concentration step, six studies used a sample reagent to sample ratio of 2:1. The manufacturer recommends a reagent to sample ratio of 3:1 for samples processed by the Kent and Kubica protocol (26) and 2:1 for unprocessed sputum.

Annex 5 shows the PRISMA diagram with the flow of the studies. From the literature search, 194 citations and reviewed 31 full-text articles were identified. Twenty two studies were identified that met the eligibility criteria for inclusion in the review. Annex 5 lists the included and excluded studies along with the reason for exclusion.
Table 13: Sample types and processing methods for included studies

<table>
<thead>
<tr>
<th>#</th>
<th>Study (Year)</th>
<th>Setting</th>
<th>Specimen Condition</th>
<th>Homogenization</th>
<th>Digestion/Decomposition</th>
<th>Concentration (Volume)</th>
<th>Resuspension</th>
<th>Sample reagent to sample ratio</th>
<th>Comments on sample Processing</th>
<th>Input volume</th>
<th>Total No. samples</th>
<th>Specimen Type and No. Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Al-Ateah (2012)</td>
<td>Tertiary Care Center</td>
<td>Fresh</td>
<td>Mechanical</td>
<td>NALC/NaOH</td>
<td>Y (2.5ml)</td>
<td>Y</td>
<td>2:1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Armand (2011)</td>
<td>Tertiary Care Center</td>
<td>Frozen</td>
<td>Mechanical</td>
<td>NALC/NaOH</td>
<td>Y (variable)</td>
<td>No</td>
<td>3:1</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Bates (2012)</td>
<td>Tertiary Care Center</td>
<td>Fresh</td>
<td>Mechanical</td>
<td>NALC/NaOH</td>
<td>Y (variable)</td>
<td>Y (2ml)</td>
<td>3:1</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Causse (2011)</td>
<td>Reference Laboratory</td>
<td>Fresh</td>
<td>N</td>
<td>NALC/NaOH</td>
<td>Y (variable)</td>
<td>N</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Friedrich (2011)</td>
<td>Tertiary Care Center</td>
<td>Fresh</td>
<td>N</td>
<td>N</td>
<td>NALC/NaOH</td>
<td>Y (50ml)</td>
<td>Y (2ml)</td>
<td>2:1</td>
<td>All samples were concentrated except for CSF</td>
<td>&gt;2</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>Hanif (2011)</td>
<td>Reference Laboratory</td>
<td>Fresh</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>2:1</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Hillemann (2011)</td>
<td>Reference Laboratory</td>
<td>Fresh</td>
<td>Mechanical (for tissue)</td>
<td>NALC/NaOH</td>
<td>Y</td>
<td>1.5ml</td>
<td>3:1</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Lighthelm (2011)</td>
<td>Tertiary Care Center</td>
<td>Fresh</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>2:1</td>
<td>If sample with &lt;0.7ml, PBS added to increase volume to 0.7ml</td>
<td></td>
<td>2</td>
<td>48</td>
</tr>
<tr>
<td>9</td>
<td>Malbruny (2011)</td>
<td>Tertiary Care Center</td>
<td>Fresh &amp; Frozen</td>
<td>Mechanical</td>
<td>NALC/NaOH</td>
<td>Y (variable)</td>
<td>N</td>
<td>3:1</td>
<td>All samples were concentrated except for CSF</td>
<td>0.5-1</td>
<td>124</td>
<td>Pleural Fluid 12</td>
</tr>
<tr>
<td>10</td>
<td>Moure (2012)</td>
<td>Tertiary Care Center</td>
<td>Frozen</td>
<td>Mechanical</td>
<td>NALC/NaOH</td>
<td>Y (variable)</td>
<td>Y (2ml)</td>
<td>2:1</td>
<td>3ml went into cartridge: 1ml sediment and 2ml SR</td>
<td>3</td>
<td>149</td>
<td>Pleural Fluid 34</td>
</tr>
<tr>
<td>11</td>
<td>Safianowska (2012)</td>
<td>Tertiary Care Center</td>
<td>Fresh</td>
<td>Mechanical</td>
<td>NALC/NaOH</td>
<td>Y (see comment)</td>
<td>Y (2.5ml)</td>
<td>3:1</td>
<td>Centrifugation, then additional washing step with</td>
<td></td>
<td>2</td>
<td>68</td>
</tr>
<tr>
<td>No.</td>
<td>Author</td>
<td>Tertiary Care Center</td>
<td>Sample Preservation</td>
<td>Nalcc/Naoh</td>
<td>Volume (ml)</td>
<td>SR:S</td>
<td>Remarks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>----------------------</td>
<td>--------------------</td>
<td>------------</td>
<td>-------------</td>
<td>------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Tortoli</td>
<td>Frozen</td>
<td>Mechanical</td>
<td>Y (variable)</td>
<td>Y (2ml)</td>
<td>2:1</td>
<td>Decontamination done only for non sterile samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Vadwai</td>
<td>Fresh</td>
<td>Mechanical (for tissue)</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>2:1</td>
<td>SR:S=2:1 except for CSF which was usually &lt;1ml and was raised to 2 ml with SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Walters</td>
<td>Fresh</td>
<td>Mechanical</td>
<td>Y (see comment)</td>
<td>Y (1.5ml)</td>
<td>2:1</td>
<td>PBS added up to volume of 40ml, then centrifuged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Zeka</td>
<td>Frozen</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>3:1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Caws</td>
<td>Fresh</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y (5-8ml)</td>
<td>Y (0.7ml)</td>
<td>2:1</td>
<td>200ul deposit was taken and 500ul PBS added</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Kohli</td>
<td>Fresh</td>
<td>Mechanical (for tissue)</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>2:1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Meldau</td>
<td>Fresh</td>
<td>N</td>
<td>N</td>
<td>NaOH</td>
<td>N</td>
<td>2:1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Oliveira</td>
<td>Frozen</td>
<td>NaOH</td>
<td>N</td>
<td>N</td>
<td>2:1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Patel</td>
<td>Frozen</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>2:1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Scott</td>
<td>Fresh</td>
<td>Mechanical (for tissue)</td>
<td>NALC/NaOH</td>
<td>Y (5ml)</td>
<td>3:1</td>
<td>Decontaminated only if contaminated. If &gt;0.5ml sample, used 2:1 SR:S ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Singh</td>
<td>Fresh</td>
<td>Mechanical (for tissue)</td>
<td>NALC/NaOH</td>
<td>Y</td>
<td>2:1</td>
<td>NALC/NaOH for all the samples except CSF; 2-3 ml deposit was used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2.2. Study Quality

The methodological quality for each individual study, separately for culture as the reference standard and the composite reference standard (CRS)\(^{16}\). The quality assessment was the same for both culture and CRS except for the assessment of ‘Flow and timing’ and the blinding of the reference standard. The overall methodological quality of all included studies is summarized in Figure 6 for the culture reference standard and in Figure 7 for the CRS.

**Figure 6: Risk of bias and applicability concerns summary for studies using a culture reference standard for TB detection.**

The graph reflects the risk of bias and applicability as judged by review authors about each QUADAS-2 domain presented as percentages across the 22 included studies.

**Figure 7: Risk of bias and applicability concerns summary for studies using an author-defined composite reference standard for TB detection.**

The graph reflects the risk of bias and applicability as judged by review authors about each QUADAS-2 domain presented as percentages across the 22 included studies.

In the patient selection domain, six studies were judged as high risk of bias because they were either case-control studies or used convenience sampling of participants for enrolment. The majority of studies used prospective data collection (17 out of 22; 77%). Concerns for applicability in respect to the intended site of use for Xpert MTB/RIF (as described for pulmonary TB, i.e. district or sub-district health care settings) were

\(^{16}\) Most analyses were performed using two reference standards: culture (the current reference standard) and a combined clinical and laboratory reference standard chosen by the study authors (given technical limitations of culture diagnosis).
judged as ‘high’ if Xpert MTB/RIF was evaluated in reference laboratories and ‘unclear’ if Xpert MTB/RIF was evaluated in a hospital laboratory.

The index test was considered blinded with respect to results of the reference standard as the interpretation of the Xpert MTB/RIF result does not require a human judgment. Similarly, the threshold for positivity for TB detection is fixed by the manufacturer and thus by definition pre-specified. In the index test domain, all studies were considered to have low risk of bias.

With respect to the domain ‘applicability’ variations in the execution of the test were considered to possibly affect estimates of the diagnostic accuracy of Xpert MTB/RIF to a different degree. There was low concern if samples were unprocessed and the test was done as per recommendation of the manufacturer for sputum samples (14%, 3/22). There were high concerns that the test performance could be altered through adding a mechanical homogenization step (13 out of 22 studies) as it was unclear whether the homogenization would be sufficient and what quantity of sample particles would ultimately be included in the sample input volume. It was also considered conceivable the particles could clog the valves and result in a higher rate of indeterminate test results. In six of the 22 studies (27%) the processing protocols were unclear.

The reference standards were considered to introduce bias due to misclassification of participants and hence all studies in this domain were rated as unclear. Blinding of the reference standard was only considered to be of relevance for culture if species identification was not done (either via the Capilia test, NAAT or sequencing). For the CRS blinding was more important, particularly if the CRS included smear without species identification or if a clinical evaluation was included. The latter applied to one study but otherwise studies showed low risk of bias.

In the domain of flow and timing, an interval of several days between the index test and reference standard was not considered problematic for the diagnosis of TB in presumptive cases (i.e. not on treatment for TB) as TB is a chronic disease and test results are unlikely to change within a few days and therefore misclassification of disease status is unlikely. In the analysis based on a culture reference standard, all patients across studies were included into the analysis; therefore partial verification bias was not considered a problem. For the CRS, three studies indicated that some patients did not receive culture as part of the CRS. These studies were rated with ‘high concern’ for risk of bias.

4.2.3. Xpert MTB/RIF for TB detection

4.2.3.1. EPTB, all sample types

The studies reviewed were very diverse with respect to both the different sample types tested and their relative percentages in each study. The heterogeneity in the performance characteristics, primarily in sensitivity, across different sample types was substantial. Therefore combining these studies to obtain an overall estimate of the accuracy of Xpert MTB/RIF in EPTB would not be meaningful.

However, when assessing all studies by smear status, the heterogeneity was restricted primarily to the smear-negative samples. For all smear-positive samples across studies (390 samples), heterogeneity was limited (Figure 8). A univariate analysis was done for sensitivity only (97.6%; 95% CI: 95.2-99.9%), as data were too limited for the estimation of specificity.
Figure 8: Forest plot of Xpert MTB/RIF sensitivity for TB detection in smear-positive subgroup.

The squares represent the sensitivity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Xpert MTB/RIF specificity was not estimated because available data was limited. Unpublished studies are labelled with year 2013 or blackened.

An analysis on predefined subgroups of sample types (i.e. pleural, lymph node aspirate or tissue, CSF, gastric fluid and tissue other than lymph node) was undertaken to account for the heterogeneity between studies. Data for smear status of samples were not available for the individual sample types. Therefore samples included in the subgroups were either smear-positive, smear-negative or of unknown smear status.

4.2.3.2. Detection of lymph node TB (biopsy or fine-needle aspirate)

14 studies were identified that tested Xpert MTB/RIF on lymph node biopsy or fine-needle aspirate against a culture reference standard (Figure 9). A meta-analysis for each sample type was performed if at least four studies were available with at least 10 samples in each study. From the 11 studies with more than 10 samples; 849 samples, estimates for sensitivity ranged from 50% to 100%. Pooled sensitivity across studies was 84.9% (95% CI 72.1%-92.4%) and specificity was 92.5% (95% CI 80.3%-97.4%). Only two studies reported any indeterminate results on Xpert MTB/RIF testing: Armand 10% (2/20) and one unpublished study 1.6% (3/193).
Figure 9: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in lymph node samples (tissue or aspirate) with culture reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are labelled with year 2013 or blackened.

One unpublished study had a much lower Xpert MTB/RIF specificity than the other studies (38%, compared to 71-100% in the other studies). If the subset of published studies was analysed separately, the pooled specificity was improved slightly to 94.4% (95% CI: 88.2-97.4%) with the lower bound of the 95% CI shifted upward by 8%. The sensitivity decreased slightly to 80.8% (95% CI: 67.9-89.4%). In the subset of studies that performed consecutive sampling of study participants, the pooled sensitivity was slightly increased 89.4% (95% CI: 74.1-96.1%) and specificity was decreased at 86.9% (95% CI: 67.5-95.5%) but the precision of these estimates also decreased. The overall estimate did not differ significantly if the two case-control studies (Armand and Moure) were removed (Table 14).

Table 14: Sensitivity analysis by sample type, TB detection

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Pooled Sensitivity (95% CI)</th>
<th>Pooled Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LYMPH NODE (overall)</strong></td>
<td>85% (72, 92)</td>
<td>93% (80, 97)</td>
</tr>
<tr>
<td>Published studies</td>
<td>81% (68, 89)</td>
<td>94% (88, 97)</td>
</tr>
<tr>
<td>Studies with consecutive sampling</td>
<td>89% (74, 96)</td>
<td>87% (68, 96)</td>
</tr>
<tr>
<td>Excluding case control studies</td>
<td>89% (77, 95)</td>
<td>90% (76, 97)</td>
</tr>
<tr>
<td><strong>PLEURAL FLUID (overall)</strong></td>
<td>44% (25, 65)</td>
<td>98% (95, 99)</td>
</tr>
<tr>
<td>Published studies</td>
<td>48% (22, 75)</td>
<td>99% (98, 100)</td>
</tr>
<tr>
<td>Studies with consecutive sampling</td>
<td>54% (39, 68)</td>
<td>98% (93, 99)</td>
</tr>
<tr>
<td>Excluding case control studies</td>
<td>46% (23, 72)</td>
<td>98% (95, 99)</td>
</tr>
<tr>
<td><strong>CEREBROSPINAL FLUID (overall)</strong></td>
<td>80% (62, 90)</td>
<td>99% (96, 100)</td>
</tr>
<tr>
<td>Published studies</td>
<td>77% (48, 100)</td>
<td>99% (97, 100)</td>
</tr>
<tr>
<td>Studies with consecutive sampling</td>
<td>70% (47, 94)</td>
<td>99% (98, 100)</td>
</tr>
<tr>
<td>Excluding case control studies</td>
<td>74% (56, 93)</td>
<td>99% (98, 100)</td>
</tr>
<tr>
<td><strong>GASTRIC FLUID (overall)</strong></td>
<td>84% (66, 93)</td>
<td>98% (92, 100)</td>
</tr>
<tr>
<td>Published studies</td>
<td>84% (73, 94)</td>
<td>99% (99, 100)</td>
</tr>
<tr>
<td>Studies with consecutive sampling</td>
<td>89% (72, 100)</td>
<td>91% (81, 100)</td>
</tr>
<tr>
<td>Excluding case control studies</td>
<td>89% (77, 100)</td>
<td>96% (92, 100)</td>
</tr>
<tr>
<td><strong>Tissue (overall)</strong></td>
<td>81% (68, 90)</td>
<td>98% (87, 100)</td>
</tr>
</tbody>
</table>
Published studies $80 \% (66, 89) \quad 99 \% (89, 100)$
Studies with consecutive sampling $80 \% (69, 88) \quad 98 \% (77, 100)$
Excluding case control studies $84 \% (76, 90) \quad 98 \% (86, 100)$

Five studies (one unpublished) assessed Xpert MTB/RIF on lymph node samples against an author-defined composite reference standard (CRS) (Figure 10). The CRS might have included a nucleic acid amplification test (NAAT other than Xpert MTB/RIF), histology, smear, culture, biochemical testing results, presenting signs or a response to treatment with anti-TB therapy. The pooled sensitivity was estimated to be 83.7\% (95% CI 73.8%-90.3) and the pooled specificity to be 99.2\% (95% CI 88.4%-100%).

Studies that used fresh samples showed a slightly higher sensitivity and a lower specificity than those that used frozen samples; however the precision of these estimates was low as data were limited. Only nine studies had information on HIV prevalence and only two studies included more than 10% HIV positive patients. Accuracy estimates for these studies did not differ substantially from those with less HIV patients (Figure 10). Given the limited amount of data in the group with HIV prevalence >10%, a summary estimate was not determined.

Figure 10: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in lymph node samples (tissue or aspirate) with composite reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are labelled with year 2013 or blackened.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligthelm 2011</td>
<td>17</td>
<td>1</td>
<td>4</td>
<td>16</td>
<td>0.81 [0.58, 0.95]</td>
<td>0.88 [0.47, 1.00]</td>
</tr>
<tr>
<td>Tortoli 2012</td>
<td>28</td>
<td>0</td>
<td>5</td>
<td>85</td>
<td>0.85 [0.68, 0.95]</td>
<td>1.00 [0.96, 1.00]</td>
</tr>
<tr>
<td>Vadwal 2011</td>
<td>49</td>
<td>0</td>
<td>17</td>
<td>122</td>
<td>0.74 [0.62, 0.84]</td>
<td>1.00 [0.97, 1.00]</td>
</tr>
<tr>
<td>Zeka 2011</td>
<td>13</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>0.76 [0.50, 0.93]</td>
<td>1.00 [0.66, 1.00]</td>
</tr>
</tbody>
</table>

4.2.3.6. Detection of pleural TB

Seventeen studies (1385 samples, 217 culture-positive) provided data to estimate Xpert MTB/RIF sensitivity and specificity in pleural fluid. Results on accuracy of Xpert MTB/RIF in pleural biopsy were integrated into the assessment of Xpert MTB/RIF on tissue biopsies of all kinds other than lymph node (see section 4.2.3.6 below).

Xpert MTB/RIF sensitivity on pleural fluid varied from 0% to 100% between studies (Figure 11). The outliers at the lower and upper end of the range were studies with few culture-confirmed TB cases. For the meta-analysis studies that did not contribute to either sensitivity or specificity and those that included less than 10 pleural specimens were excluded. The pooled sensitivity was low, 43.7% with a wide 95% CI (24.8-64.7%) and the pooled specificity was high, 98.1% (95% CI: 95.3-99.2%).
Seven studies (4 published and 3 unpublished studies) involving 698 samples (188 culture positive) evaluated Xpert MTB/RIF in pleural fluid using a composite reference standard (CRS). The CRS might have included a nucleic acid amplification test (NAAT other than Xpert MTB/RIF), histology, smear, culture, biochemical testing results, presenting signs or a response to treatment with anti-TB therapy. Compared with studies that used culture as the reference standard, the CRS subgroup yielded an even lower pooled sensitivity (17.0%, 95% CI 7.5%-34.2%) with a high specificity (99.9%, 95% CI 93.7%-100.0%).

Sensitivity was increased in studies with a low rate of HIV-coinfection (48% compared to 31% in studies with more than 10% HIV, however confidence intervals were wide and overlapping. There was also no difference in the results of studies that used a concentration step. In assessing the condition of the specimen, an improved sensitivity was observed for fresh samples (50%, 95% CI 36%-64%) compared to frozen samples (26%, 95% CI 14%-40%), while specificity was lower (95%, 95% CI 93-98% versus 99%, 95% CI 97%-100% for frozen).

4.2.3.4. Detection of TB in cerebrospinal fluid

In total, 709 CSF samples were tested with Xpert MTB/RIF against a culture reference standard across 16 studies (13 with more than 10 samples and 10 that provided information on both sensitivity and specificity). Only 117 culture-confirmed cases of TB were found. Estimates for sensitivity varied widely and ranged from 51% to 100%, with one study comprising 19 samples (3 false negatives) being an outlier at 0% (Figure 12). Pooled sensitivity across studies was 79.5% (95% CI: 62.0-90.2%) and specificity was 98.6% (95% CI: 95.8-99.6%) suggesting good performance of Xpert MTB/RIF for detection of TB in CSF when tested against a culture reference standard.
Figure 12: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in cerebrospinal fluid with culture reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are labelled with year 2013 or blackened.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Ateah 2012</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>1.00 [0.77, 1.00]</td>
<td></td>
</tr>
<tr>
<td>Armand 2011</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>1.00 [0.46, 1.00]</td>
<td></td>
</tr>
<tr>
<td>Causse 2011</td>
<td>5</td>
<td>0</td>
<td>14</td>
<td></td>
<td>0.83 [0.36, 1.00]</td>
<td>1.00 [0.92, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caws 2013</td>
<td>43</td>
<td>0</td>
<td>5</td>
<td>93</td>
<td>0.90 [0.77, 0.97]</td>
<td>1.00 [0.96, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haif 2011</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td></td>
<td>1.00 [0.03, 1.00]</td>
<td>1.00 [0.40, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillemann 2011</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td>1.00 [0.82, 1.00]</td>
<td></td>
</tr>
<tr>
<td>Malbrunru 2011</td>
<td>2</td>
<td>1</td>
<td>37</td>
<td></td>
<td>1.00 [0.16, 1.00]</td>
<td>0.97 [0.86, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moure 2012</td>
<td>2</td>
<td>0</td>
<td>12</td>
<td></td>
<td>1.00 [0.16, 1.00]</td>
<td>1.00 [0.74, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patel 2013</td>
<td>18</td>
<td>7</td>
<td>17</td>
<td>107</td>
<td>0.51 [0.34, 0.69]</td>
<td>0.94 [0.88, 0.97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salfianowska 2012</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>1.00 [0.54, 1.00]</td>
<td></td>
</tr>
<tr>
<td>Tortoli 2012</td>
<td>11</td>
<td>2</td>
<td>118</td>
<td></td>
<td>0.85 [0.55, 0.98]</td>
<td>0.98 [0.94, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vadwai 2011</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>16</td>
<td>0.00 [0.00, 0.71]</td>
<td>1.00 [0.79, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zeka 2011</td>
<td>3</td>
<td>0</td>
<td>28</td>
<td></td>
<td>1.00 [0.29, 1.00]</td>
<td>1.00 [0.85, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the subset of published studies (8 studies; only a univariate analysis was feasible), sensitivity and specificity were largely unchanged from the overall estimate (Table 14). Sensitivity in the studies that used consecutive sampling (rather than convenience; 10 studies) was decreased at 70.1% (95% CI 46.6%-93.7%) while specificity remained largely the same. All estimates from the above sensitivity analyses were accompanied by wide and overlapping confidence intervals.

Only six studies (3 unpublished) assessed Xpert MTB/RIF on CSF samples against an author-defined CRS with sensitivity estimates ranging from 20% to 86% (Figure 13). The pooled sensitivity was estimated to be 55.5% (95% CI 44.2%-66.3) and specificity to be 98.8% (95% CI 94.5%-99.8) The reduced sensitivity of Xpert MTB/RIF compared with the CRS versus a culture reference standard suggests that either the CRS is too broad or that culture as the single reference standard is inadequate.

Figure 13: Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in cerebrospinal fluid with composite reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are labelled with year 2013 or blackened.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caws 2013</td>
<td>43</td>
<td>0</td>
<td>28</td>
<td>70</td>
<td>0.61 [0.48, 0.72]</td>
<td>1.00 [0.95, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patel 2013</td>
<td>20</td>
<td>5</td>
<td>23</td>
<td>101</td>
<td>0.47 [0.31, 0.62]</td>
<td>0.95 [0.89, 0.98]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tortoli 2012</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td>118</td>
<td>0.86 [0.57, 0.98]</td>
<td>0.99 [0.95, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vadwai 2011</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>14</td>
<td>0.20 [0.01, 0.72]</td>
<td>1.00 [0.77, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zeka 2011</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>26</td>
<td>0.60 [0.15, 0.95]</td>
<td>1.00 [0.87, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prevalence of HIV and condition of specimen did not have an effect on the performance estimates of Xpert MTB/RIF in CSF. Ten of the sixteen studies comparing Xpert MTB/RIF with a culture reference standard used a concentration step in the processing of the sample. Six studies did not use a concentration step. A concentration step appeared to increase sensitivity of the Xpert MTB/RIF (82%, 95% CI 71%-93% versus 56%, 95% CI 36%-77% for un-concentrated samples) although confidence intervals overlapped. A concentration steps did not affect the specificity.
### 4.2.3.5. Detection of TB in gastric fluid

12 studies examined the performance of Xpert MTB/RIF in gastric fluid against a culture reference standard (8 with more than 10 samples, total of 1258 samples). Two studies included children only. The remaining 10 studies included adults and children (proportion of children included across sample types ranged between 0% to 33.5%). One study with 788 samples with valid results on Xpert MTB/RIF accounted for 62.6% of all samples of this specimen type. Estimates for sensitivity varied from 69% to 100% while specificity varied from 98 to 100% with one at 52% (Figure 14). Pooled sensitivity across studies was 83.8% (95% CI: 65.9-93.2%) and specificity was 98.1% (95% CI: 92.3-99.5%) which suggests good accuracy of Xpert MTB/RIF for detection of TB in gastric fluid. Indeterminate results on Xpert MTB/RIF were reported for 2 studies (Bates, 1.5% and one unpublished study, 2.3%).

**Figure 14:** Forest plot of Xpert MTB/RIF sensitivity and specificity for TB detection in gastric fluid with culture reference standard.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are labelled with year 2013 or blackened.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armand 2011</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0.70 [0.53, 0.93]</td>
<td>1.00 [0.16, 1.00]</td>
<td>0.70 [0.53, 0.93]</td>
<td>1.00 [0.16, 1.00]</td>
</tr>
<tr>
<td>Bates 2012</td>
<td>33</td>
<td>15</td>
<td>735</td>
<td>17</td>
<td>0.69 [0.54, 0.81]</td>
<td>0.99 [0.98, 1.00]</td>
<td>0.69 [0.54, 0.81]</td>
<td>0.99 [0.98, 1.00]</td>
</tr>
<tr>
<td>Caussé 2011</td>
<td>8</td>
<td>0</td>
<td>46</td>
<td>46</td>
<td>1.00 [0.63, 1.00]</td>
<td>1.00 [0.92, 1.00]</td>
<td>1.00 [0.63, 1.00]</td>
<td>1.00 [0.92, 1.00]</td>
</tr>
<tr>
<td>Hanif 2011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.00 [0.03, 1.00]</td>
<td>1.00 [0.93, 1.00]</td>
<td>1.00 [0.03, 1.00]</td>
<td>1.00 [0.93, 1.00]</td>
</tr>
<tr>
<td>Hilleman 2011</td>
<td>7</td>
<td>0</td>
<td>19</td>
<td>26</td>
<td>0.88 [0.47, 1.00]</td>
<td>1.00 [0.82, 1.00]</td>
<td>0.88 [0.47, 1.00]</td>
<td>1.00 [0.82, 1.00]</td>
</tr>
<tr>
<td>Malbruny 2011</td>
<td>5</td>
<td>0</td>
<td>28</td>
<td>33</td>
<td>1.00 [0.48, 1.00]</td>
<td>1.00 [0.88, 1.00]</td>
<td>1.00 [0.48, 1.00]</td>
<td>1.00 [0.88, 1.00]</td>
</tr>
<tr>
<td>Moure 2012</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.00 [0.09, 0.99]</td>
<td>1.00 [0.29, 1.00]</td>
<td>1.00 [0.09, 0.99]</td>
<td>1.00 [0.29, 1.00]</td>
</tr>
<tr>
<td>Torlatti 2012</td>
<td>45</td>
<td>3</td>
<td>13</td>
<td>163</td>
<td>0.78 [0.65, 0.87]</td>
<td>0.98 [0.95, 1.00]</td>
<td>0.78 [0.65, 0.87]</td>
<td>0.98 [0.95, 1.00]</td>
</tr>
<tr>
<td>Walters 2012</td>
<td>3</td>
<td>0</td>
<td>17</td>
<td>17</td>
<td>1.00 [0.29, 1.00]</td>
<td>1.00 [0.99, 1.00]</td>
<td>1.00 [0.29, 1.00]</td>
<td>1.00 [0.99, 1.00]</td>
</tr>
</tbody>
</table>

The pooled estimates for published studies (7 studies; univariate analysis) were largely unchanged for both sensitivity and specificity. Univariate sensitivity for studies that used consecutive sampling (4 studies) was slightly increased (89.1%, 95% CI: 72.1-100%) while specificity was decreased at 90.5 (95% CI: 81-100%). Both estimates had wide confidence intervals, which highlights the lack of precision due to the limited amount of data.

All studies included a concentration step therefore this was not considered a source of heterogeneity within the data. The condition of the gastric fluid specimen (fresh versus frozen) did not appear to have a strong effect on the performance of Xpert MTB/RIF.

### 4.2.3.6. Detection of TB in tissue

There were 12 studies that tested Xpert MTB/RIF on tissue (from any site other than lymph node) against a culture reference standard (10 with more than 10 samples; 699 samples total). Estimates for sensitivity varied widely and ranged from 42% to 100% (Figure 15). The pooled sensitivity estimate was calculated as 81.2% (95% CI: 67.7-89.9%). Specificity was 98.1% (95% CI: 87.0-99.8%). One study represented an outlier with respect to specificity with an estimate of 61.1%, while other studies ranged from 84.6 to 100%. Indeterminate results on Xpert MTB/RIF were present in the studies by Hilleman (3.4%, 6/176) and Tortoli (1.6%; 4/254). The condition of the specimen did not have an impact on the Xpert MTB/RIF performance characteristics in tissue samples.
4.2.4. Rifampicin resistance detection

Data for rifampicin resistance detection was only used from published studies as for some unpublished studies data collection on resistance detection was incomplete. Furthermore, data from a study was only included if DST was done for all culture and Xpert MTB/RIF positive samples because a selective confirmation of results could have introduced bias.

In total, data on resistance testing were available for 566 samples from 13 studies. Forty-one samples were confirmed to be rifampicin resistant on phenotypic DST. Given the limited amount of data, no summary estimate was calculated. The average prevalence of rifampicin resistance across the studies was 5.4% with the highest prevalence reported from India (25.6%). Xpert MTB/RIF did not identify two of the 41 phenotypically rifampicin-resistant samples. Six of the 41 samples which were identified to be rifampicin resistant on Xpert MTB/RIF testing were found to be susceptible on phenotypic DST. Five of these six samples underwent sequencing of the \textit{rpoB} gene and four were found to have a mutation in the same region of the \textit{rpoB} gene at codon 533. Hence, Xpert MTB/RIF detected four additional rifampicin resistant strains that would have been missed by phenotypic DST alone.

4.2.5. Summary of findings

The review of the diagnostic accuracy of Xpert MTB/RIF in non-respiratory samples identified 15 published and 7 unpublished studies. An analysis was performed against either a culture reference standard or separately against an author-defined CRS (that included culture in all studies).

Good performance of Xpert MTB/RIF in was observed in smear-positive samples across sample types (sensitivity of 97.6%; 95% CI: 95.2-99.9 %) and the low number of indeterminate results (1.4%) supports the use of the technique in specific non-respiratory samples. However, the variability of performance characteristics, in particular sensitivity, between non-respiratory sample types suggests that a different approach might be needed for different sample types. While Xpert MTB/RIF could be considered as a diagnostic tool for the evaluation of TB in tissue and lymph node samples, gastric fluid and CSF, the benefit for testing pleural fluid is limited. Furthermore, these recommendations do not apply to stool, urine or blood, given data on the utility of Xpert MTB/RIF on these specimens are still limited.

In respect to rifampicin resistance testing, the available data were limited and did not allow for the calculation of pooled estimates. However, given the mechanism of detection used with Xpert MTB/RIF it is unlikely that accuracy will differ from that estimated in respiratory samples.
The results demonstrated that for TB detection in non-respiratory samples:

- Xpert MTB/RIF sensitivity varied widely across different sample types if the smear-result was negative;
- Xpert MTB/RIF had good sensitivity if done on smear-positive samples;
- Xpert MTB/RIF had good sensitivity in comparison with culture in lymph node tissues or aspirates, gastric fluid, CSF and other tissue samples;
- Xpert MTB/RIF had good sensitivity in comparison with an author-defined CRS in lymph node tissues or aspirates;
- Xpert MTB/RIF had poor sensitivity in pleural fluid;
- The proportion of indeterminate results was low; and
- There was substantial heterogeneity even within subgroups classified by sample type, and therefore the pooled estimates must be interpreted with caution.

Figure 16 shows the summary estimates for different sample types.

**Figure 16: Summary estimates across sample types for (A) sensitivity and (B) specificity**
4.2.6. Strengths and limitations of the evidence base

Strengths of the review include the use of a standard protocol, strict inclusion criteria, standardized data extraction, independent reviewers, a bivariate model for meta-analysis, and pre-specified subgroups to account for heterogeneity.

This data set involved comprehensive searching and correspondence with expertise in the field and the test manufacturer to identify additional published and unpublished studies, as well as repeated correspondence with study authors to obtain additional data and information that was missing in the papers. The search strategy included studies published in all languages. The majority of studies used consecutive selection of participants and interpreted the reference standard results without knowledge of Xpert MTB/RIF results. Xpert MTB/RIF results are generated automatically, without requiring subjective interpretation.

However, the review also had several limitations. The meta-analysis was limited by the small number of studies for the different sample types, particularly those using a CRS. Also, low event rates (i.e. number of confirmed TB cases) limited the precision of our sensitivity estimates. Furthermore the sample processing was highly variable across studies and within studies.

The sensitivity and specificity estimates in the meta-analysis might be overly optimistic for the following reasons: (1) study quality for some studies suffered from lack of a representative patient spectrum, which could result in exaggerated estimates of test accuracy; and (2) all of the studies were performed in tertiary care centers or reference laboratories, where performance characteristics might be better and where patients might present later in their disease process.
4.2.7. GRADE Evaluations and Recommendations

GRADE evidence profiles are provided in Tables 15 to 20. The GRADE evaluation supports the use of Xpert MTB/RIF in the diagnosis of extrapulmonary TB and rifampicin resistance. The Expert Group therefore concluded that:

1. Xpert MTB/RIF should be used in preference to conventional microscopy and culture as the initial diagnostic test in testing CSF from patients presumed to have TB meningitis (strong recommendation given the urgency for rapid diagnosis, very low quality of evidence).

   Note: The Expert Group noted that a negative CSF Xpert MTB/RIF result should be followed up by other tests. The Expert Group also noted that concentration methods should be used to enhance yield when sufficient volume of CSF is available. These recommendations apply to both children and adults.

2. Xpert MTB/RIF may be used as a replacement test for usual practice, including conventional microscopy and culture, in testing lymph nodes and tissues from patients presumed to have extrapulmonary TB (conditional recommendation, very low quality of evidence).

   Note: The Expert Group noted that a negative Xpert MTB/RIF result should be followed by other tests. The Expert Group also noted that sample processing methods for lymph nodes and tissues need to be standardised to optimise yield. These recommendations apply to both adults and children.

3. Xpert MTB/RIF should not be used in the diagnostic workup in patients presumed to have pleural TB (conditional recommendation; very low quality of evidence).

   Note: The Expert Group noted that pleural fluid is a suboptimal sample for the diagnosis of pleural TB in general and that a pleural biopsy is the preferred sample for bacteriological confirmation. These recommendations apply to both adults and children.
**Table 15: Xpert MTB/RIF assay for tuberculosis detection in lymph node fluid and tissue: A. Evidence profile B. Summary of findings**

**PICO question 1:** What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in lymph node fluid and tissue, where Xpert MTB/RIF is used as a replacement test for usual practice?

**A) Evidence profile**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
</tr>
<tr>
<td><strong>True positives (patients with TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
</tr>
<tr>
<td><strong>True negatives (patients without TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
</tr>
<tr>
<td><strong>False positives (patients incorrectly classified as having TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
</tr>
<tr>
<td><strong>False negatives (patients incorrectly classified as not having TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
</tr>
</tbody>
</table>

Footnotes: For each outcome, quality of evidence started high when there were randomized controlled trials or high quality observational studies (cross-sectional studies with diagnostic uncertainty and direct comparison of index test results with a reference standard) and low for case-control studies. The evidence was downgraded one point when a serious issue
was identified and two points when a very serious issue was identified in any of the five factors that may decrease the quality of evidence: limitations, indirectness, inconsistency, imprecision, and publication bias.

1 The QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled patients consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and considered blinded in all studies. However the variable use of the composite reference standard across studies was a concern for the possible introduction of bias. Evidence was downgraded 1 point.

2 The majority of studies ran Xpert MTB/RIF in tertiary care centers or reference laboratories. As obtaining the sample requires an invasive procedure, the test will likely be performed at higher levels of care than feasible for pulmonary TB. Thus, the study populations are probably representative of the population that will receive the test. Evidence was not downgraded.

3 Unexplained heterogeneity in findings might originate from differences in sample processing, sample condition and differences in study population (varying prevalence, HIV). Evidence was downgraded 1 point.

4 Confidence intervals are large likely in parts due to unexplained heterogeneity as above and in parts due to verification bias because of the imperfect reference standard. Evidence was downgraded 1 point.

5 Unpublished studies were included. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there is going to be considerable attention and scrutiny, reporting bias was considered to be minimal.
### B) Summary of findings

**Reference standard culture:**

*Number of studies (number of samples):* 14 studies total, 11 with more than 10 samples (849 samples)
- **Pooled sensitivity:** 85% (95% CI 72, 92)
- **Pooled specificity:** 93% (95% CI 80, 97)

**Composite reference standard:**

*Number of studies (number of samples):* 5 studies total, 409 samples
- **Pooled sensitivity:** 84% (95% CI 74, 90)
- **Pooled specificity:** 99% (95% CI 88, 100)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Prevalence 2.5%</th>
<th>Prevalence 5%</th>
<th>Prevalence 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Culture</td>
<td>CRS</td>
<td>Culture</td>
</tr>
<tr>
<td><strong>True positives (patients with TB)</strong></td>
<td>21 (18, 23)</td>
<td>21 (19, 23)</td>
<td>43 (36, 46)</td>
</tr>
<tr>
<td><strong>True negatives (patients without TB)</strong></td>
<td>907 (780, 946)</td>
<td>965 (858, 975)</td>
<td>884 (769, 922)</td>
</tr>
<tr>
<td><strong>False positives (patients incorrectly classified as having TB)</strong></td>
<td>68 (29, 195)</td>
<td>10 (0, 117)</td>
<td>67 (29, 190)</td>
</tr>
<tr>
<td><strong>False negatives (patients incorrectly classified as not having TB)</strong></td>
<td>4 (2, 7)</td>
<td>3 (4, 7)</td>
<td>8 (4, 14)</td>
</tr>
</tbody>
</table>

**Quality of Evidence**
- Very low 
- Very low 
- Very low
### Table 16: Xpert MTB/RIF assay for tuberculosis detection in pleural fluid: A. Evidence profile B. Summary of findings

**PICO question 2:** What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in pleural fluid, where Xpert MTB/RIF is used as a replacement test for usual practice?

#### A) Evidence profile

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True positives (patients with TB)</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
<tr>
<td>True negatives (patients without TB)</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having TB)</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having TB)</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
</tbody>
</table>

**Footnotes:**
For each outcome, quality of evidence started high when there were randomized controlled trials or high quality observational studies (cross-sectional studies with diagnostic uncertainty and direct comparison of index test results with a reference standard) and low for case-control studies. We then downgraded one point when a serious issue was
identified and two points when a very serious issue was identified in any of the five factors that may decrease the quality of evidence: limitations, indirectness, inconsistency, imprecision, and publication bias.

1 The QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled patients consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and considered blinded in all studies. However the variable use of the composite reference standard across studies was a concern for the possible introduction of bias. Evidence was downgraded 1 point.

2 The majority of studies ran Xpert MTB/RIF in tertiary care centers or reference laboratories. As obtaining the sample requires an invasive procedure, the test will likely be performed at higher levels of care than feasible for pulmonary TB. Thus, the study populations are probably representative of the population that will receive the test. Evidence was not downgraded.

3 Unexplained heterogeneity in findings might originate from differences in sample processing, sample condition and differences in study population (varying prevalence, HIV). Evidence was downgraded 1 point.

4 Confidence intervals are large likely in parts due to unexplained heterogeneity as above and in parts due to verification bias because of the imperfect reference standard. Evidence was downgraded 1 point.

5 Unpublished studies were included. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there is going to be considerable attention and scrutiny, reporting bias was considered to be minimal.
B) Summary of findings

Reference standard culture:
Number of studies (number of samples): 17 studies total, 16 with more than 10 samples (1384 samples)
- Pooled sensitivity: 44% (95% CI 25, 65)
- Pooled specificity: 98% (95% CI 95, 99)

Composite reference standard:
Number of studies (number of samples): 7 studies total, (698 samples)
- Pooled sensitivity: 17% (95% CI: 8, 34)
- Pooled specificity: 100% (95% CI 94, 100)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Effect per 1000 patients with presumed TB for varying prevalence settings comparing Xpert MTB/RIF against culture and composite reference standard (CRS)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 2.5%</td>
<td>Prevalence 5%</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>True positives</strong> (patients with TB)</td>
<td>Culture</td>
<td>CRS</td>
</tr>
<tr>
<td></td>
<td>11 (6, 16)</td>
<td>4 (2, 9)</td>
</tr>
<tr>
<td><strong>True negatives</strong> (patients without TB)</td>
<td>956 (926, 965)</td>
<td>975 (917, 975)</td>
</tr>
<tr>
<td><strong>False positives</strong> (patients incorrectly classified as having TB)</td>
<td>20 (10, 49)</td>
<td>0 (0, 59)</td>
</tr>
</tbody>
</table>

Quality of Evidence:
- Very low: ☹️☹️☹️☹️
- Low: ☹️☹️☹️
**False negatives (patients incorrectly classified as not having TB)**

<table>
<thead>
<tr>
<th></th>
<th>14 (9, 19)</th>
<th>21 (17, 23)</th>
<th>28 (18, 38)</th>
<th>42 (33, 46)</th>
<th>56 (35, 75)</th>
<th>83 (66, 92)</th>
<th>Very low</th>
</tr>
</thead>
</table>

Table 17: Xpert MTB/RIF assay for tuberculosis detection in cerebrospinal fluid: A. Evidence profile B.Summary of findings

**PICO question 3.** What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in cerebrospinal fluid (CSF), where Xpert MTB/RIF is used as a replacement test for usual practice?

**A) Evidence profile**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
</tr>
<tr>
<td><strong>True positives (patients with TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
<tr>
<td><strong>True negatives (patients without TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
<tr>
<td><strong>False positives (patients incorrectly classified as having TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
<tr>
<td><strong>False negatives (patients incorrectly classified as not having TB)</strong></td>
<td>Majority cross-sectional</td>
<td>Serious (-1)</td>
<td>None</td>
</tr>
</tbody>
</table>
Footnotes:
For each outcome, quality of evidence started high when there were randomized controlled trials or high quality observational studies (cross-sectional studies with diagnostic uncertainty and direct comparison of index test results with a reference standard) and low for case-control studies. We then downgraded one point when a serious issue was identified and two points when a very serious issue was identified in any of the five factors that may decrease the quality of evidence: limitations, indirectness, inconsistency, imprecision, and publication bias.

1 The QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled patients consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and considered blinded in all studies. However, the variable use of the composite reference standard across studies was a concern for the possible introduction of bias. Evidence was downgraded 1 point.

2 The majority of studies ran Xpert MTB/RIF in tertiary care centers or reference laboratories. As obtaining the sample requires an invasive procedure, the test will likely be performed at higher levels of care than feasible for pulmonary TB. Thus, the study populations are probably representative of the population that will receive the test. Evidence was not downgraded.

3 Unexplained heterogeneity in findings might originate from differences in sample processing, sample condition and differences in study population (varying prevalence, HIV). Evidence was downgraded 1 point.

4 Confidence intervals are large likely in parts due to unexplained heterogeneity as above and in parts due to verification bias because of the imperfect reference standard. Evidence was downgraded 1 point.

5 Unpublished studies were included. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there is going to be considerable attention and scrutiny, reporting bias was considered to be minimal.
### B) Summary of findings

**Reference standard culture:**
- **Number of studies (number of samples):** 16 studies total, 13 with more than 10 samples (709 samples)
  - Pooled sensitivity: 80% (95% CI 62, 90)
  - Pooled specificity: 99% (95% CI 96, 100)

**Composite reference standard:**
- **Number of studies (number of samples):** 7 studies total, (698 samples)
  - Pooled sensitivity: 56% (95% CI: 44, 66)
  - Pooled specificity: 99% (95% CI 95, 100)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Effect per 1000 patients with presumed TB for varying prevalence settings comparing Xpert MTB/RIF against culture and composite reference standard (CRS)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 2.5%</td>
<td>Prevalence 5%</td>
</tr>
<tr>
<td></td>
<td>Culture</td>
<td>CRS</td>
</tr>
<tr>
<td>True positives (patients with TB)</td>
<td>20 (16, 23)</td>
<td>14 (11, 17)</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having TB)</td>
<td>10 (0, 39)</td>
<td>10 (0, 49)</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having TB)</td>
<td>31 (3, 10)</td>
<td>9 (11, 14)</td>
</tr>
</tbody>
</table>
Table 18: Xpert MTB/RIF assay for tuberculosis detection gastric fluid.

**PICO question 4.** What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in gastric fluid, where Xpert MTB/RIF is used as a replacement test for usual practice?

**Reference standard culture**

**Number of studies (number of samples):** 12 studies total, 8 with more than 10 samples, (1258 samples)

**Pooled sensitivity:** 84% (95% CI 66, 93)

**Pooled specificity:** 98% (95% CI 92, 100)

| Outcome | Study Design | Factors that may decrease the quality of evidence | Quality of Evidence | Effect per 1000 patients with presumed TB for varying prevalence settings comparing Xpert MTB/RIF against culture and CRS |
|---------|--------------|--------------------------------------------------|---------------------|-----------------------------------------------------------------------------------------------------------------
| True positives (patients with TB) | Majority cross-sectional | Serious (-1) ¹  None ²  Serious (-1) ³  Serious (-1) ⁴  Undetected ⁵ | Very low ☒  || 21 (17, 23) 42 (33, 47) 84 (66, 93) |
| True negatives (patients without TB) | Majority cross-sectional | Serious (-1) ¹  None ²  Serious (-1) ³  Serious (-1) ⁴  Undetected ⁵ | Very low ☒ || 956 (897, 975) 931 (874, 950) 882 (828, 900) |
| False positives (patients incorrectly classified as having TB) | Majority cross-sectional | Serious (-1) ¹  None ²  Serious (-1) ³  Serious (-1) ⁴  Undetected ⁵ | Very low ☒ || 20 (0, 78) 19 (0, 76) 18 (0, 72) |
| False negatives (patients incorrectly classified as having TB) | Majority cross-sectional | Serious (-1) ¹  None ²  Serious (-1) ³  Serious (-1) ⁴  Undetected ⁵ | Very low ☒ || 4 (2, 9) 8 (4, 17) 16 (7, 34) |
Footnotes: For each outcome, quality of evidence started high when there were randomized controlled trials or high quality observational studies (cross-sectional studies with diagnostic uncertainty and direct comparison of index test results with a reference standard) and low for case-control studies. We then downgraded one point when a serious issue was identified and two points when a very serious issue was identified in any of the five factors that may decrease the quality of evidence: limitations, indirectness, inconsistency, imprecision, and publication bias.

1 The QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled patients consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and considered blinded in all studies. However the variable use of the composite reference standard across studies was a concern for the possible introduction of bias. Evidence was downgraded 1 point.

2 The majority of studies ran Xpert MTB/RIF in tertiary care centers or reference laboratories. As obtaining the sample requires an invasive procedure, the test will likely be performed at higher levels of care than feasible for pulmonary TB. Thus, the study populations are probably representative of the population that will receive the test. Evidence was not downgraded.

3 Unexplained heterogeneity in findings might originate from differences in sample processing, sample condition and differences in study population (varying prevalence, HIV). Evidence was downgraded 1 point.

4 Confidence intervals are large likely in parts due to unexplained heterogeneity as above and in parts due to verification bias because of the imperfect reference standard. Evidence was downgraded 1 point.

5 Unpublished studies were included. Data included did not allow for formal assessment of publication bias was conducted using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there is going to be considerable attention and scrutiny, reporting bias was considered to be minimal.
Table 19: Xpert MTB/RIF assay for tuberculosis detection in tissue samples.

**PICO question 5.** What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in tissue samples, where Xpert MTB/RIF is used as a replacement test for usual practice?

**Reference standard culture**

**Number of studies (number of samples):** 12 studies total, 10 with more than 10 samples (699 samples)

- **Pooled sensitivity:** 81% (95% CI 68, 90)
- **Pooled specificity:** 98% (95% CI 87, 100)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Effect per 1000 patients with presumed TB for varying prevalence settings comparing Xpert MTB/RIF against culture and CRS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirect-ness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (patients with TB)</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
<td>Serious (-1)¹</td>
</tr>
<tr>
<td>True negatives (patients without TB)</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
<td>Serious (-1)¹</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having TB)</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
<td>Serious (-1)¹</td>
</tr>
<tr>
<td>False negatives (patients incorrectly</td>
<td>Majority cross-sectional</td>
<td>Serious (-1)¹</td>
<td>None²</td>
<td>Serious (-1)¹</td>
</tr>
</tbody>
</table>
The QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled patients consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and considered blinded in all studies. However, the variable use of the composite reference standard across studies was a concern for the possible introduction of bias. Evidence was downgraded 1 point.

The majority of studies ran Xpert MTB/RIF in tertiary care centers or reference laboratories. As obtaining the sample requires an invasive procedure, the test will likely be performed at higher levels of care than feasible for pulmonary TB. Thus, the study populations are probably representative of the population that will receive the test. Evidence was not downgraded.

Unexplained heterogeneity in findings might originate from differences in sample processing, sample condition and differences in study population (varying prevalence, HIV). Evidence was downgraded 1 point.

Confidence intervals are large likely in parts due to unexplained heterogeneity as above and in parts due to verification bias because of the imperfect reference standard. Evidence was downgraded 1 point.

Unpublished studies were included. Data included did not allow for formal assessment of publication bias was conducted using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there is going to be considerable attention and scrutiny, reporting bias was considered to be minimal.
**Table 20**: Xpert MTB/RIF assay for rifampicin resistance detection in non-respiratory specimens. A. Evidence profile

**PICO question 6.** What is the diagnostic accuracy of Xpert MTB/RIF for rifampicin resistance detection in non-respiratory specimens, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing?

**Number of studies (number of samples):** 13 (566 samples)
Given the limited amount of data to address this question, we did not calculate summary estimates. Six studies reported a sensitivity of 100% (34 rifampicin resistant cases), 1 study with 50% sensitivity (2 rifampicin resistant cases), 1 study with 0% sensitivity (1 rifampicin resistant case).

A) Evidence profile

<table>
<thead>
<tr>
<th>Outcome (patients with rifampicin-resistant TB)</th>
<th>Study Design</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>Majority cross-sectional</td>
<td>Very low</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having rifampicin-resistant TB)</td>
<td>Majority cross-sectional</td>
<td>Very low</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having rifampicin-resistant TB)</td>
<td>Majority cross-sectional</td>
<td>Very low</td>
</tr>
</tbody>
</table>

Factors that may decrease the quality of evidence:

- **Limitations**
- **Indirectness**
- **Inconsistency**
- **Imprecision**
- **Publication Bias**

<table>
<thead>
<tr>
<th>Outcome (patients with rifampicin-resistant TB)</th>
<th>Study Design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>Majority cross-sectional</td>
<td>None ¹</td>
<td>Serious (-1)²</td>
<td>Serious (-1)¹</td>
<td>Serious (-1)⁴</td>
<td>Undetected ⁵</td>
<td>Very low</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having rifampicin-resistant TB)</td>
<td>Majority cross-sectional</td>
<td>None ¹</td>
<td>Serious (-1)²</td>
<td>Serious (-1)³</td>
<td>Serious (-1)⁴</td>
<td>Undetected ⁵</td>
<td>Very low</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having rifampicin-resistant TB)</td>
<td>Majority cross-sectional</td>
<td>None ¹</td>
<td>Serious (-1)²</td>
<td>Serious (-1)³</td>
<td>Serious (-1)⁴</td>
<td>Undetected ⁵</td>
<td>Very low</td>
</tr>
</tbody>
</table>
Footnotes:
For each outcome, quality of evidence started high when there were randomized controlled trials or high quality observational studies (cross-sectional studies with diagnostic uncertainty and direct comparison of index test results with a reference standard) and low for case-control studies. Evidence was downgraded one point when a serious issue was identified and two points when a very serious issue was identified in any of the five factors that may decrease the quality of evidence: limitations, indirectness, inconsistency, imprecision, and publication bias.

1 The QUADAS-2 tool was used to assess risk of bias. The majority of studies enrolled patients consecutively and assessed the reference standard result blinded to the Xpert MTB/RIF result. The Xpert MTB/RIF result is automated and considered blinded in all studies. The reference standard of phenotypic drug-susceptibility was used across all studies. Evidence was not downgraded.

There are concerns regarding phenotypic drug-susceptibility testing not being a perfect reference standard for detection of rifampicin resistance as sequencing data has shown that Xpert MTB/RIF might correctly identify resistance that is not identified based on phenotypic DST. Across all studies included here, six false positives results on Xpert MTB/RIF were identified. Of those 5 were tested with sequencing and 4 out of 5 were found to have a mutation in codon 533.

2 The majority of studies ran Xpert MTB/RIF in tertiary care centers or reference laboratories. As obtaining the sample requires an invasive procedure, the test will likely be performed at higher levels of care than feasible for pulmonary TB. Thus, the study populations are probably representative of the population that will receive the test. Evidence was not downgraded.

3 Unexplained heterogeneity in findings might originate from differences in specimen types tested and sample processing. Evidence was downgraded 1 point.

4 There was insufficient data to obtain a summary estimate for rifampin resistance. The individual study results varied widely; therefore, evidence was downgraded 1 point.

5 Unpublished studies were not included in the assessment of rifampicin resistance detection because data collection on drug-susceptibility testing was not complete for some of these studies. However, with Xpert MTB/RIF being a new test for which there is going to be considerable attention and scrutiny, reporting bias was considered to be minimal.
4.2.8. Further research needs

4.2.8.1. Optimised sample processing

There was evidence that concentration methods enhanced the sensitivity of Xpert MTB/RIF for the detection of TB in CSF specimens although detailed investigation of optimized sample processing was not possible within this review. As a priority, the Expert Group highlighted the need to develop standardised protocols for processing samples for use in subsequent studies employing Xpert MTB/RIF in non-respiratory specimens.

The Expert Group recommended that processing protocols not only focus on one sample type in particular but assess differences in performance with modified processing in the individual steps: homogenization, concentration, decontamination, and sample reagent to sample ratio. Optimization of sample preparation and DNA extraction as well as purification might further enhance accuracy. The specific steps necessary will likely vary across sample types.

4.2.8.2. Sample types for testing

All of the studies published to date examined samples from routine care in hospital or reference laboratories. Further research is needed to assess the difference in performance of Xpert MTB/RIF after a freeze-thaw cycle, as centralized testing combined with specimen transport would involve specimen preservation through freezing.

4.2.8.3. Reference standards for EPTB

Additional research would be beneficial to elucidate the performance of Xpert MTB/RIF in non-respiratory samples within subgroups at high risk of EPTB (e.g. HIV-infected patients). Effort should also be undertaken to optimize the reference standard for evaluation of Xpert MTB/RIF and other new diagnostics in EPTB. An optimized and standardized case definition for the diagnosis of EPTB can facilitate comparisons across studies. A case definition has already been published for TB meningitis, however for pleural TB, musculoskeletal TB or TB lymphadenitis no case definition is currently available.
4.3. Xpert MTB/RIF for the diagnosis of pulmonary TB, peripheral lymph node TB and TB meningitis in children

4.3.1. Study characteristics

Studies that assessed Xpert MTB/RIF for the diagnosis of pulmonary TB (PTB), peripheral lymph node TB (pLN TB), TB meningitis (TBM), and rifampicin resistance in children 0-15 years of age with presumptive TB were included. All published articles, articles in press and unpublished studies (with authors' agreement) were shared confidentially with the Expert Group members.

Cross-sectional studies, cohort studies, and randomized controlled trials (RCTs) were included if they compared Xpert MTB/RIF to an acceptable reference standard (see below). Case-control studies, case reports as well as studies only presented as abstracts were excluded.

Studies had to include children aged 0-15 years presumed to have TB without pre-defined criteria for presumptive TB. Studies that recruited children from both in- and outpatient settings were considered as well as studies performed at any level of the health care system, or in research laboratories.

Studies that included and excluded HIV-infected populations, as well as children with other co-morbidities such as malnutrition so as to improve generalizability were also included. Authors of studies including both children and adults were contacted if the pediatric data were not presented individually. Studies that employed respiratory samples, rifampicin resistance testing on respiratory samples and non-respiratory samples, i.e. CSF or lymph node biopsies/aspirates were included.

The reference standard for PTB, pLN TB and TB meningitis was TB confirmed by at least one positive culture on solid media or a commercial liquid culture system such as BACTEC™ MGIT™ (mycobacterial growth indicator tube) 960 Mycobacterial Detection System, (BD, USA). The reference for rifampicin resistance was WHO recommended conventional phenotypic drug susceptibility testing (DST) on solid or liquid media (WHO Policy DST 2008) or molecular lineprobe assays. Recognizing the limitations of mycobacterial culture in children (culture sensitivity is approximately 30% to 60%), a second reference standard, clinical TB, was applied only in culture negative children. Children were categorized as positive for the clinical TB reference standard if they were culture negative and had started anti-TB therapy based on a clinical diagnosis of TB.

An electronic literature search was performed initially on January 24, 2013. In total, 39 articles were identified. Of these, 26 were excluded based on title and abstract review. Of the remaining 13 articles that underwent full-text review, 10 were considered in this review. An additional 2 published studies were included that were identified via a final electronic search on April 3, 2013. An additional 4 unpublished studies were included that were identified by querying child TB networks and contacting authors as previously described. In total, 16 studies were included (Annex 6).

4.3.2. Study Quality

The quality of included studies was assessed with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing.

Within the Patient Selection domain, the majority of studies (75%) was considered low risk of bias as the studies recruited children in a consecutive manner and avoided inappropriate exclusions. The remaining studies were considered high risk of bias because patient recruitment was either by convenience or unclear. There was high concern for applicability in this domain as 1) most studies were performed among inpatients, and 2) all studies were performed at higher levels of care (university hospitals, research laboratories) or were laboratory-based studies with unclear patient selection criteria and limited clinical information (3 studies).
Risk of bias for the index test was judged as low concern for all studies as Xpert MTB/RIF is fully automated. The applicability of the index test was rated as low concern for the majority of studies. Two studies were rated as unclear regarding applicability of the index test since specimen preparation was not clearly described.

The risk of bias for the reference standard was rated unclear for all studies, because mycobacterial culture as well as clinical case definitions are considered imperfect gold standards for childhood TB. However, applicability of the reference standard was considered to be of low concern for all studies.

Risk of bias for flow and timing was considered high in two studies. In the first study, 99 children were excluded from analysis as they were lost to follow-up. In the second study, children with a clinical diagnosis of TB and negative culture results were excluded from participation. In contrast, culture positive children with TB were included. Figure 17 shows the concerns regarding risk of bias and applicability for each study included.

**Figure 17: Risk of bias and applicability concerns graph: review authors’ judgments about each domain presented as percentages across included studies**

4.3.3. TB Detection- Xpert MTB/RIF as the initial test

Pulmonary TB was evaluated in 13 studies including 2603 participants. Studies either collected the same specimen type from all children or different types of specimen from different subgroups of children (e.g. expectorated sputum in older, induced sputum or gastric lavage or aspiration in younger children). In three studies different types of specimen were collected in each child. As a result, a total number of 3347 specimens were assessed (median number of specimens per study was 69, range 3-788): expectorated sputum (ES, 4 studies, 270 children), induced sputum (IS, 7 studies, 1279 children), nasopharyngeal aspirate (NPA, 1 study, 474 children), gastric lavage or aspiration (GLA, 6 studies, 1324 children).

The individual sensitivities and specificities of Xpert MTB/RIF per specimen type and study against a reference standard of culture are expressed as Forest plots in Figure 18. Sensitivities varied from 55 to 90% for ES, 40 to 100% for IS, and 40 to 100% for GLA. Specimen type confidence intervals were overlapping and suggest no superior specimen type. Specificities for all studies and specimen types ranged from 93 to 100%.

One study examined the yield of nasopharyngeal specimens. The sensitivity of Xpert MTB/RIF in NPA was 44% (95% CI 33 - 55%). In the same group of children, the sensitivity of IS was 60% (95% CI 49 - 70).
4.3.3.1 Differences in reference standards used

Among all studies included in the review, 13.2% of children had culture confirmed TB. The proportion of children with culture confirmed TB varied per study and specimen type (range 0-54.2%). In the majority of studies (9/13), multiple cultures were performed on single participants. Hence, the definition ‘culture positive’ was based on the presence of at least one positive culture result out of as many as 6 cultures performed. The average bacteriologic yield in studies using multiple cultures was increased compared to the group of four studies that based ‘culture positive’ on one culture result only. Studies also used differing culture techniques but the impact of this potential source of bias was not evaluated.

Children were categorized as positive for the clinical TB reference standard if they were culture negative and had started anti-TB therapy based on a clinical diagnosis of TB. This broad clinical reference standard was chosen in order to accommodate the heterogeneous study methods and clinical definitions expected in the studies. Children assigned to the group ‘clinical not TB’ (i.e. negative clinical TB reference standard) had to a) have another diagnosis assigned, and/or b) did not start anti-TB treatment (ATT) and improved or did not worsen after at least 1 month follow-up (from enrolment).

4.3.3.2 Accuracy of Xpert MTB/RIF versus Culture

The pooled sensitivity of Xpert MTB/RIF against culture was 66% for ES/IS (95% CrI 52%-77%), and 66% for GLA (95% CrI 51%-81%). The width of the confidence intervals indicated a high level of heterogeneity between studies. The specificity values of Xpert MTB/RIF against the reference standard ‘Culture’ were all ≥98% with narrow confidence intervals.

4.3.3.3 Accuracy of Xpert MTB/RIF versus Clinical TB

The sensitivity of Xpert MTB/RIF in culture-negative paediatric samples against the reference standard ‘Clinical TB’ was 4% for ES/IS and 15% for GLA with all confidence intervals being wide and therefore indicating a high level of heterogeneity. It is likely that the apparent poor performance of Xpert MTB/RIF was the result of a clinical TB reference standard that lacked specificity. The specificity values of Xpert MTB/RIF against the reference standard “clinical TB” were ≥99% with narrow confidence intervals.
Estimated sensitivities and specificities for Xpert MTB/RIF against the reference standard ‘Culture’ (published and unpublished studies) as well as against the reference standard ‘Clinical TB’ are given in Table 21. The exclusion of unpublished studies increased sensitivities to 69% for ES/IS and 75% for GLA, but the confidence intervals remained wide and overlapped between the two estimates suggesting the heterogeneity was retained.

Table 21: Meta-analysis. Estimated Xpert MTB/RIF Sensitivity and Specificity against the reference standard ‘Culture’ (published and unpublished studies) as well as against the reference standard ‘Clinical TB’.

<table>
<thead>
<tr>
<th>Category</th>
<th>Specimen type (No. of studies, No. of children)</th>
<th>Pooled sensitivity Median (pooled 95% credible interval)</th>
<th>Pooled specificity Median (pooled 95% credible interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xpert MTB/RIF against Reference standard ‘Culture’ Published and unpublished studies</td>
<td>ES and IS (10, 1546)*</td>
<td>66% (52, 77)</td>
<td>98% (96, 99)</td>
</tr>
<tr>
<td></td>
<td>GLA (7, 1319)†</td>
<td>66% (51, 81)</td>
<td>98% (96, 99)</td>
</tr>
<tr>
<td>Xpert MTB/RIF against Reference standard ‘Culture’ Published studies only</td>
<td>ES and IS (7, 1075)</td>
<td>69% (55, 81) (34, 91)</td>
<td>98% (97, 99)</td>
</tr>
<tr>
<td></td>
<td>GLA (5, 1045)</td>
<td>75% (59, 90) (41, 95)</td>
<td>99% (97, 100)</td>
</tr>
<tr>
<td>Xpert MTB/RIF against Reference standard ‘Clinical TB’ Published and unpublished studies</td>
<td>ES and IS (8, 995)‡</td>
<td>4% (01, 12)</td>
<td>100% (99, 100)</td>
</tr>
<tr>
<td></td>
<td>GLA (3, 269)§</td>
<td>15% (05, 31)</td>
<td>99% (96, 100)</td>
</tr>
</tbody>
</table>

* Bates 2013, Chisti unpublished, LaCourse unpublished, Nhu 2013, Nicol 2011, Rachow 2012 (ES and IS cohorts included as separate studies), Sekadde 2013, Walters unpublished, Zar 2012
‡ Studies included: Chisti unpublished, LaCourse unpublished, Nhu 2013, Nicol 2011, Rachow 2012 (ES and IS cohorts included as separate studies), Walters unpublished, Zar 2012
§ Studies included: Chisti unpublished, Nhu 2013, Walters unpublished

4.3.4. Xpert MTB/RIF compared with smear microscopy

The diagnostic accuracy of smear microscopy was calculated against a reference standard ‘Culture’ for the same studies and specimen types as the accuracy of Xpert MTB/RIF, above. The Forest plot in Figure 19 shows the individual estimates. Sensitivities varied from 30 to 60% for ES, 0 to 50% for IS, 0 to 50% for GLA, and was 18% in the one NPA cohort. Confidence intervals were wide and overlapping. The specificity was high (>93) for all studies and specimen types.
Figure 19: Forest plot. Sensitivity and specificity of smear microscopy against a reference standard ‘Culture’ by study and specimen type.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are blackened.

The pooled sensitivity was 29% (95% CrI 16% - 42%) for ES/IS and 22% (95% CrI 12% - 35%) for GLA with wide pooled confidence intervals indicating a high level of heterogeneity. Similar to Xpert MTB/RIF analyses, the specificity of Xpert MTB/RIF was >99% in both comparison groups (Table 22).

Table 22: Meta-analysis: Estimated sensitivity and specificity of smear microscopy against a reference standard ‘Culture’

<table>
<thead>
<tr>
<th>Category</th>
<th>Specimen type (No. of studies, No. of children)</th>
<th>Pooled sensitivity</th>
<th>Pooled specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median (pooled 95% credible interval)</td>
<td>Median (pooled 95% credible interval)</td>
</tr>
<tr>
<td>Smear microscopy against reference standard ‘Culture’</td>
<td>ES and IS (10, 1546)</td>
<td>29% (16, 42)</td>
<td>100% (99, 100)</td>
</tr>
<tr>
<td>Published and unpublished studies</td>
<td>GLA (7, 1319)</td>
<td>22% (12, 35)</td>
<td>99% (97, 100)</td>
</tr>
</tbody>
</table>

These data suggest that, in comparison to smear microscopy, the sensitivity of Xpert MTB/RIF was 37% higher if performed on ES/IS and 44% if performed on GLA.
4.3.5. Investigations of heterogeneity

Factors that might cause heterogeneity of results associated with smear status, HIV-infection, age as well as the approach to confirm TB were investigated.

4.3.5.1. Xpert MTB/RIF performance in smear-positive and smear-negative children

7 studies collectively containing data from 1083 children were included in the analysis of Xpert MTB/RIF in ES and IS. Six studies with data from 1259 children were included for the analysis of Xpert MTB/RIF in GLA. All included studies reported results for both smear positive and smear negative children (Figure 20).

Figure 20: Forest plot. Xpert MTB/RIF sensitivities in smear positive and smear negative children

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are blackened.

Xpert in smear positive children

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES_Bates 2013</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>2.00 (0.29, 1.00)</td>
<td>0.85 (0.42, 1.00)</td>
</tr>
<tr>
<td>ES_Nhau 2013</td>
<td>14</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.93 (0.68, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>ES_Rachow 2012</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.00 (0.54, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Bates 2013</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.00 (0.6, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Causse 2012</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0.86 (0.36, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Malbruny 2011</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Nhau 2013</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.00 (0.54, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Rachow 2012</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.00 (0.54, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Walters 2012</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.00 (0.54, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Sekkade 2013</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.00 (0.54, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Zar 2012</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
</tbody>
</table>

Xpert in smear negative children

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES_Bates 2013</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>124</td>
<td>0.86 (0.42, 1.00)</td>
<td>0.99 (0.96, 1.00)</td>
</tr>
<tr>
<td>ES_Nhau 2013</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td>0.70 (0.35, 0.93)</td>
<td>1.00 (0.85, 1.00)</td>
</tr>
<tr>
<td>ES_Rachow 2012</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>56</td>
<td>0.36 (0.13, 0.65)</td>
<td>0.92 (0.86, 1.00)</td>
</tr>
<tr>
<td>GLA_Bates 2013</td>
<td>22</td>
<td>4</td>
<td>14</td>
<td>73</td>
<td>0.61 (0.43, 0.77)</td>
<td>0.99 (0.99, 1.00)</td>
</tr>
<tr>
<td>GLA_Causse 2012</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>1.00 (0.40, 1.00)</td>
<td>1.00 (0.91, 1.00)</td>
</tr>
<tr>
<td>MAL_Causse 2012</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>200</td>
<td>0.40 (0.05, 0.85)</td>
<td>0.96 (0.93, 0.98)</td>
</tr>
<tr>
<td>GLA_Malbruny 2011</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>Not estimable</td>
<td>1.00 (0.48, 1.00)</td>
</tr>
<tr>
<td>GLA_Nhau 2013</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>1.00 (0.54, 1.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>GLA_Rachow 2012</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>3</td>
<td>0.50 (0.35, 0.65)</td>
<td>0.98 (0.94, 1.00)</td>
</tr>
<tr>
<td>GLA_Walters 2012</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>1.00 (0.16, 1.00)</td>
<td>1.00 (0.85, 1.00)</td>
</tr>
<tr>
<td>GLA_Sekkade 2013</td>
<td>2</td>
<td>14</td>
<td>3</td>
<td>192</td>
<td>0.40 (0.05, 0.85)</td>
<td>0.93 (0.89, 0.96)</td>
</tr>
<tr>
<td>GLA_Zar 2012</td>
<td>1</td>
<td>24</td>
<td>19</td>
<td>9</td>
<td>Not estimable</td>
<td>1.00 (0.99, 1.00)</td>
</tr>
<tr>
<td>GLA_Nhau 2013</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>30</td>
<td>0.50 (0.12, 0.88)</td>
<td>1.00 (0.96, 1.00)</td>
</tr>
<tr>
<td>MAL_Nhau 2012</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>42</td>
<td>0.25 (0.03, 0.65)</td>
<td>0.98 (0.85, 1.00)</td>
</tr>
<tr>
<td>MAL_Sekkade 2013</td>
<td>14</td>
<td>7</td>
<td>6</td>
<td>194</td>
<td>0.70 (0.46, 0.88)</td>
<td>0.97 (0.93, 0.99)</td>
</tr>
<tr>
<td>GLS 2013</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>41</td>
<td>0.43 (0.18, 0.71)</td>
<td>0.93 (0.81, 0.99)</td>
</tr>
<tr>
<td>GLS_Sekkade 2013</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>41</td>
<td>0.43 (0.18, 0.71)</td>
<td>0.93 (0.81, 0.99)</td>
</tr>
</tbody>
</table>

Xpert MTB/RIF sensitivities in ES/IS among children with smear negative results ranged from 25 to 86%. In contrast, Xpert MTB/RIF sensitivity among children with smear positive results ranged from 92 to 100%. The pooled estimate in smear positive children was 96% (95% CrI 90%-99%) and 55% (95% CrI 41%-69%) in smear negative children. The estimates were similar for Xpert MTB/RIF in GLA.
with an overall sensitivity of 95% (95% CrI 83%-99%) among smear positive and 62% (95% CrI 44%-80%) among smear negative children. The confidence intervals were wide (indicating variability) but did not overlap (Table 23).

Table 23: Meta-analysis. Xpert MTB/RIF against a reference standard ‘Culture’ in smear negative and smear positive children

<table>
<thead>
<tr>
<th>Category</th>
<th>Specimen type (No. of studies, No. of children)</th>
<th>Pooled sensitivity (pooled 95% credible interval)</th>
<th>Pooled specificity (pooled 95% credible interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ES and IS (7, 68)*</td>
<td>96% (90, 99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GLA (6, 32)†</td>
<td>95% (83, 99)</td>
<td></td>
</tr>
<tr>
<td>Xpert MTB/RIF in smear +</td>
<td>ES and IS (7, 1008)*</td>
<td>55% (41, 69)</td>
<td>98% (96, 99)</td>
</tr>
<tr>
<td></td>
<td>GLA (6, 1204)†</td>
<td>62% (44, 80)</td>
<td>99% (97, 99)</td>
</tr>
</tbody>
</table>

* Bates 2013, Nhu 2013, Nicol 2011, Rachow 2012 (ES only), Sekadde 2013, unpubl., Zar 2012
‡ There was not enough data to calculate specificity.

As expected, smear status was associated with Xpert MTB/RIF performance, indicating that Xpert MTB/RIF sensitivity was greater in children with higher mycobacterial burden compared to those with paucibacillary disease. Xpert MTB/RIF detected 55% (ES/IS) and 62% (GLA) of smear-negative, culture-positive children.

The data indicated that smear status is associated with Xpert MTB/RIF performance. This finding suggests that Xpert MTB/RIF sensitivity was greater in children with higher mycobacterial burden compared to those with paucibacillary disease.

Xpert MTB/RIF detected 55% (ES/IS) and 62% (GLA) of smear negative, culture positive children. This finding suggests that Xpert MTB/RIF has potential to effectively contribute in the diagnostic algorithm as an add-on following negative smear microscopy.

4.3.5.2. Xpert MTB/RIF in children aged 0-4 and 5-15 years
Five of the seven studies reported results for the 0-4 and 5-15 age groups. Data from 976 children were included for the analysis of Xpert MTB/RIF in ES and IS. Xpert MTB/RIF in GLA estimated accuracy in the 0-4 age group only (5 studies, 957 children).

Xpert MTB/RIF sensitivity in both age groups ranged from 0 to 100%. The pooled sensitivity among children aged 0-4 years was 57% for Xpert MTB/RIF in both ES/IS and GLA. The pooled sensitivities for ES/IS were higher in children aged 5-15 years (83% (95% CrI 68%-92%). Pooled specificity was ≥98% for all groups assessed with relatively narrow confidence intervals.
4.3.5.3. Xpert MTB/RIF in HIV-infected and HIV-uninfected children

7 studies with data from 1074 children were included for the analysis of Xpert MTB/RIF in ES and IS. All studies considered reported results for both HIV-infected and -uninfected children.

Sensitivities among HIV-infected children ranged from 20 to 100%, and among HIV-uninfected children ranged from 33 to 100% (Figure 22).
The pooled sensitivity among HIV-infected children (75%, 95% CrI 57%-88%) was higher compared with HIV-uninfected children (57%, 95% CrI 41%-71%), although confidence intervals were wide and overlapping. Specificity was 98% in both groups.

A stratified analysis comparing Xpert MTB/RIF performance according to smear status with the HIV-infected and HIV-uninfected strata demonstrated high sensitivity among smear-positive children regardless of HIV status. Xpert MTB/RIF sensitivity was lowest among HIV-uninfected children although confidence intervals were wide and overlapping (Table 24).

Table 24: Meta-analysis. Xpert MTB/RIF against a reference standard 'Culture' in HIV-infected and -uninfected children, subdivided by smear status.
A meta-regression model simultaneously controlling for smear and HIV status for Xpert MTB/RIF on ES and IS showed that the odds of test positivity was four fold greater in smear-positive compared with smear-negative children. The odds of Xpert MTB/RIF positivity was not statistically significant for HIV-infected children compared with HIV-uninfected children (Table 25).

Table 25: Meta-regression model for Xpert MTB/RIF on ES/IS, controlling for smear and HIV status.

<table>
<thead>
<tr>
<th>Node</th>
<th>Mean</th>
<th>SD</th>
<th>MC error</th>
<th>2.5%</th>
<th>Median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BETA 0</td>
<td>0.06201</td>
<td>0.385</td>
<td>0.0054</td>
<td>-0.8159</td>
<td>-0.064</td>
<td>0.7059</td>
</tr>
<tr>
<td>BETA 1 (HIV)</td>
<td>0.5863</td>
<td>0.5551</td>
<td>0.00862</td>
<td>-0.4919</td>
<td>0.5789</td>
<td>1.705</td>
</tr>
<tr>
<td>BETA 2 (Smear)</td>
<td>3.98</td>
<td>1.076</td>
<td>0.02855</td>
<td>2.159</td>
<td>3.878</td>
<td>6.399</td>
</tr>
<tr>
<td>SM-/HIV-</td>
<td>0.485</td>
<td>0.09264</td>
<td>0.0013</td>
<td>0.3066</td>
<td>0.484</td>
<td>0.6695</td>
</tr>
<tr>
<td>Sm+/HIV-</td>
<td>0.9694</td>
<td>0.03031</td>
<td>6.402</td>
<td>0.8873</td>
<td>0.9785</td>
<td>0.9983</td>
</tr>
<tr>
<td>Sm-/HIV+</td>
<td>0.6213</td>
<td>0.1101</td>
<td>0.001197</td>
<td>0.3944</td>
<td>0.6257</td>
<td>0.8216</td>
</tr>
<tr>
<td>Sm+/HIV+</td>
<td>0.988</td>
<td>0.07977</td>
<td>9.51</td>
<td>0.9284</td>
<td>0.9879</td>
<td>0.9991</td>
</tr>
</tbody>
</table>

There was insufficient data to perform a meta-analysis comparing Xpert MTB/RIF performance when using GLA in HIV-infected and HIV-uninfected children.

4.3.6. Xpert MTB/RIF for the detection of peripheral lymph node TB in children

Lymph node fine needle aspirates (FNAs) or biopsies for the diagnosis of peripheral lymph node (pLN) TB were evaluated in five studies. Two studies were excluded from the meta-analysis because the sample size was less than five. The analysis therefore included 3 studies with data from 172 children (Figure 23). The pooled sensitivity and specificity for Xpert MTB/RIF against culture as the reference standard were 86% (95% CrI 65%-96%) and 81% (95% CrI 54%-93%), respectively. Confidence intervals were wide for both sensitivity and specificity, indicating heterogeneity.

4.3.7. Xpert MTB/RIF for the detection of TB meningitis in children

CSF for the diagnosis of TB meningitis was evaluated in five studies among 61 children. In total 7/61 (11.5%) children had TB meningitis confirmed on CSF cultures. Of these, 3 children were Xpert MTB/RIF positive (3/61, 4.9%). Two studies were excluded from the meta-analysis because they did not include any culture-positive children. One of the remaining studies had a sub-group sample size <5. Hence, there was insufficient data to calculate sensitivity from the two remaining studies, in which 2/6 culture positive children had positive Xpert MTB/RIF results (Figure 23). Pooled specificity including 3 studies and 51 children was 95% (95% CrI 81%-99%), with relatively wide confidence intervals.
4.3.8. Xpert MTB/RIF for the detection of rifampicin resistance in children

In total, seven studies provided data on Xpert MTB/RIF for RIF resistance testing (Figure 23). Four studies used conventional phenotypic DST, and three used lineprobe assays. A meta-analysis of 3 studies including 176 participants showed a pooled sensitivity of 86% (95% CrI 53%-98%) and a pooled specificity of 98% (95% CrI 94%-100%).

Figure 23: Forest plot. Xpert MTB/RIF for the diagnosis of RIF resistance, peripheral lymph node TB and TB Meningitis.

The squares represent the sensitivity and specificity of one study, the black line its confidence interval. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Unpublished studies are blackened.

4.3.9. Summary of findings

- Xpert MTB/RIF shows moderate sensitivity for the detection of pulmonary TB in children on ES/IS and GLA (66%) against the reference standard culture.
- Xpert MTB/RIF shows high sensitivity in smear positive children.
- The specificity of Xpert MTB/RIF was consistently high (>93%). Few additional clinically confirmed, culture-negative TB cases can be detected by Xpert MTB/RIF, but the challenge of collecting good quality specimens from children may explain the low yield in culture negative children.
- The sensitivity of Xpert MTB/RIF was poor against a reference standard “clinical TB” which highlights the need for a composite reference standard that could be used universally for the evaluation of tests for the diagnosis of TB in children.
- The performance of Xpert MTB/RIF is superior to smear microscopy when both are examined against the reference standard of culture. Hence, Xpert MTB/RIF identifies additional TB cases if used as an add-on in children with negative smear results.
- The performance of Xpert MTB/RIF appears similar in HIV-positive and HIV-negative children and may have higher sensitivity as disease severity, associated with smear positivity, increases.
- The sensitivity of Xpert MTB/RIF was higher among children aged 5-15 compared to children aged 0-4 and is driven by smear status.
- Xpert MTB/RIF detected 86% of children with culture confirmed pLN TB on FNA specimens.
• There was insufficient data to calculate the pooled sensitivity of Xpert MTB/RIF on CSF for the detection of TB meningitis in children. However, Xpert MTB/RIF showed good sensitivity for the detection of TB meningitis in adults.
• Xpert MTB/RIF detected 86% of culture confirmed RIF resistance with a high specificity and has a potential to increase children’s access to DST.

4.3.10. Strengths and limitations of the evidence base

The findings of this review are based on comprehensive searching, strict inclusion criteria, and standardized data extraction. Despite limitations in the number of studies and participants included as well as heterogeneous methodological approaches amongst these studies, this review for the first time provides data on the accuracy of Xpert MTB/RIF for the diagnosis of PTB, RIF resistance, pLN TB and TBM in children.

The data set involved comprehensive searching and correspondence with expertise in the field to identify additional studies, as well as repeated correspondence with study authors to obtain additional data and information that was not provided in the papers. The search strategy included studies published in all languages. However, some studies may have been missed, particularly on-going work or articles in press, despite the comprehensive search.

Culture is regarded as the best available reference standard for active TB and was the reference standard used for TB in this review. Yet, its accuracy is recognized to be suboptimal in children. The limitations and implications of using culture as our reference standard are described and discussed throughout the review. A second, also suboptimal, reference standard, clinical TB, was applied to culture negative children only. Although this approach mimics clinical practice, it is methodologically flawed. Ideally, studies would apply each reference standard to all included children.

The majority of studies used consecutive selection of participants. Xpert MTB/RIF results are generated automatically, without requiring subjective interpretation. The majority of studies performed Xpert MTB/RIF at higher levels of care and among inpatients. In general, studies were fairly well reported, though we corresponded with all authors for additional data and missing information.

4.3.11. GRADE Evaluations and Recommendations

GRADE evidence profiles are provided in Tables 26 to 32. The GRADE process supports the use of Xpert MTB/RIF for the diagnosis of pulmonary and extrapulmonary TB and rifampicin resistance in children. The Expert Group therefore concluded that:

1. Xpert MTB/RIF should be used rather than conventional microscopy, culture and DST as the initial diagnostic test in children presumed to have MDR-TB or HIV-associated TB (strong recommendation given the difficulties in diagnosing paediatric TB, very low quality of evidence).

2. Xpert MTB/RIF may be used rather than conventional microscopy and culture in all other children presumed to have pulmonary TB. (conditional recommendation acknowledging resource implications, very low quality of evidence).

Note: The Expert Group noted that Xpert MTB/RIF should not be used as the only test in the diagnostic pathway of children with presumed TB, and that a child with high clinical suspicion for TB should be treated even if the Xpert MTB/RIF result is negative or if the test is not available.
3. Xpert MTB/RIF should be used in preference to conventional microscopy and culture as the initial diagnostic test in testing CSF from patients presumed to have TB meningitis (strong recommendation given the urgency for rapid diagnosis, very low quality of evidence).

*Note: The Expert Group noted that a negative Xpert MTB/RIF result should be followed up by other tests and that a child with high clinical suspicion for TB should be treated even if the Xpert MTB/RIF result is negative or if the test is not available. The Expert Group also noted that concentration methods should be used to enhance yield when sufficient volume of CSF is available.*

4. Xpert MTB/RIF may be used as a replacement test for usual practice, including conventional microscopy and culture, in testing lymph node fluid and tissue from children presumed to have peripheral lymph node TB (conditional recommendation, very low quality of evidence).

*Note: The Expert Group noted that a negative Xpert MTB/RIF result should be followed by other tests and that a child with high clinical suspicion for TB should be treated even if the Xpert MTB/RIF result is negative or if the test is not available. The Expert Group also noted that sample processing methods for lymph nodes and tissues need to be standardised to optimise yield.*
Table 26: GRADE evidence profile: Diagnostic accuracy of Xpert MTB/RIF for the detection TB in children against culture, where Xpert MTB/RIF is used as a replacement test for usual practice. A. ES/IS specimens B. GLA

**PICO question 1:** What is the diagnostic accuracy of Xpert MTB/RIF for the detection TB in children against culture, where Xpert MTB/RIF is used as a replacement test for usual practice?

**Participants:** Children 0-15 years with presumed TB  
**Setting:** Mainly tertiary care referral hospitals, university hospitals.  
**Target condition:** Pulmonary TB  
**Reference test:** Solid or liquid culture  
**Number of studies (number of participants):**

A  
**Expectorated sputum (ES) and induced sputum (IS) combined:** 10* (1546)  
**Pooled sensitivity:** 66% (95% CrI 52, 77)  
**Pooled specificity:** 98% (95% CrI 96, 99)  
*One study reporting one cohort with ES and one with IS was counted as two studies. (Rachow 2012)

**Evidence profile Xpert MTB/RIF performed ES/IS specimens**

<table>
<thead>
<tr>
<th>Outcome Description</th>
<th>Study Design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ²</td>
<td>Serious ³ -1</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Low ★★★★★</td>
</tr>
<tr>
<td>False negatives</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ² -1</td>
<td>Serious ³ -1</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Low ★★★★★</td>
</tr>
<tr>
<td>False positives</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ² -1</td>
<td>None ³</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Moderate ★★★</td>
</tr>
<tr>
<td>True negatives</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ³</td>
<td>None ³</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Moderate ★★★</td>
</tr>
</tbody>
</table>
The following considerations were not taken into account for the judgment of Limitations but are of great importance: Culture is an imperfect reference standard. In children, culture sensitivity varies between approx. 20-70%, depending on factors such as the smear status, severity of disease and age of the children, culture methods and specimen types. In order to improve the diagnostic yield 9/10 of the studies performed at least 2 but up to 6 cultures per child. Xpert MTB/RIF sensitivity may be underestimated against a culture reference standard. Evidence was not downgraded.

The quality of evidence may be lowered if there are important differences in the populations and tests studied, and the expertise of those applying them in the studies compared to the settings for which the recommendations are intended. All studies were performed at higher levels of care: higher-level referral and university hospitals. 6/10 (60%) studies included only inpatients, 4/10 studies included in- and outpatients. In the pediatric population, inpatients may present with more severe disease and have a higher likelihood of being smear positive. They are one group that might benefit from Xpert MTB/RIF, but outpatients may present very differently with less severe disease (less likely to be smear/culture positive). This population is not represented by the studies included in the review. Only one study (Rachow 2012) represents a broader group of children: approximately 20% of child TB suspects were included identified through contact tracing. Two unpublished studies (Chisti and LaCourse) included children that were severely malnourished, and TB was one differential diagnosis. (Children in the Chisti study additionally had to have cough >2 weeks and/or pneumonia on X-ray). Serious concerns about indirectness (Evidence was be downgraded by 1).

The point estimates for sensitivity range from 25 to 100%, indicating a high level of heterogeneity. 95% Confidence intervals however are overlapping. Subgroup analysis suggests an effect of smear status on overall sensitivity, with a pooled estimate for sensitivity among smear positive of 96% (95%CrI 90, 99) and smear negative of 55% (41, 69). There was no relevant heterogeneity among smear positives (pooled estimates ranged from 92-100%) but considerable heterogeneity among smear negatives (range 25-100%). Subgroup analysis for age groups (0-4 years: pooled estimate 57% 95%CrI 36.74, and 5-15 years: pooled estimate 83%, 95%CrI 68,92) and HIV status (HIV-infected, pooled estimate 75%, 95%CrI 57,88, and HIV-uninfected, pooled estimate 57%, 95%CrI 41,71) indicated differences between these groups, but these estimates are heterogeneous themselves and seem to be driven by smear status. Overall heterogeneity cannot be fully explained by the subgroup analysis; therefore inconsistency is rated as serious concern for sensitivity estimates (Evidence was downgraded by 1). Specificity estimates did not show serious heterogeneity: point estimates range from 93 to 100%. No concern.

Five of ten studies include <100 children and the yield of culture among all studies varies between 2 and 52%. The confidence interval for sensitivity is wide (crossing +/- 10%), which is of concern. Evidence was not downgraded further given the downgrading for inconsistency. No concern for Specificity.

Three unpublished studies were included. A few other studies were on-going but data was not available for inclusion in the analysis. Data included did not allow for formal assessment.
of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.

*Time to diagnosis was mainly reported for the turnaround time in the laboratory rather than from specimen collection: 6 days to 7 weeks for culture (7 studies) and 2 to 2.5 hours for Xpert MTB/RIF. Three studies describe time from enrolment/specimen collection to diagnosis for Xpert MTB/RIF (1 day) and Culture (21 days MGIT, 30days LJ).

B  
**Gastric aspirate/lavage (GLA):** 7 (1319)  
**Pooled sensitivity (GLA):** 66% (95% Crl 51, 81)  
**Pooled specificity:** 98% (95% Crl 96, 99)

Evidence profile: Xpert MTB/RIF performed on GLA specimens

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ²</td>
<td>Serious ³</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Low ⊕⊕⊕⊕</td>
</tr>
<tr>
<td>False negatives (as not having TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ²</td>
<td>Serious ³</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Low ⊕⊕⊕</td>
</tr>
<tr>
<td>False positives (as not having TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ²</td>
<td>None ³</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Moderate ⊕⊕⊙</td>
</tr>
<tr>
<td>True negatives (without TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ²</td>
<td>None ³</td>
<td>None ⁴</td>
<td>Undetected ⁵</td>
<td>Moderate ⊕⊕⊙</td>
</tr>
</tbody>
</table>

¹ The QUADAS-2 tool was used to assess the risk of bias. 3/7 of studies enrolled individuals consecutively. The Xpert MTB/RIF result is automated and considered blinded in all studies. Exclusions of enrolled patients from analysis were well explained in most studies except: in one study (Bates 2013) a number of children that were already initiated on anti-TB treatment (77/1037, ES/IS and GLA combined) were enrolled. 66 of these children were culture negative and excluded 11 were culture positive and included in the analysis.

The following considerations were not taken into account for the judgment of Limitations but are of great importance: Culture is an imperfect reference standard. In children,
culture sensitivity varies between approx. 20-70%, depending on factors such as the severity of disease and age of the children, culture methods and specimen types. In order to improve the diagnostic yield 9/10 of the studies performed at least 2 but up to 6 cultures per child. Xpert MTB/RIF sensitivity may be underestimated against a culture reference standard. The evidence was not downgraded.

The quality of evidence may be lowered if there are important differences in the populations and tests studied, and the expertise of those applying them in the studies compared to the settings for which the recommendations are intended. All studies were performed at higher levels of care: higher level referral and university hospitals. 4/7 studies included only inpatients, 1 included in- and outpatients, 2 were laboratory studies without information on the patient cohort. In the paediatric population, inpatients may present with more severe disease and have a higher likelihood of being smear positive. They are one group that might benefit from Xpert MTB/RIF, but outpatients may present very differently with less severe disease (less likely to be smear/culture positive). This population was not represented by the studies included in the review. One unpublished study (Chist) included severely malnourished children with cough >2 weeks and/or pneumonia on Xray, TB was one differential diagnosis. (Serious concerns about indirectness (Evidence was downgraded by 1).

The point estimates for sensitivity range from 40 to 100%, indicating a high level of heterogeneity. 95% Confidence intervals however were overlapping. Subgroup analysis suggests en effect of smear status on overall sensitivity, with a pooled estimate for sensitivity among smear positive of 95% (95% CrI 83, 99) and smear negative of 62% (44, 80). There was no relevant heterogeneity among smear positives (pooled estimates ranged from 92-100%) but considerable heterogeneity among smear negatives (range 36-100%). Subgroup analysis for age groups was only possible for 0-4 years-olds (pooled estimate 57% 95%CrI 38,75). Three studies were only performed in 0-5 year old children, the number of GLA's performed in children 5-15 in the remaining studies was small. Subgroup analysis for HIV status was not possible among GLA specimens. Overall heterogeneity could not be fully explained by the subgroup analysis; therefore inconsistency was rated as serious concern for sensitivity estimates (Evidence was downgraded by 1). Specificity estimates did not show serious heterogeneity: point estimates range from 93 to 100%. No concern.

Three of 7 studies include <100 children and the yield of culture among all studies varies between 2.8 and 32.7%. The confidence interval for sensitivity is wide (crossing +/- 10%), which is of concern. Evidence was not downgraded further given the downgrading for inconsistency. No concern for Specificity.

Two unpublished studies were included. A few other studies were on-going but data was not available for inclusion in the analysis. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
Summary of findings table: Xpert MTB/RIF performed on ES/IS and GLA

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Xpert MTB/RIF performed on ES/IS specimens</th>
<th>Xpert MTB/RIF performed on GLA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of results per 1000 individuals tested (95% CrI)</td>
<td>Number of results per 1000 individuals tested (95% CrI)</td>
</tr>
<tr>
<td></td>
<td>Prevalence 10 per 1000&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Prevalence 50 per 1000&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>7 (5,8)</td>
<td>33 (26, 39)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>3 (2,5)</td>
<td>17 (12, 24)</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>20 (10, 40)</td>
<td>19 (10, 38)</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>970 (950, 980)</td>
<td>931 (912, 941)</td>
</tr>
</tbody>
</table>

<sup>1</sup> Assumed numbers with Xpert MTB/RIF evaluated against culture based on the pooled estimates for sensitivity and specificity of Xpert MTB/RIF.

<sup>2</sup> Estimates of TB prevalence were provided by WHO.
Table 27: diagnostic accuracy of Xpert MTB/RIF for detection TB in children compared with a clinical reference standard, where Xpert MTB/RIF is used as a replacement test for usual practice

PICO question 2: What is the diagnostic accuracy of Xpert MTB/RIF for detection TB in children compared with a clinical reference standard, where Xpert MTB/RIF is used as a replacement test for usual practice?

Participants: Culture negative children 0-15 years with presumed TB
Setting: Mainly tertiary care referral hospitals, university hospitals.
Target condition: Pulmonary TB
Reference test: Clinical TB
Number of studies (number of participants):
- Expectorated sputum (ES) and induced sputum (IS) combined: 8* (995)
- Gastric aspirate/lavage (GLA): 3 (269)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>Serious (^1) (-1)</td>
<td>Serious (^2) (-1)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>Serious (^1) (-1)</td>
<td>Serious (^2) (-1)</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>Serious (^1) (-1)</td>
<td>Serious (^2) (-1)</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>Cross-sectional</td>
<td>Serious (^1) (-1)</td>
<td>Serious (^2) (-1)</td>
</tr>
</tbody>
</table>
The QUADAS-2 tool was used to assess the risk of bias. Six of 7 studies included in the meta-analysis (7 provided specimens for ES and 0 for IS and 3 for GLA) enrolled individuals consecutively. The Xpert MTB/RIF result is automated and considered blinded in all studies. Clinical TB is an imperfect reference standard. In this review it was defined on the basis of whether culture negative child TB suspects were initiated on anti-TB treatment or not. There is a lack of data on how well this clinical reference standard identifies true cases, on the level of over- as well as under-diagnosis. The risk of bias for the reference standard was rated as unclear for all studies. Very serious concern about limitations. (Evidence was downgraded by 1)

The quality of evidence may be lowered if there are important differences in the tests studied and the expertise of those applying them in the studies compared to the settings for which the recommendations are intended. All studies were performed at higher levels of care: higher-level referral and university hospitals. 5/7 (71%) studies included only inpatients, two studies included in- and outpatients. In the paediatric population, inpatients may present with more severe disease and have a higher likelihood of being smear positive. Approximately 20% of child TB suspects included in one study were identified through contact tracing, and this study most represents the overall paediatric population that might benefit from Xpert MTB/RIF (Rachow 2012). Two unpublished studies (Chisti and LaCourse) included children that were severely malnourished, and TB was one differential diagnosis. (Children in the Chisti study additionally had to have cough >2 weeks and/or pneumonia on Xray). Serious concerns about indirectness (Evidence was downgraded by 1).

The point estimates for sensitivity range from 0 to 35% (ES/IS) and 0 to 20% (GLA), indicating moderate heterogeneity. 95% Confidence intervals however were overlapping. No subgroup analysis was performed for Xpert MTB/RIF against a clinical reference standard. Heterogeneity may partly be defined by the differences in patient population (e.g. inclusion criteria). Inconsistency was rated as serious concern for sensitivity estimates (Evidence was downgraded by 1). Specificity estimates did not show heterogeneity; point estimates range from 99 to 100%. No concern.

The majority of cohorts included have a sample size of <60 (5/8 for ES/IS and 2/3 for GLA), and the yield of culture among all cohorts varies between 2.4 and 54.2%. The confidence interval for sensitivity for ES/IS was within the +/-10% margin, it is wide for GLA. No concern for ES/IS, serious concern for GLA. Evidence was not downgraded further given the downgrading for inconsistency. No concern for Specificity.

Two unpublished studies were included. A few other studies were on-going but data was not available for inclusion in the analysis. Data included do not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
### Summary of findings table: Xpert MTB/RIF performed on ES/IS and GLA against clinical Reference standard

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Xpert MTB/RIF performed on ES/IS specimens</th>
<th>Xpert MTB/RIF performed on GLA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of results per 1000 individuals tested (95% CrI)</td>
<td>Quality of Evidence</td>
</tr>
<tr>
<td></td>
<td>Prevalence 10 per 1000²</td>
<td>Prevalence 50 per 1000²</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>0 (0, 1)</td>
<td>2 (1, 6)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>10 (9, 10)</td>
<td>48 (44, 50)</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>0 (1, 10)</td>
<td>0 (0, 10)</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>0 (0, 1)</td>
<td>2 (1, 6)</td>
</tr>
</tbody>
</table>

1. Assumed numbers with Xpert MTB/RIF evaluated against culture based on the pooled estimates for sensitivity and specificity of Xpert MTB/RIF.
2. Estimates of TB prevalence were provided by WHO.

Page | 100
**Pico Question 3**: What is the diagnostic accuracy of Xpert MTB/RIF for detection of TB in children, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result?

**Participants**: Smear negative children 0-15 years old with presumed TB

**Setting**: Mainly tertiary care referral hospitals, university hospitals.

**Target condition**: Pulmonary TB

**Reference test**: Solid or liquid culture

**Number of studies (number of participants)**:

- Expectorated sputum (ES) and induced sputum (IS) combined: 7 (1008)
- Gastric aspirate/lavage (GLA): 6 (1204)

**Pooled sensitivity** 44% (95% CrI 41, 69)

**Pooled specificity**: 98% (95% CrI 96, 99)

**Evidence profile IS/ES and GLA**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>Serious ² (-1)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>Serious ² (-1)</td>
</tr>
<tr>
<td>False positives ⁷ (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None ³</td>
<td>Serious ² (-1)</td>
</tr>
</tbody>
</table>

---

---

Page | 102
<table>
<thead>
<tr>
<th>True negatives (individuals without TB)</th>
<th>Cross-sectional</th>
<th>None $^1$</th>
<th>Serious $^2$ (-1)</th>
<th>None $^3$</th>
<th>None $^4$</th>
<th>Undetected $^5$</th>
<th>Moderate $\text{ΘΘΘΘ}$</th>
</tr>
</thead>
</table>

1 The QUADAS-2 tool was used to assess the risk of bias for all 11 studies that were included in this estimate (7 contributing cohorts with ES/IS, 4 with GLA specimens). The majority of studies enrolled individuals consecutively. The Xpert MTB/RIF result is automated and considered blinded in all studies. Exclusions of enrolled patients from analysis were well explained in most studies except: in one study contributing GLA specimens (Bates 2013) a number of children that were already initiated on Anti TB treatment (77/1037 ES/IS and GLA combined) were enrolled. 66 of these children were culture negative and excluded 11 were culture positive and included in the analysis.

The following considerations were not taken into account for the judgment of Limitations but are of great importance: Culture is an imperfect reference standard. In children, culture sensitivity varies between approx. 20-70%, depending on factors such as the severity of disease and age of the children, culture methods and specimen types. In order to improve the diagnostic yield 2/11 of the studies (both GLA) performed at least 2 but up to 6 cultures per child. Xpert MTB/RIF sensitivity may be underestimated against a culture reference standard. Evidence was not downgraded.

2 All studies were performed at higher levels of care: higher-level referral and university hospitals. 8 studies included only inpatients, 2 studies (2 ES/IS, 1 GLA) included in- and outpatients, 2 studies (GLA) were laboratory-based with little clinical information. In the paediatric population, inpatients may present with more severe disease and have a higher likelihood of being smear positive. They are one group that might benefit from Xpert MTB/RIF, but outpatients may present very differently with less severe disease (less likely to be smear/culture positive). This population is not represented by the studies included in the review. Only one study (Rachow 2012, ES) represents a broader group of children: approximately 20% of child TB suspects were identified through contact tracing. One unpublished study (Chisti) included severely malnourished children with cough >2 weeks and/or pneumonia on Xray, TB was one differential diagnosis. Serious concerns about indirectness, evidence was downgraded by 1.

3 No study directly addressed this question by performing prior testing with microscopy and then subsequently performing Xpert MTB/RIF. Sensitivity estimates were variable across studies and specimen types and ranged from 36% to 86% (ES), 43% to 70% (IS), and 40% to 100% (GLA). The heterogeneity could not be explained and evidence was therefore downgraded 1 point. Specificity estimates did not show serious heterogeneity; point estimates range from 93 to 100%. No concern.

4 The pooled sensitivity estimate for ES/IS had a wide 95% confidence interval (\( \pm >10\% \) of point estimate), the estimate for GLA had a very wide 95% confidence interval (\( \pm >20\% \) of point estimate). Concerns were serious (ES/IS)/very serious (GLA). Evidence was not downgraded further given the downgrading for inconsistency. No concerns for specificity.

5 Two unpublished studies were included. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there is going to be considerable attention and scrutiny, reporting bias was considered to be minimal.
### Summary of findings table: Xpert MTB/RIF performed on ES/IS and GLA against culture in smear negatives

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Xpert MTB/RIF performed on ES/IS specimens</th>
<th>Xpert MTB/RIF performed on GLA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of results per 1000 individuals tested (95% CrI)</td>
<td>Number of results per 1000 individuals tested (95% CrI)</td>
</tr>
<tr>
<td></td>
<td>Prevalence 10 per 1000²</td>
<td>Prevalence 50 per 1000²</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>6 (4, 7)</td>
<td>28 (21, 35)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>5 (3, 6)</td>
<td>23 (16, 30)</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>20 (10, 40)</td>
<td>19 (10, 38)</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>970 (950, 980)</td>
<td>931 (912, 941)</td>
</tr>
</tbody>
</table>

1 Assumed numbers with Xpert MTB/RIF evaluated against culture based on the pooled estimates for sensitivity and specificity of Xpert MTB/RIF.
2 Estimates of TB prevalence were provided by WHO.
What is the additional yield of Xpert MTB/RIF over microscopy in smear-negative, culture positive TB?

### Xpert MTB/RIF on ES/IS

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of results per 1000 smear-negative (culture-positive) individuals tested (95% CrI)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Prevalence 10 per 1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smear Microscopy</td>
</tr>
<tr>
<td>True positives (patients with TB)</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>TP absolute difference</td>
<td></td>
<td>6 more</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having TB)</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>FN absolute difference</td>
<td></td>
<td>5 less</td>
</tr>
</tbody>
</table>

### Xpert MTB/RIF on GLA

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of results per 1000 smear-negative (culture-positive) individuals tested (95% CrI)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Prevalence 10 per 1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smear Microscopy</td>
</tr>
<tr>
<td>True positives (patients with TB)</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>TP absolute difference</td>
<td></td>
<td>6 more</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having TB)</td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

Page | 105
Table 29: Incremental yield of Xpert MTB/RIF compared with smear microscopy in children with culture confirmed TB

PICO question 4: What is the incremental yield of Xpert MTB/RIF compared with smear microscopy in children with culture confirmed TB?

**Participants:** Children 0-15 years with culture-confirmed TB  
**Setting:** Mainly tertiary care referral hospitals, university hospitals.  
**Target condition:** Pulmonary TB  
**Reference test:** Solid or liquid culture  
**Number of studies assessing microscopy (number of participants):** These studies are the same as the studies in PICO question 1.

<table>
<thead>
<tr>
<th>Microscopy</th>
<th>Pooled sensitivity</th>
<th>Pooled specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smear Microscopy</td>
<td>29% (95% CrI 16, 42)</td>
<td>100% (95% CrI 99, 100)</td>
</tr>
<tr>
<td>Xpert MTB/RIF</td>
<td>66% (95% CrI 52, 77)</td>
<td>98% (95% CrI 96, 99)</td>
</tr>
</tbody>
</table>

**Outcome ES/IS**

<table>
<thead>
<tr>
<th>Outcome ES/IS</th>
<th>Number of results per 1000 culture-positive individuals tested (95% CrI)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 10 per 1000</td>
<td>Prevalence 50 per 1000</td>
</tr>
<tr>
<td>Smear Microscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xpert MTB/RIF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| True positives (patients with TB) | 3 (2, 4) | 15 (8, 21) | 33 (26, 39) | 29 (16, 42) | 66 (52, 77) | Low ⊕⊕⊕\(\)
<p>| TP absolute difference | 4 more | 18 more | 37 more |</p>
<table>
<thead>
<tr>
<th><strong>False negatives</strong> (patients incorrectly classified as not having TB)</th>
<th>7 (6, 8)</th>
<th>3 (2, 5)</th>
<th>36 (29, 42)</th>
<th>17 (12, 24)</th>
<th>71 (58, 84)</th>
<th>34 (23, 48)</th>
<th>Low ☺☺☺</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FN absolute difference</strong></td>
<td>4 less</td>
<td>19 less</td>
<td>37 less</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GLA

<table>
<thead>
<tr>
<th><strong>Outcome</strong></th>
<th><strong>Smear Microscopy</strong></th>
<th><strong>Xpert MTB/RIF</strong></th>
<th><strong>Smear Microscopy</strong></th>
<th><strong>Xpert MTB/RIF</strong></th>
<th><strong>Smear Microscopy</strong></th>
<th><strong>Xpert MTB/RIF</strong></th>
<th><strong>Quality of Evidence</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True positives</strong> (patients with TB)</td>
<td>2 (1, 4)</td>
<td>7 (5, 8)</td>
<td>11 (6, 18)</td>
<td>33 (26, 41)</td>
<td>22 (12, 35)</td>
<td>66 (51, 81)</td>
<td>Low ☺☺☺</td>
</tr>
<tr>
<td><strong>TP absolute difference</strong></td>
<td>5 more</td>
<td>22 more</td>
<td>44 more</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>False negatives</strong> (patients incorrectly classified as not having TB)</td>
<td>8 (7, 9)</td>
<td>3 (2, 5)</td>
<td>39 (33, 44)</td>
<td>17 (10, 25)</td>
<td>78 (65, 88)</td>
<td>34 (19, 49)</td>
<td>Low ☺☺☺</td>
</tr>
<tr>
<td><strong>FN absolute difference</strong></td>
<td>5 less</td>
<td>22 less</td>
<td>44 less</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

1. Number of results per 1000 culture-positive individuals tested (95% CrI)
2. Quality of Evidence
Table 30: diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance in respiratory specimens in children

PICO question 5: What is the diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance in respiratory specimens in children?

**Participants:** Children 0-15 years old with presumed TB or MDR-TB  
**Setting:** Mainly tertiary care referral hospitals, university hospitals.  
**Target condition:** RIF resistance  
**Reference test:** Culture and culture-based phenotypic DST (1 studies) or MDRTBplus (2 studies)  
**Number of studies (number of participants):**

- All respiratory specimens: 3 (176)  
- **Pooled sensitivity:** 86% (95% CrI 53, 98);  
- **Pooled specificity:** 98% (95% CrI 94, 100)

**Evidence profile**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Study</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives  (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Very serious 4 (-2)</td>
</tr>
<tr>
<td>False negatives  (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Very serious 4 (-2)</td>
</tr>
<tr>
<td>False positives  (individuals incorrectly classified)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>True negatives (individuals without TB)</td>
<td>Cross-sectional</td>
<td>None ¹</td>
<td>None ²</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>--------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Cost compared with culture and phenotypic DST</td>
<td>No data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ The QUADAS-2 tool was used to assess the risk of bias. All three studies enrolled individuals consecutively. The Xpert MTB/RIF result is automated and considered blinded in all studies. No concerns

² The quality of evidence may be lowered if there are important differences in the populations and tests studied, and the expertise of those applying them in the studies compared to the settings for which the recommendations are intended. All studies were performed at tertiary level referral university hospitals, and all children included were inpatients. No concerns

³ The point estimates for sensitivity are 67%, 83%, 100%, indicating some heterogeneity. 95% Confidence intervals were wide and overlapping. Specificity estimates did not show serious heterogeneity: point estimates range from 94 to 100%. No concerns

⁴ Only 3 studies with a total of 176 children are included, of which 121 come from one study. The confidence interval for sensitivity was wide (crossing +/− 20%). Very serious concerns, evidence downgraded -2. No concern for Specificity.

⁵ One unpublished study was included. A few other studies were on-going but data was not available for inclusion in the analysis. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
Table 31: Diagnostic accuracy of Xpert MTB/RIF for detection of peripheral lymph node TB in children, where Xpert MTB/RIF is used as a replacement test for usual practice

Pico question 6: What is the diagnostic accuracy of Xpert MTB/RIF for detection of peripheral lymph node TB in children, where Xpert MTB/RIF is used as a replacement test for usual practice?

Participants: Children 0-15 years old with presumed peripheral lymph node TB
Setting: Mainly tertiary care referral hospitals, university hospitals.
Target condition: peripheral lymph node TB
Reference test: Solid or liquid culture on Fine needle aspirates/biopsies
Number of studies (number of participants): 3 (172)
Pooled sensitivity: 86% (95% CrI 65, 96);  Pooled specificity: 81% (95% CrI 54, 93)

Evidence profile

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>Serious (1)</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>Serious (1)</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>Serious (1)</td>
</tr>
<tr>
<td>True negatives</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>Serious (1)</td>
</tr>
<tr>
<td>Individuals without TB</td>
<td>sectional</td>
<td>5</td>
<td>907</td>
<td>884</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>

2 The QUADAS-2 tool was used to assess the risk of bias. Two of three studies enrolled individuals consecutively; one was a laboratory-based study with no clinical information. The Xpert MTB/RIF result is automated and considered blinded in all studies. It has to be taken into account that culture is an imperfect reference standard. **No concerns**

2 The quality of evidence may be lowered if there are important differences in the populations and tests studied, and the expertise of those applying them in the studies compared to the settings for which the recommendations are intended. All studies were performed higher-level care facilities. The patient population that received the index test probably reflects the population of intended use. **No concerns**

3 The point estimates for sensitivity are 77%, 100% and 100%, indicating some heterogeneity. 95% Confidence intervals overlap. Specificity estimates are 50%, 71%, 96%. There is no explanation for the heterogeneity. **Serious concern, evidence was downgraded -1**

4 Only 3 studies with a total of 172 children were included, of which one has 5 participants. The confidence intervals for both sensitivity and specificity are wide (crossing +/- 20%). **Very serious concerns.** Evidence was not downgraded further given the downgrading for inconsistency.

5 One unpublished study was included. A few other studies were on-going but data was not available for inclusion in the analysis. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
Pico question 7: What is the diagnostic accuracy of Xpert MTB/RIF for detection of TB meningitis in children, where Xpert MTB/RIF is used as a replacement test for usual practice?

**Participants:** Children 0-15 years old with presumed TB meningitis

**Setting:** Mainly tertiary care referral hospitals, university hospitals

**Target condition:** TB meningitis

**Reference test:** Solid or liquid culture on CSF

**Number of studies (number of participants):** 3 (51)

**Pooled sensitivity:** not enough data; **Pooled specificity:** 95% (95% CrI 81, 99)

Evidence profile and summary of findings table

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study Design</th>
<th>Factors that may decrease the quality of evidence</th>
<th>Quality of Evidence</th>
<th>Number of results per 1000 individuals tested (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Limitations</td>
<td>Indirectness</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>True positives (individuals with TB)</td>
<td>Cross-sectional</td>
<td>Not enough data</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>False negatives (individuals incorrectly classified as not having TB)</td>
<td>Cross-sectional</td>
<td>Not enough data</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>False positives (individuals incorrectly classified as having TB)</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>True negatives</td>
<td>Cross-sectional</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>(individuals without TB)</td>
<td>sectional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Three studies provided data on Xpert MTB/RIF performed on CSF for the diagnosis of TB meningitis. There was insufficient data to calculate the sensitivity of Xpert MTB/RIF. Among 2 studies, 2/6 culture positive children were Xpert MTB/RIF positive. One study did not have any positive culture or Xpert MTB/RIF.

2 The studies were all performed at higher level of care (referral hospitals), which reflects the population that would benefit from Xpert MTB/RIF performed on CSF. No concerns

3 The point estimates for specificity are 86%, 97% and 100%, indicating little heterogeneity. 95% Confidence intervals overlap. No concern

4 Only 3 studies with a total of 51 children were included. The confidence intervals for both sensitivity and specificity are wide (crossing +/- 20%). Very serious concerns. Evidence was not downgraded further given the downgrading for inconsistency.

5 One unpublished study was included. A few other studies were on-going but data was not available for inclusion in the analysis. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been found to be helpful for diagnostic test accuracy studies. However, being a new test for which there has been considerable attention and scrutiny, reporting bias was considered to be minimal.
4.4. Affordability and cost-effectiveness of Xpert MTB/RIF for the diagnosis of tuberculosis

Twelve published papers were identified which focused on comparing the costs of using Xpert MTB/RIF with follow-on tests for the diagnosis of TB and MDR-TB versus the current diagnostic algorithm for diagnosing TB and MDR-TB. The setting for most of these analyses was South Africa (10), two of these include also other countries of Sub-Saharan Africa (Botswana, Lesotho, Namibia, Swaziland and Uganda); one study covered the countries of the Former Soviet Union; and one Global Analysis which included all countries. A list of included studies for the review are given in Annex 7.

Seven (7/12) of these studies were cost analysis (CA) and 5 were cost-effectiveness analysis (CEA).

4.4.1. Summary of Cost Analysis) studies:
Of the seven cost-analyses six were from South Africa and one was a global analysis. One study adopted a societal perspective. Five focused on the screening algorithm using Xpert MTB/RIF for all individuals with signs and symptoms of TB (“all TB suspects”); and one on using Xpert MTB/RIF for HIV-positive individuals suspected of having TB who were Xpert MTB/RIF-negative initially. Two of these studies assessed the budget impact for South Africa of scaling-up at national level the use of Xpert MTB/RIF for all TB suspects. All 7 studies used non-empirical data collection, the main source of costs for Xpert MTB/RIF were the South African National Health Laboratory Service and the ‘WHO Rapid implementation of the Xpert MTB/RIF diagnostic test – Technical and Operational How-to’ practical guidance (http://whqlibdoc.who.int/publications/2011/9789241501569_eng.pdf).

Main conclusions of these analyses are:

- Cost of using Xpert MTB/RIF added 35% more to the TB budget of South Africa, but these represented only 2% of the public national health budget.
- Cost of using Xpert MTB/RIF at the level of point-of-treatment (clinic) was 51% more expensive than placement at sub-district laboratories.
- Cost of using Xpert MTB/RIF on smear-negative suspects is lower than using Xpert MTB/RIF to all TB suspects.
- Using Xpert MTB/RIF to smear-negative suspects can be cost-saving for patients.
- Cost of using Xpert MTB/RIF to Xpert MTB/RIF-negatives was less than using culture.
- At the global level, the cost of using Xpert MTB/RIF to diagnose MDR-TB and TB in HIV-positive individuals was less than the cost of conventional diagnostics. On the other hand, the cost of using Xpert MTB/RIF to diagnose TB in all TB suspects was almost five times higher than the cost of using conventional diagnostics.

4.4.2. Summary of Cost-effectiveness analysis studies:
The screening algorithm for three of these studies was for all individuals with signs and symptoms of TB, and the two other focused on HIV-positive individuals about to start ART. All five adopted a provider cost perspective. Only one study (Vassall 2011) empirically collected cost data from pilot sites of Xpert MTB/RIF; the other 4 used the costs of Xpert MTB/RIF found by Vassall 2011. All 5 CEA found that used Xpert MTB/RIF to diagnose TB and with follow-on test to diagnose MDR-TB was cost-effective compared with the current approach in the setting where the study was conducted. The current approach differs in each setting and the costs methods included are different in each study. Therefore, although all 5 CEA included incremental cost-effectiveness ratios, the comparison among them is challenging.

4.4.2.1. Cost per Xpert MTB/RIF test
The only study that directly measured cost data (full-site costing) found that the overall cost per Xpert MTB/RIF test was between US$ 23 and US$ 28 with a cartridge price of US$ 19.4; and with a cartridge price
of US$ 11.7 the cost per test is between US$ 15 and US$ 20. The other 3 cost analysis studies estimated the cost per Xpert MTB/RIF test based on market prices of the test; these found a cost per Xpert MTB/RIF test between US$ 22 and 39, the range reflected different cartridge prices and different placements of the machines.

4.4.2.2. Affordability

Two studies analysed affordability: one in South Africa and the other one at global level focusing on the 36 high TB and high MDR-TB countries. In South Africa, the increased annual costs including a national cover of Xpert MTB/RIF, represented only 2% of the public health budget.

Worldwide, the cost of Xpert MTB/RIF for all TB suspects is about 5-times higher than for WHO-recommended diagnostics. But, in high MDR-TB countries in Europe, as well as Brazil and South Africa, Xpert MTB/RIF cost is less than 10% of available TB funding and less than 0.2% of national health spending.

Worldwide, Xpert MTB/RIF costs less than WHO-recommended diagnostics for TB in people living with HIV. In high TB burden African countries, Xpert MTB/RIF cost is only 1-3% of funding approved for PEPFAR operational plans. Worldwide, cost of using Xpert MTB/RIF (with follow-on tests) for diagnosing MDR-TB is less than WHO-recommended diagnostics for MDR-TB. In high MDR-TB countries, Xpert MTB/RIF cost is only 4% of available TB funding.

4.4.3. Summary of findings

- Use of Xpert MTB/RIF to diagnose TB and MDR-TB found to be cost-effective compared with current practices for all individuals suspected of having TB and for HIV-positive individuals suspected of having TB.
- Majority of existing published studies refer to South Africa only.
- Only one study out of the 5 cost-effectiveness analysis directly measured costs of laboratory resources.
- Cost of Xpert MTB/RIF per test around US$ 15-39, depending on the cost of cartridge, placement.
- Use of Xpert MTB/RIF could be cost-saving for TB patients.
- Use of Xpert MTB/RIF more costly compared to current practices but increased costs represent small share of available TB funding, even smaller share of approved PEPFAR operational plans, and even a smaller share of national health spending.

4.4.4. Recommendations

- More directly measured costing evidence is needed for improving cost-effectiveness analysis and their recommendations.
- Need to enlarge the sample of countries where studies are done; need to include low-income countries.
- Budget impact and affordability analysis are needed when studies are done at country level.
- Need for a proper and standardized design of costing and cost-effectiveness studies.
Annex 1: Meeting participants

Sevim Ahmedov  
Senior Tuberculosis Technical Advisor  
USAID - RRB 5.07.16  
1300 Pennsylvania Ave, NW  
Washington, DC 20523  
E-mail: sahmedov@usaid.gov

Catharina Boehme  
Senior Medical Officer  
FIND  
Foundation for Innovative New Diagnostics  
Avenue de Budé 16  
1202 Geneva  
Switzerland  
E-mail: catharina.boehme@finddiagnostics.org

Lucia Barrera  
Servicio Micolobacterias (Mycobacteria Laboratory)  
INEI ANLIS Dr C G Malbran  
Velez Sarsfield 563  
1281 Buenos Aires  
ARGENTINA  
E-mail: lubarrera2000@yahoo.com.ar

Lucy Cheshire  
Tuberculosis Advocacy Consortium  
Adalyn Court  
Suite C6  
Ngong Road  
Nairobi  
Kenya  
E-mail: lucy@tbadvocacy.org

Bill Coggin  
Department of State / PEPFAR  
Office of the US Global AIDS Coordinator  
2100 Pennsylvania Ave NW  
Washington, 20522  
USA  
E-mail: CogginWL@state.gov

Gavin Churchyard  
The Aurum Institute NPC  
The Ridge  
29 Queens Road  
Parktown  
Johannesburg  
South Africa  
Email: gchurchyard@auruminstitute.org

Daniela Maria Cirillo  
Head, Emerging Bacterial Pathogens Unit  
San Raffaele Scientific Institute  
Via Olgettina 58  
20132 Milan Italy  
E-mail: cirillo.daniela@hsr.it

Frank Cobelens  
Amsterdam Institute of Global Health and Development  
& Department of Global Health  
Academic Medical Center  
Amsterdam, Netherlands  
E-mail: f.cobelens@aighd.org

Colleen Daniels  
Director TB/HIV  
Treatment Action Group  
New York, USA  
E-mail: colleen.daniels@treatmentactiongroup.org

Claudia Denkinger  
McGill University  
4075 Rue Cartier  
Montreal, Quebec, Canada  
ddenking@bidmc.harvard.edu

Anne Detjen  
Technical Consultant, The Union North America Office, International Union Against Tuberculosis and Lung Disease (The Union)  
61 Broadway, Suite 1720  
New York, NY 10006, U.S.A  
E-mail: adetjen@theunion.org
Mildred Fernando  
Management Sciences for Health, Inc.  
181 Floro Subdivision, Malhacan, Meycauayan, Bulacan  
Philippines  
Email: mildspirit@gmail.com

Nazir Ismail  
Head of Centre for Tuberculosis (incorporating the NTBRL)  
National Institute for Communicable Diseases  
1 Modderfontein Road, Sandringham  
Johannesburg  
South Africa  
Email: naziri@nicd.ac.za

Moses Joloba  
Department of Medical Microbiology  
Makerere University College of Health Sciences  
2nd Floor Pathology/Microbiology BL Room C31, Upper Mulago Hill Road  
256 Kampala  
Uganda  
Email: moses.joloba@case.edu

Anna Mandalakas  
Director, Global Tuberculosis and Mycobacteriology Program  
The Tuberculosis Initiative Texas Children’s Hospital  
1102 Bates St, FC 630  
Houston, Texas 77030, USA  
Email: anna.mandalakas@bcm.edu

Andrea Pantoja  
Ankenhofstrasse 23  
Oberengstringen, 8102  
Switzerland  
Email: andreapantoja@gmail.com

Holger Schünemann  
Chair, Department of Clinical Epidemiology & Biostatistics  
Department of Clinical Epidemiology & Biostatistics  
McMaster University Health Sciences Centre, Room 2C10B  
1200 Main Street West  
Hamilton, ON L8N 3Z5, Canada  
E-mail: schuneh@mcmaster.ca

Moorine Penniah Sekadde  
National Tuberculosis and Leprosy Programme  
Kampala  
Uganda  
Email: moorine.sekadde@gmail.com

Thomas M. Shinnick  
Associate Director for Global Laboratory Activities  
Centres for Disease Control and Prevention  
1600 Clifton Road  
MS-G35, NE  
30333 Atlanta, GA  
United States of America  
E-mail: tms1@cdc.gov

Karen Steingart  
Editor, Cochrane Infectious Diseases Group  
USA  
Email: karen.steingart@gmail.com

Sabira Tahseen  
National Coordinator (NRL) National TB Control Programme  
Islamabad  
Pakistan  
Email: sabira.tahseen@gmail.com

Maarten van Cleeff  
KNCV Tuberculosis Foundation  
Parkstraat 17  
2514 JD The Hague  
The Netherlands  
Email: vancleeffm@kncvtbc.nl

Francis Varaine  
 Médecins sans Frontières  
8 rue St Sabin  
75011 Paris  
France  
Email: francis.varaine@paris.msf.org
WHO Staff

Haileyesus Getahun  
E-mail: getahunh@who.int

Christopher Gilpin  
E-mail: gilpinc@who.int

Jean Iragena  
E-mail: iragenaj@who.int

Knut Lonnroth  
Email: lonnrothk@who.int

Lisa Nelson  
Email: nelsonl@who.int

Fuad Mirzayev  
E-mail: mirzayevf@who.int

Wayne van Gemert  
Email: vangemertw@who.int

Karin Weyer  
E-mail: weyerk@who.int

Matteo Zignol  
E-mail: zignolm@who.int
Annex 2: Meeting agenda

WHO Policy Guidance

on the utility of the Xpert MTB/RIF assay for the diagnosis of pulmonary and extra-pulmonary TB in adults and children

- EXPERT GROUP MEETING -

Date: 20-21 May 2012
Venue: Les Pensierès, Veyrier-du-Lac, France

BACKGROUND

The World Health Organization (WHO) in 2011 issued a policy statement recommending the Xpert MTB/RIF assay (Cepheid, Ca., USA) for the diagnosis of tuberculosis (TB) and Rifampicin (RIF) resistance as a proxy for multi drug-resistant TB (MDR-TB). Xpert MTB/RIF should be used as an initial diagnostic test in individuals suspected of MDR or HIV-associated TB. It should be used as an add-on test to smear microscopy in settings where MDR or HIV are of lesser concern, especially in smear-negative specimens. Generalizing from adult data, the recommendation includes the use of Xpert MTB/RIF in children, acknowledging the difficulties in the microbiological diagnosis of childhood TB.

Since that time, additional studies on Xpert MTB/RIF for both pulmonary and extrapulmonary TB have been published or are being performed. An update of the 2011 Xpert MTB/RIF policy guidance is planned for 2013.

In adults, Xpert MTB/RIF has shown a high sensitivity (99%) in smear and culture positive individuals with pulmonary TB. HIV co-infection did not significantly affect Xpert MTB/RIF performance to detect smear-negative/culture-positive patients (Sensitivity 86%). The ability of Xpert MTB/RIF to detect TB in smear-negative individuals is encouraging and suggests that it may be a valuable TB diagnostic tool in children. There are emerging data available on the utility of Xpert MTB/RIF in children. Some of these studies have shown a better performance of Xpert MTB/RIF compared to microscopy and similar performance when compared to liquid culture.

WHO has commissioned three systematic reviews which will include available data regarding the use of Xpert MTB/RIF for the diagnosis of pulmonary and extrapulmonary TB in adults and children and will provide evidence to inform future policy recommendations.

In accordance with current WHO standards for evidence assessment in the formulation of policy recommendations, WHO engages in a systematic, transparent process using the GRADE approach (http://www.gradeworkinggroup.org). GRADE provides a structured framework for evaluating diagnostic test accuracy and the patient/public health impact of new diagnostic tests.
MEETING OBJECTIVES

- To review the evidence base and evaluate data from an updated systematic review to assess the accuracy of Xpert MTB/RIF assay for the diagnosis of pulmonary TB and rifampicin resistance in adults;
- To review the evidence base and evaluate data from a systematic review on the accuracy of the Xpert MTB/RIF assay for the diagnosis of tuberculosis on non-respiratory samples;
- To review the evidence base and evaluate data from a systematic review on the accuracy of the Xpert MTB/RIF assay for the diagnosis of tuberculosis and rifampicin resistance in children;
- To review the evidence on the cost-effectiveness and affordability of the Xpert MTB/RIF in different epidemiological and resource settings;
- To outline issues to be addressed by WHO in subsequent policy recommendations.

EXPECTED OUTCOMES

- Evidence-based recommendations on the accuracy of Xpert MTB/RIF assay for the diagnosis of pulmonary TB and rifampicin resistance in adults
- Evidence-based recommendations on the accuracy of the Xpert MTB/RIF assay for the diagnosis of tuberculosis on non-respiratory samples
- Evidence-based recommendations on the accuracy of the Xpert MTB/RIF assay for the diagnosis of tuberculosis and rifampicin resistance in children
- Evidence-based recommendations on the cost-effectiveness and affordability of the Xpert MTB/RIF in different epidemiological and resource settings;
- Consensus on issues to be addressed in development of subsequent WHO policy recommendations.
PROVISIONAL AGENDA

Monday, 20 May 2013: Expert Group Meeting on Xpert MTB/RIF DAY 1

Chairs: H Schünemann

WHO Secretariat: C Gilpin

Rapporteur: W van Gemert

<table>
<thead>
<tr>
<th>Opening session</th>
<th>Time</th>
<th>Session 1: Xpert MTB/RIF for diagnosis of pulmonary TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 – 09:10</td>
<td>Welcome</td>
<td>K Weyer</td>
</tr>
<tr>
<td>09:10 – 09:20</td>
<td>Meeting scope and objectives</td>
<td>C Gilpin</td>
</tr>
<tr>
<td>09:20 – 09:30</td>
<td>Declaration of Interest by Expert Group members</td>
<td>C Gilpin</td>
</tr>
<tr>
<td>09:30 – 10:00</td>
<td>Grading quality of evidence and strength of recommendations: Brief overview of GRADE</td>
<td>H Schünemann</td>
</tr>
<tr>
<td>10:00 – 10:15</td>
<td>Questions</td>
<td></td>
</tr>
<tr>
<td>10:15 – 10:45</td>
<td>BREAK</td>
<td></td>
</tr>
<tr>
<td>10:45 – 11:00</td>
<td>Current WHO policy guidance on Xpert MTB/RIF</td>
<td>Wayne van Gemert</td>
</tr>
<tr>
<td>11:00 – 11:15</td>
<td>PICO questions and rating outcomes</td>
<td>Chris Gilpin</td>
</tr>
<tr>
<td>11:15 – 12:00</td>
<td>Updated systematic review: Xpert MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults</td>
<td>Karen Steingart</td>
</tr>
<tr>
<td>12:00 – 13:00</td>
<td>Discussion</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>1. What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults, where Xpert MTB/RIF is used as a replacement test for smear microscopy?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. What is the diagnostic accuracy of Xpert MTB/RIF for detection of smear-positive pulmonary TB in adults?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. What is the diagnostic accuracy of Xpert MTB/RIF for detection of smear-negative (culture-positive) pulmonary TB in adults?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in people living with HIV (adults)?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. What is the diagnostic accuracy of Xpert MTB/RIF for detection of pulmonary TB in adults without HIV infection?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. What is the diagnostic accuracy of Xpert MTB/RIF for detection</td>
<td></td>
</tr>
</tbody>
</table>
of rifampicin resistance, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing?

<table>
<thead>
<tr>
<th>Time</th>
<th>Session 2: Xpert MTB/RIF for diagnosis of extra-pulmonary TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00 – 14:00</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:00 – 15:00</td>
<td>Draft recommendations: Xpert MTB/RIF in pulmonary TB</td>
</tr>
<tr>
<td>15:00 – 15:45</td>
<td>Systematic review: Xpert MTB/RIF assay for tuberculosis on non-respiratory samples</td>
</tr>
<tr>
<td>15:45 – 16:00</td>
<td>Break</td>
</tr>
<tr>
<td>16:00 – 17:00</td>
<td>Discussion&lt;br&gt;1. What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in lymph node fluid and tissue, where Xpert MTB/RIF is used as a replacement test for usual practice? &lt;br&gt;2. What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in pleural fluid, where Xpert MTB/RIF is used as a replacement test for usual practice? &lt;br&gt;3. What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in cerebrospinal fluid (CSF), where Xpert MTB/RIF is used as a replacement test for usual practice? &lt;br&gt;4. What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in gastric fluid, where Xpert MTB/RIF is used as a replacement test for usual practice? &lt;br&gt;5. What is the diagnostic accuracy of Xpert MTB/RIF for TB detection in tissue samples, where Xpert MTB/RIF is used as a replacement test for usual practice? &lt;br&gt;6. What is the diagnostic accuracy of Xpert MTB/RIF for rifampicin resistance detection in non-respiratory specimens, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing?</td>
</tr>
<tr>
<td>17:00 – 18:00</td>
<td>Draft recommendations: Xpert MTB/RIF in extrapulmonary TB</td>
</tr>
<tr>
<td>19:00–</td>
<td>Dinner reception</td>
</tr>
</tbody>
</table>

17:00 – 18:00 Draft recommendations: Xpert MTB/RIF in extrapulmonary TB

19:00– Dinner reception
**Tuesday, 21 May 2013: Expert Group Meeting on Xpert MTB/RIF DAY 2**

**Chairs:** H Schünemann  
**WHO Secretariat:** C Gilpin  
**Rapporteur:** W van Gemert

### Session 3: Xpert MTB/RIF in paediatric TB

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 – 9:45</td>
<td>Systematic review: Xpert MTB/RIF assay for tuberculosis and rifampicin resistance in children</td>
<td>A. Detjen</td>
</tr>
</tbody>
</table>
| 9:45 – 10:45 | Discussion  
What is the diagnostic accuracy of Xpert MTB/RIF for detection TB in children compared with culture, where Xpert MTB/RIF is used as a replacement test for usual practice?  
2. What is the diagnostic accuracy of Xpert MTB/RIF for detection TB in children compared with a combined clinical and laboratory reference standard, where Xpert MTB/RIF is used as a replacement test for usual practice?  
3. What is the diagnostic accuracy of Xpert MTB/RIF for detection of TB in children, where Xpert MTB/RIF is used as an add-on test following a negative smear microscopy result?  
4. What is the diagnostic accuracy of Xpert MTB/RIF compared with smear microscopy for detection of TB in children?  
5. What is the diagnostic accuracy of Xpert MTB/RIF for detection of rifampicin resistance in children, where Xpert MTB/RIF is used as an initial test replacing phenotypic culture-based drug susceptibility testing?  
6. What is the diagnostic accuracy of Xpert MTB/RIF for detection of peripheral lymph node TB in children, where Xpert MTB/RIF is used as a replacement test for usual practice?  
7. What is the diagnostic accuracy of Xpert MTB/RIF for detection of TB meningitis in children, where Xpert MTB/RIF is used as a replacement test for usual practice? | All |
| 10:45 – 11:00 | BREAK | |
| 11:00 – 12:00 | Draft recommendations:  
Xpert MTB/RIF in paediatric TB | H Schünemann |
| 12:00 – 13:00 | LUNCH | |

### Session 4: Cost-effectiveness and affordability of Xpert MTB/RIF MTB/RIF

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00 – 13:45</td>
<td>Review: Cost-effectiveness and resource implications for Xpert MTB/RIF implementation</td>
<td>A Pantoja</td>
</tr>
<tr>
<td>Time</td>
<td>Activity</td>
<td>Presenter</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>13:45 – 14:45</td>
<td>Discussion</td>
<td>All</td>
</tr>
<tr>
<td>14:45 – 15:30</td>
<td>Draft recommendations:</td>
<td>H Schünemann</td>
</tr>
<tr>
<td>15:30 – 15:45</td>
<td>BREAK</td>
<td></td>
</tr>
<tr>
<td><strong>Session 5:</strong></td>
<td>Review of GRADE summaries and formulation of final recommendations</td>
<td></td>
</tr>
<tr>
<td>15:45 – 16:15</td>
<td>Discussion on issues to be addressed by WHO in subsequent policy recommendations on Xpert MTB/RIF</td>
<td>All</td>
</tr>
<tr>
<td>16:15 – 16:45</td>
<td>Review of GRADE process and summaries</td>
<td>H Schünemann</td>
</tr>
<tr>
<td>16:45 – 17:30</td>
<td>Final recommendations</td>
<td>H Schünemann</td>
</tr>
</tbody>
</table>
Annex 3: Declarations of Interest

None declared

Lucia Barrera, Lucy Cheshire, Colleen Daniels, Holger Schünemann (Chair), Moorine Sekadde, Sabira Tahseen, Maarten van Cleeff

Declared, insignificant

Daniela Cirillo - Meeting participation sponsored by Cepheid, research grant support from FIND and ST microelectronics.

Mildred Fernando - MSH contributed approx. USD 80 for her participation in the Expert Group meeting. Xpert MTB/RIF MTB/RIF may have been used to determine if her TB had relapsed.

Thomas Shinnick - US Government funding for participation in the meeting

Francis Varaine - Leader of the MSF working group on TB and required to defend positions related to TB diagnostics

Declared, significant (observer status)

Catharina Boehme - FIND employment, co-funding agreement with Cepheid around the development of Xpert MTB/RIF MTB/RIF

Bill Coggin - US Government employee funding the cartridge price reduction

Sevim Ahmedov - US Government employee funding the cartridge price reduction

Moses Joloba - FIND demonstration site

Gavin Churchyard - Employee of Aurum Institute which received funding from BMGF for trials on the Xpert MTB/RIF assay for diagnosis of TB in South Africa and for evaluating impact and cost-effectiveness in routine roll-out.

Frank Cobelens - Consultancy on research for roll-out and scale-up MTB/RIF (BMGF) and cost-effectiveness analysis of Xpert MTB/RIF MTB/RIF (FIND)

Nazir Ismail - Employee NHLS rolling out Xpert MTB/RIF in South Africa, SA purchased >1 million cartridges. No financial gain

Karen Steingart - Systematic reviewer (pulmonary TB)

Claudia Denkinger - Systematic reviewer (extrapulmonary TB)

Anne Detjen - Systematic reviewer (paediatric TB)

Anna Mandalakas - Systematic reviewer (paediatric TB)

Andrea Pantoja - Reviewer – affordability and cost-effectiveness review
Annex 4: Selection of studies evaluating Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults

Flow of studies in the initial literature searches

- 139 records identified through database searching 25 September 2011
- 5 additional records identified through other sources
- 137 records screened after duplicates removed
- 77 records excluded based on title and abstract
- 60 full-text articles assessed for eligibility
- 42 full-text articles excluded, with reasons:
  - Abstract (10)
  - Case control (1)
  - Correspondence (1)
  - Cost-effectiveness analysis (1)
  - Could not obtain (1)
  - Duplicate data (1)
- 81 records identified through database searching 15 December 2011; no new records were identified
- 18 studies included in qualitative synthesis
- 15 studies included in meta-analysis for TB detection; 11 studies included in meta-analysis for rifampicin resistance detection
Flow of studies in the updated literature search

343 records identified through database searching 7 February 2013

2 additional records identified through other sources

130 records screened after duplicates removed

76 records excluded based on title and abstract

54 full-text articles assessed for eligibility

45 full-text articles excluded, with reasons:
- Extrapulmonary TB (7)
- Paediatric TB (6)
- Correspondence (5)
- Duplicate data (4)
- Impact study (4)
- Study did not enrol patients suspected of TB (3)
- Technical (3)
- Case control study (2)

9 studies included in qualitative synthesis

7 studies included in quantitative synthesis (meta-analysis)
Included and excluded studies from the reviews

**Included studies (Published)**


Unpublished studies


Excluded studies:
Excluded studies identified from electronic searches 25 September 2011, 15 December 2011, and 7 February 2013 (studies may have appeared in more than one search) (Reasons for exclusion in parenthesis after reference)


   3. Armand S, Vanhuls P, Delcroix G, Courcel R, Lemaitre N. Comparison of the Xpert MTB/RIF test with an IS6110-TaqMan real-time PCR assay for direct detection of Mycobacterium tuberculosis in respiratory and nonrespiratory specimens. Journal of Clinical Microbiology 2011;49(5):1772-6. This was a case-control study that compared Xpert MTB/RIF assay with an in-house IS6110-based real-time PCR using TaqMan probes (IS6110-TaqMan assay) for TB detection


33. Lawn SD, Nicol MP. Xpert MTB/RIF® assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. Future Microbiology 2011;6(9):1067-82. Narrative review


43. Melzer Mark. An automated molecular test for Mycobacterium tuberculosis and resistance to rifampin (Xpert MTB/RIF) is sensitive and can be carried out in less than 2 h. Evidence Based Medicine 2011;16:19. Editorial and comment


53. Ntinginya EN. Squire SB, Millington KA, Mtafya B, Saathoff E, Heinrich N, et al. Performance of the Xpert MTB/RIF assay in an active case-finding strategy: a pilot study from Tanzania. Int J Tuberc Lung Dis 16:1468-70. This study included both adults and children. The study used an active case finding strategy involving previously known TB cases
and identified five additional culture-confirmed TB cases (5/219). Xpert MTB/RIF showed a positive result in all 5 culture-confirmed TB cases (sensitivity = 100%). We considered the study design to be different from a diagnostic test accuracy study and therefore did not include this study in the review.


57. Peter JG, Theron G, Muchinga TE, Govender U, Dheda K. The diagnostic accuracy of urine-based Xpert MTB/RIF in HIV-infected hospitalized patients who are smear-negative or sputum scarce. PLoS One 2012;7:e39966. This study evaluated Xpert MTB/RIF for the diagnosis of extrapulmonary TB.


74. Van Rie A, Page-Shipp L, Scott L, Sanne I, Stevens W. Xpert MTB/RIF(*) MTB/RIF for point-of-care diagnosis of TB in high-HIV burden, resource-limited countries: hype or hope? Xpert MTB/RIF Review of Molecular Diagnostics 2010;10(7):937-46. This was a review that covered technical details of Xpert MTB/RIF and the test's potential value as a point-of-care test


Annex 5: Selection of studies evaluating Xpert MTB/RIF for extrapulmonary tuberculosis in adults and children

Flow of studies included in the review

Potentially relevant citations identified from electronic databases: 194

Excluded screen 1: 143
Reason: Not relevant based on assessment of title and abstract

Full papers retrieved for more detailed evaluation: 51

Excluded screen 2: 36
Reasons:
• < 10 samples per EPTB type: 10
• Specificity results lacking: 0
• Abstract: 12
• Cost effectiveness: 0
• Does not contain EPTB samples: 5
• Duplicate data: 1
• Editorial or comment: 2

Unpublished data/papers added: 8
• About to be published: 6
• Ongoing: 2

Unpublished data/papers excluded: 1
Reason: Did not include any sample type that contributed to the subgroups analyzed

Papers (studies) included in the systematic review: 22
List of included studies

Published studies


Unpublished

2. Comparison of rapid tools, including Xpert MTB/RIF F assay, for the diagnosis of pleural tuberculosis (interim analysis).
3. Comparison of rapid tools, including Xpert MTB/RIF F assay, for the diagnosis of pleural tuberculosis (interim analysis).
7. Xpert MTB/RIF for the diagnosis of extrapulmonary tuberculosis (interim analysis).

List of excluded studies and reasons for exclusion


33. Theron, G. 2012. Cycle-threshold (CT) values of an automated TB-specific PCR platform (Xpert MTB/RIF) as a predictor of smear status and grade. Tropical Medicine & International Health 17:24-24. Does not contain non-respiratory samples


Annex 6: Selection of studies evaluating Xpert MTB/RIF for pulmonary and extrapulmonary tuberculosis and rifampicin resistance in children

Flow diagram of studies in the review

Inclusion and exclusion criteria:

1. Studies evaluating the Xpert MTB/RIF assay for the diagnosis of tuberculosis in children
2. Studies involving pediatric populations
3. Studies published in English
4. Studies published between January 2011 and January 2013

List of included studies:


**Unpublished studies**

2. Xpert for the diagnosis of peripheral lymphnode TB in children. Ongoing study May 2013

**Excluded studies**

1. Armand S, Vanhuls P, Delcroix G, Courcol R, Lemaitre N. Comparison of the Xpert MTB/RIF test with an IS6110-TaqMan real-time PCR assay for direct detection of Mycobacterium tuberculosis in respiratory and nonrespiratory specimens. Journal of Clinical Microbiology 2011;49(5):1772-6. This was a case-control study that compared Xpert MTB/RIF assay with an in-house IS6110-based real-time PCR using TaqMan probes (IS6110-TaqMan assay) for TB detection
18. Ntinginya EN. Squire SB, Millington KA, Mtasya B, Saathoff E, Heinrich N, et al. Performance of the Xpert(R) MTB/RIF assay in an active case-finding strategy: a pilot study from Tanzania. Int J Tuberc Lung Dis 16:1468-70. This study included both adults and children. The study used an active case finding strategy involving previously known TB cases. The study design was considered to be different from a diagnostic test accuracy study and therefore was not included in the review


