Systematic reviews of diagnostic test accuracy studies
McGill Summer Institute, Montreal June 2017

Karen R Steingart, MD, MPH
Cochrane Infectious Diseases Group
Liverpool School of Tropical Medicine

karen.steingart@gmail.com
Declarations of Interest

• Editor, Cochrane Infectious Diseases Group and Cochrane Diagnostic Test Accuracy Editorial Team

• No financial disclosures
Overview

• Systematic reviews (DTA systematic reviews): what, why, and how
• Formulate a review question, PICO
• Describe quality assessment of DTA studies
• Describe methods for analyzing data, including meta-analysis
• Interpret results

• DTA = diagnostic test accuracy
A systematic review starts with a **clearly formulated question** and uses **systematic** and **explicit methods** to identify, select, and **critically appraise relevant research**, and to **collect** and **analyse** data from the studies that are included in the review.

Egg slide adapted from Madhu Pai
A meta-analysis is the use of statistical techniques in a systematic review to integrate the results of included studies and analyse data from the studies that are included in the review.

Not all systematic reviews include a meta-analysis.
Why undertake a systematic review?
It is surely a great criticism of our profession that we have not organized a critical summary, by specialty and subspecialty, adapted periodically, of all relevant randomised controlled trials”

• The growing amount of information is challenging to manage
• Decision makers need to integrate critical pieces of information
• Systematic reviews are an efficient scientific technique

Mulrow BMJ 1994
Systematic reviews of diagnostic test accuracy studies - terminology

- **Index test(s)** - new test

- **Target condition** - is a particular disease that the index test is intended to identify

- **Reference standard (gold standard)** - an agreed-upon, accurate method for identifying patients who have the target condition
Diagnostic test accuracy - definition

• Diagnosis asks, “does this person have this disease (more generally, this target condition) at this point in time?”

• Diagnostic test accuracy refers to the ability of a test to distinguish between patients with disease and those without disease
Reference standard - a few comments

- The accuracy of an index test cannot be measured without a reference standard.
- There should be general agreement that the reference standard is more accurate than the index test.
- There may be more than one reference standard.
- The reference standard may comprise several pieces of information.
- The most accurate reference standard may not be feasible or ethical.
Diagnostic test accuracy – what are we measuring?

- Determine agreement between the results of the index test and the reference standard
  - Usually estimate sensitivity and specificity
  - Requires a 2×2 table

<table>
<thead>
<tr>
<th>True positive</th>
<th>False positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>False negative</td>
<td>True negative</td>
</tr>
</tbody>
</table>
### 2x2 Table - Sensitivity and Specificity

<table>
<thead>
<tr>
<th>Disease (Reference standard)</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Index test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>-</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

**Sensitivity**
- $\frac{TP}{TP+FN}$

**Specificity**
- $\frac{TN}{TN+FP}$
Diagnostic accuracy cross-sectional study design

1. Series of patients
2. Index test
3. Reference standard
4. Blinded cross-classification

P Bossuyt http://srdda.cochrane.org/presentations
Tests do not make patients better

Diagnostic accuracy ≠ patient outcome
<table>
<thead>
<tr>
<th>Pathway component and mechanism</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Diagnostic test delivered</td>
<td></td>
</tr>
<tr>
<td>Timing of test</td>
<td>Speed with which a test is performed within the management strategy</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Completion of test process. Reasons for non-completion are: patient acceptability (patient’s refusal to have test), test was contraindicated (clinical reason not to administer test), and technical failure (ability of diagnostic equipment to produce data)</td>
</tr>
<tr>
<td>Test process</td>
<td>Patients’ interaction with test procedure, potentially causing physical or psychological harms or benefits</td>
</tr>
<tr>
<td>(2) Test result produced</td>
<td></td>
</tr>
<tr>
<td>Interpretability</td>
<td>Degree to which test data can be used to inform a diagnostic classification</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Ability of a test to distinguish between patients who have disease and those who do not</td>
</tr>
<tr>
<td>Timing of results</td>
<td>Speed with which test results are available</td>
</tr>
<tr>
<td>(3) Diagnosis made</td>
<td></td>
</tr>
<tr>
<td>Timing of diagnosis</td>
<td>Speed with which a diagnostic decision is made</td>
</tr>
<tr>
<td>Diagnostic yield</td>
<td>Degree to which the test contributes to a patient diagnosis in any form, including: provision of a definitive diagnosis, confirmation of a suspected diagnosis, ruling out a working diagnosis, and distinguishing between alternative diagnoses with different treatment implications. Diagnostic yield is different from accuracy because it also incorporates any other information used by a doctor to make a diagnosis (such as previous test results)</td>
</tr>
<tr>
<td>Diagnostic confidence</td>
<td>Degree of confidence that doctors and patients have in the validity or applicability of a test result</td>
</tr>
<tr>
<td>(4) Management decided</td>
<td></td>
</tr>
<tr>
<td>Therapeutic yield</td>
<td>Degree to which diagnostic decisions affect treatment plans</td>
</tr>
<tr>
<td>Therapeutic confidence</td>
<td>Certainty with which doctors and patients pursue a course of treatment</td>
</tr>
<tr>
<td>(5) Treatment implemented</td>
<td></td>
</tr>
<tr>
<td>Timing of treatment</td>
<td>Speed with which patients receive treatment</td>
</tr>
<tr>
<td>Treatment efficacy</td>
<td>Ability of the treatment intervention to improve patient outcomes</td>
</tr>
<tr>
<td>Adherence</td>
<td>Extent to which patients participate in the management plan, as advised by their doctor, to attain therapeutic goal</td>
</tr>
</tbody>
</table>

di Ruffano, BMJ 2012
Road map for diagnostic accuracy reviews

A “road map” for systematic reviews of diagnostic test evaluations

1. Define a focused diagnostic review question (Patient/Disease, Index test, Reference standard, and Outcomes)
2. Search directly or via reference manager software: avoid language restrictions at this stage, involve a librarian
3. Use sensitive filters for diagnostic studies* if the number of citations is too large
4. Use Medline, Reference Manager, ProCite
5. Need clear inclusion and exclusion criteria
6. Software suggestions: Medline, Reference Manager, ProCite
7. Reviewers meet and resolve disagreements on citations they do not agree on
8. The final number (N) selected after this process is ready for second screen (review of full text articles)
9. Articles considered eligible after full text review (by 2 reviewers) is the final set of studies for inclusion (N)
10. Studies included in the final analysis (n)
11. Each article gets a unique ID number
12. Reviewers meet and resolve disagreements on data
13. The final data after this process are ready for data entry
14. Enter data into database manager software
15. Conduct authors for missing data (emailed may be more effective than letters)
16. Software suggestions: Access, Excel
17. Explanatory heterogeneity: graphical methods, subgroup analyses, and meta-regression
18. Use QUOROM** or MOOSE** as general guides for report writing (acknowledging that they are not meant for diagnostic reviews)
19. Collect outcomes on TR, PP, NP, and TN, or raw ROC data
20. Enter data and analyze using software* Tabulate study characteristics
21. Forest and ROC plots of SE and SP
22. Look for correlation between TPR and FPR
23. Search for threshold effect
24. Perform SROC analysis
25. Pool measures like LR and DOR only if appropriate
26. Search for heterogeneity, and reasons for heterogeneity* Consider subgroup and sensitivity analyses
27. Consider blinded data extraction (blinding author names, etc)
28. Quality criteria: patient spectrum, blinding, verification, sampling, appropriate reference standard, and other criteria
29. You made it! Celebrate!!!


©2004 by BMJ Publishing Group Ltd
10 steps in performing a systematic review of diagnostic test accuracy studies

1. Define the review question and selection criteria
2. Write a protocol (plan the methods)
3. Search for studies
4. Apply selection criteria
5. Collect the data
6. Assess methodological quality
7. Analyze the data
8. Interpret results
9. Draw conclusions
10. Improve and update the review
Define the review question and selection (inclusion and exclusion) criteria
The index test was Alere Determine™ TB LAM Ag test. We evaluated the test at two different cut-off values for positivity (grade 1 and grade 2) based on the original manufacturer reference card.
Begin with a well-framed question - 2

- How do the participants present?

- What is the purpose of the test (such as screening, diagnosis, prognosis, monitoring)?

- What is the role of the new test (replacement, triage, add-on, combination with other tests?)
Role of tests

• Replacement: new test replaces existing test

• Triage: new test goes before existing test and only patients with a particular result go on to receive further testing

• Add-on: new test follows existing test

• Parallel: new test is used in combination with existing test(s)

Bossuyt et al. BMJ 2006
Most situations are replacement

Why would you want to replace an existing test with a new test?

• More accurate
• More rapid
• Less invasive
• Technically easier to do
• Easier to interpret
• Other: storage, type of specimen, environment, infection control
PICO for systematic review of diagnostic test accuracy

- Participants
- Index test(s)
- Purpose of test
- Role of testing
- Target condition
- Reference standard
- Comparator (may or may not be included)
- Outcomes (usually sensitivity and specificity)
Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

Maunank Shah et al
The Cochrane Library
2016
Open access
Professor Stephen Lawn
Urine lateral-flow lipoarabinomannan (LF-LAM) 
Background - 1

- LAM: 17.5 kilodalton structural component of mycobacterial cell walls
- Detectable in urine of patients with TB
- Meta-analysis of ELISA-LAM*
  - Increased sensitivity in HIV-positive compared with HIV-negative patients
  - Pooled sensitivity 56% (40-71%)
  - Pooled specificity 95% (77-99%)
  - Higher sensitivity with more severe immunosuppression

Urine LF-LAM is a new diagnostic test that may overcome limitations of other approaches:

- Point of care, lateral flow format
- Fast (< 20 minutes)
- No equipment needed
- Low cost (~$3.50)
- Accuracy has varied across published studies
Where’s PICO?
Objectives

• To assess the accuracy of LF-LAM for the diagnosis of active TB disease in HIV-positive adults who have signs and symptoms suggestive of TB (TB diagnosis)

• To assess the accuracy of LF-LAM as a screening test for active TB disease in HIV-positive adults irrespective of signs and symptoms suggestive of TB (TB screening)
Secondary objectives

• To compare diagnostic accuracy of LF-LAM and existing tests, sputum smear microscopy or sputum Xpert® MTB/RIF, as well as determine the diagnostic accuracy of LF-LAM when added to existing tests

• To investigate heterogeneity of test accuracy in the included studies (CD4 count and clinical setting)
Selection criteria - 1

- We included randomized controlled trials, cross-sectional studies, and cohort studies that determined LF-LAM accuracy for TB against a microbiological reference standard (culture or nucleic acid amplification test from any body site)
- We included abstracts with sufficient data
- We excluded case-control studies
Selection criteria - 2

• Participants were HIV positive adults
• We required studies to diagnose TB using at least one of the following two reference standards
  - Microbiological reference standard
  - Composite reference standard with (1) a positive culture, or (2) a positive NAAT, or (3) a positive smear, or (4) a clinical decision to start TB treatment
PICO for LF-LAM ?

• Participants =
• Presentation =
• Index test(s) =
• Purpose of test =
• Role of testing =
• Target condition =
• Reference standard =
• Comparator =
• Outcomes =
The medical literature can be compared to a jungle. It is fast growing, full of deadwood, sprinkled with hidden treasure and infested with spiders and snakes. Morgan. Can Med Assoc J, 134, Jan 15, 1986

Assess methodological quality (bias and applicability)
Nothing will corrode public trust more than a creeping awareness that scientists are unable to live up to the standards that they have set for themselves.
What is bias?

- Bias is **any process at any stage** of inference tending to produce **results** that **differ systematically** from the **true** values. *Murphy EA. The logic of medicine 1976*

- This is the tendency of some (poor) study designs **systematically** to produce results that are **better** (rarely if ever worse) than those with a robust design. Bias for diagnostic tests works in different ways to bias in trials of treatment. *Bandolier*
Three key sources of bias in diagnostic studies

1. Inclusion of right spectrum of patients
2. Verification of patients
   - choice of reference standard
   - complete verification
3. Independent interpretation of index test and reference standard results (blinding)
Quality Assessment of Diagnostic Accuracy Studies, QUADAS-2

• Domain list
  - patient selection
  - index test
  - reference standard
  - flow and timing

• Signalling questions are used for judgments of risk of bias

• First three domains are also assessed for applicability

Whiting et al. Annals Internal Medicine. 2011
Are you concerned about risk of bias?

Risk of Bias

<table>
<thead>
<tr>
<th>Patient Selection</th>
<th>Index Test</th>
<th>Reference Standard</th>
<th>Flow and Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrews 2014</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Drain 2014c</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Lawn 2014a</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Nakiyingi 2014</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Peter 2012a</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Peter 2015</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Applicability Concerns

<table>
<thead>
<tr>
<th>Patient Selection</th>
<th>Index Test</th>
<th>Reference Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrews 2014</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Drain 2014c</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Lawn 2014a</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Nakiyingi 2014</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Peter 2012a</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Peter 2015</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

- High
- Low
? Unclear

Shah Cochrane 2016
Could the reference standard, its conduct, or its interpretation have introduced bias?

Signalling question 1: is the reference standard likely to correctly classify the target condition?

Due to the difficulties in diagnosing HIV-associated TB, it is recommended that multiple cultures from sputum and other specimens be evaluated.
STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies

http://www.equator-network.org/reporting-guidelines/stard/
Analyze the data
Statistical analysis and data synthesis

- Calculate estimates of sensitivity and specificity and 95% confidence intervals for individual studies
- Visually examine results of individual studies
- Calculate summary (pooled) accuracy estimates using recommended methods for meta-analysis
- Investigate possible reasons for heterogeneity
## Test: 2 Diagnosis of TB against microbiological reference at Grade 2: all participants

<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>Andrews 2014</td>
<td>16</td>
<td>14</td>
<td>3</td>
<td>60</td>
<td>0.84 [0.60, 0.97]</td>
<td>0.81 [0.70, 0.89]</td>
</tr>
<tr>
<td>Lawn 2014a</td>
<td>53</td>
<td>3</td>
<td>83</td>
<td>274</td>
<td>0.39 [0.31, 0.48]</td>
<td>0.99 [0.97, 1.00]</td>
</tr>
<tr>
<td>Nakinyi 2014</td>
<td>136</td>
<td>21</td>
<td>231</td>
<td>609</td>
<td>0.37 [0.32, 0.42]</td>
<td>0.97 [0.95, 0.98]</td>
</tr>
<tr>
<td>Peter 2012a</td>
<td>58</td>
<td>31</td>
<td>58</td>
<td>94</td>
<td>0.50 [0.41, 0.59]</td>
<td>0.75 [0.67, 0.82]</td>
</tr>
<tr>
<td>Peter 2015</td>
<td>41</td>
<td>27</td>
<td>140</td>
<td>361</td>
<td>0.23 [0.17, 0.29]</td>
<td>0.93 [0.90, 0.95]</td>
</tr>
</tbody>
</table>

http://tech.cochrane.org/revman
Forest plots of urine LAM sensitivity and specificity for TB diagnosis measured against a microbiological reference standard. The studies are ordered by decreasing sensitivity.

<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>Andrews 2014</td>
<td>16</td>
<td>14</td>
<td>3</td>
<td>60</td>
<td>0.84 [0.60, 0.97]</td>
<td>0.81 [0.70, 0.89]</td>
</tr>
<tr>
<td>Peter 2012a</td>
<td>58</td>
<td>31</td>
<td>58</td>
<td>94</td>
<td>0.50 [0.41, 0.59]</td>
<td>0.75 [0.67, 0.82]</td>
</tr>
<tr>
<td>Lawn 2014a</td>
<td>53</td>
<td>3</td>
<td>83</td>
<td>274</td>
<td>0.39 [0.31, 0.48]</td>
<td>0.99 [0.97, 1.00]</td>
</tr>
<tr>
<td>Nakiyingi 2014</td>
<td>136</td>
<td>21</td>
<td>231</td>
<td>609</td>
<td>0.37 [0.32, 0.42]</td>
<td>0.97 [0.95, 0.98]</td>
</tr>
<tr>
<td>Peter 2015</td>
<td>41</td>
<td>27</td>
<td>140</td>
<td>361</td>
<td>0.23 [0.17, 0.29]</td>
<td>0.93 [0.90, 0.95]</td>
</tr>
</tbody>
</table>

TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative
Meta-analysis of diagnostic test accuracy

• Calculate the diagnostic accuracy of a test

• Compare the diagnostic accuracy of two or more tests

• Investigate the variability of results between studies (heterogeneity is to be expected)
How is a meta-analysis performed?

• The bivariate model
  - gives average sensitivity and specificity
  - use when studies report a common threshold for a positive result

• Hierarchical summary ROC curve
  - gives a summary ROC curve
  - use when studies report several different thresholds

• Both models use random effects

Meta-analysis with STATA

Stata command: metandi
Practical Stata tutorial:
http://methods.cochrane.org/sdt/software-meta-analysis-dta-studies
Heterogeneity

• Refers to variation in results among studies

• May be caused by variation in
  – test thresholds (unique to meta-analyses of diagnostic tests)
  – prevalence of disease
  – patient spectrum
  – study quality
  – chance
  – unexplained
Visual inspection, do the confidence intervals overlap?
Heterogeneity, sensitivity and specificity of urine LAM for TB diagnosis stratified by CD4 count
When would you not do a meta-analysis?
It has been said that a fellow with one leg frozen in ice and the other leg in boiling water is comfortable - on average.

Interpret results and draw conclusions
**Summary of Findings Table - 1**

- **Question**: what is the diagnostic accuracy of LF-LAM for diagnosing TB in adults living with HIV?

- **Participants**: HIV-positive adults with symptoms of TB

- **Index test**: LF-LAM

- **Role**: a replacement test or test in combination with sputum smear microscopy or sputum Xpert® MTB/RIF

- **Reference standard**: microbiological (mainly mycobacterial culture)
Summary of Findings Table - 2

- **Studies**: cross-sectional
- **Setting**: inpatient and outpatient
- **Limitations**: the main limitations of the review were the use of a lower quality reference standard in most included studies, and the small number of studies and participants included in the analyses
- **Pooled sensitivity**: 45% (95% CrI: 29 to 63); pooled specificity: 92% (95% CrI: 80 to 97)
### Summary of Findings Table - 3

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 1000 patients tested (95% CrI)</th>
<th>Number of participants (studies)</th>
<th>Quality of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True positives</strong> (patients with tuberculosis)</td>
<td>45 (29 to 63)</td>
<td>135 (87 to 189)</td>
<td>819 (5)</td>
</tr>
<tr>
<td><strong>False negatives</strong> (patients incorrectly classified as not having tuberculosis)</td>
<td>55 (37 to 71)</td>
<td>165 (111 to 213)</td>
<td>Moderate¹</td>
</tr>
<tr>
<td><strong>True negatives</strong> (patients without tuberculosis)</td>
<td>828 (720 to 873)</td>
<td>644 (560 to 679)</td>
<td>1494 (5)</td>
</tr>
<tr>
<td><strong>False positives</strong> (patients incorrectly classified as having tuberculosis)</td>
<td>72 (27 to 180)</td>
<td>56 (21 to 140)</td>
<td>Low²,4,5</td>
</tr>
</tbody>
</table>

Pooled sensitivity: 45% (29, 63)
Pooled specificity: 92% (80, 97)
Do you have confidence in the results, why or why not?

- Completeness of evidence?
- Accuracy of the reference standard?
- Quality and quality of reporting of included studies?
- Applicability of studies?

Pooled sensitivity: 45% (29, 63)
Pooled specificity: 92% (80, 97)
Completeness of evidence?

• 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 and unpublished data.

• The search strategy included studies published in all languages.

• We acknowledge that we may have missed some studies despite the comprehensive search.
Accuracy of the reference standard?

- HIV-positive TB patients may have pulmonary TB, extrapulmonary TB, or both pulmonary and extrapulmonary TB.
- Due to the difficulties in diagnosing HIV-associated TB, it is recommended that multiple cultures from sputum and other specimen types be evaluated.
- We therefore considered a reference standard using two or more specimen types to be of higher quality than a reference standard using one specimen type.
Using GRADE approach, we judged the certainty of evidence for accuracy of LF-LAM for TB diagnosis to be low. This means that our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.

Studies were fairly well reported, though we corresponded with almost all study authors for additional data.

We had low concern about the applicability of the included studies to our review question.
• The most common error was to misinterpret the role of the reference standard as a comparator
• Participants could not recall definitions for sensitivity and specificity
• The Summary of Findings format was understood
In summary

• Express all components of the review question (PICO and more)
• Heterogeneity is to be expected in meta-analyses of diagnostic test accuracy; investigate the reasons for heterogeneity
• The Summary of Findings Table is useful for interpretation because it brings together the key elements of a review’s findings
References

• Cochrane Diagnostic Test Accuracy srtdta.cochrane.org/
• Leeflang. Clinical Microbiology and Infection, Volume 20 Number 2, February 2014
Initiatives to improve quality and reporting

- STARD: reporting of diagnostic studies
- PRISMA: reporting of systematic reviews/meta-analyses of RCTs
- STROBE: reporting of observational studies
- MOOSE: reporting of meta-analyses of observational studies
- AMSTAR: assessing quality of systematic reviews

www.equator-network.org/
Acknowledgements

- Clare Davenport
- Mariska Leeflang
- Hans Reitsma
- Maunank Shah
- Penny Whiting
<table>
<thead>
<tr>
<th>Analysis</th>
<th>Participants (number of studies)</th>
<th>Pooled estimates (95% CrI)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>LF-LAM, ALL</td>
<td>2313 (5 studies)</td>
<td>45% (29, 63)</td>
<td>92% (80, 97)</td>
</tr>
<tr>
<td>LAM alone</td>
<td>1876 (4 studies)</td>
<td>38% (34, 42)</td>
<td>98% (93, 100)</td>
</tr>
<tr>
<td>Microscopy alone</td>
<td>1876 (4 studies)</td>
<td>40% (27, 54)</td>
<td>95% (94, 97)</td>
</tr>
<tr>
<td>LAM and microscopy*</td>
<td>1876 (4 studies)</td>
<td>59% (47, 70)</td>
<td>92% (73, 97)</td>
</tr>
</tbody>
</table>

* Either test positive
<table>
<thead>
<tr>
<th>Analysis</th>
<th>Number of studies</th>
<th>Pooled estimates (95% CrI)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>LAM alone</td>
<td>3</td>
<td>36% (31, 42)</td>
<td>96% (94, 98)</td>
</tr>
<tr>
<td>Xpert alone</td>
<td>3</td>
<td>61% (39, 77)</td>
<td>97% (94, 99)</td>
</tr>
<tr>
<td>LAM and Xpert*</td>
<td>3</td>
<td>75% (61, 87)</td>
<td>93% (81, 97)</td>
</tr>
</tbody>
</table>

* Either test positive