



Cochrane
Library

Cochrane Database of Systematic Reviews

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults (Review)

Shah M, Hanrahan C, Wang ZY, Dendukuri N, Lawn SD, Denkinger CM, Steingart KR

Shah M, Hanrahan C, Wang ZY, Dendukuri N, Lawn SD, Denkinger CM, Steingart KR.

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults.

Cochrane Database of Systematic Reviews 2016, Issue 5. Art. No.: CD011420.

DOI: 10.1002/14651858.CD011420.pub2.

www.cochranelibrary.com

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults (Review)

Copyright © 2016 The Authors. Cochrane Database of Systematic Reviews published by John Wiley & Sons, Ltd. on behalf of The Cochrane Collaboration.

WILEY

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
BACKGROUND	4
Figure 1.	6
OBJECTIVES	7
METHODS	7
RESULTS	12
Figure 2.	13
Figure 3.	14
Figure 4.	15
Figure 5.	16
Figure 6.	17
Figure 7.	18
Figure 8.	19
Figure 9.	20
Figure 10.	21
Figure 11.	23
Figure 12.	24
Figure 13.	24
Figure 14.	25
Figure 15.	26
DISCUSSION	32
AUTHORS' CONCLUSIONS	33
ACKNOWLEDGEMENTS	34
REFERENCES	35
CHARACTERISTICS OF STUDIES	41
DATA	67
Test 1. Diagnosis of TB against microbiological reference at Grade 1: all participants.	70
Test 2. Diagnosis of TB against microbiological reference at Grade 2: all participants.	70
Test 3. Diagnosis of TB against composite reference at Grade 1: all participants.	71
Test 4. Diagnosis of TB against composite reference at Grade 2: all participants.	71
Test 5. Diagnosis of TB against microbiological reference at Grade 1, CD4 > 200.	72
Test 6. Diagnosis of TB against microbiological reference at Grade 1, CD4 < 200.	72
Test 7. Diagnosis of TB against microbiological reference at Grade 1, CD4 > 100.	73
Test 8. Diagnosis of TB against microbiological reference at Grade 1, CD4 < 100.	73
Test 9. Diagnosis of TB against microbiological reference at Grade 1, CD4 > 50.	74
Test 10. Diagnosis of TB against microbiological reference at Grade 1, CD4 < 50.	74
Test 11. Diagnosis of TB against microbiological reference at Grade 2, CD4 > 200.	75
Test 12. Diagnosis of TB against microbiological reference at Grade 2, CD4 < 200.	75
Test 13. Diagnosis of TB against microbiological reference at Grade 1, CD4 > 100.	76
Test 14. Diagnosis of TB against microbiological reference at Grade 1, CD4 < 100.	76
Test 15. Diagnosis of TB against microbiological reference at Grade 1, CD4 > 50.	77
Test 16. Diagnosis of TB against microbiological reference at Grade 1, CD4 < 50.	77
Test 17. Diagnosis of TB against microbiological reference at Grade 1: inpatients.	78
Test 18. Diagnosis of TB against microbiological reference at Grade 2: inpatients.	78
Test 19. Diagnosis of TB against microbiological reference at Grade 1: outpatients.	79
Test 20. Diagnosis of TB against microbiological reference at Grade 2: outpatients.	79
Test 21. Diagnosis of TB with LAM at Grade 2 (among studies comparing with smear).	80
Test 22. Diagnosis of TB with smear (among studies comparing with LAM).	80
Test 23. Diagnosis of TB with combination of LAM at Grade 2 with smear.	81

Test 24. Diagnosis of TB with LAM at Grade 2 (among studies comparing with Xpert®).	81
Test 25. Diagnosis of TB with Xpert® (among studies comparing with LAM).	82
Test 26. Diagnosis of TB with combination of LAM at Grade 2 with Xpert®.	82
Test 27. Screening for TB against microbiological reference at Grade 1: all participants.	83
Test 28. Screening for TB against microbiological reference at Grade 2: all participants.	83
Test 29. Screening for TB against composite reference at Grade 1: all participants.	84
Test 30. Screening for TB against composite reference at Grade 2: all participants.	84
Test 31. Screening for TB against microbiological reference at Grade 1, CD4 > 200.	85
Test 32. Screening for TB against microbiological reference at Grade 1, CD4 < 200.	85
Test 33. Screening for TB against microbiological reference at Grade 1, CD4 > 100.	86
Test 34. Screening for TB against microbiological reference at Grade 1, CD4 < 100.	86
Test 35. Screening for TB against microbiological reference at Grade 1, CD4 > 50.	87
Test 36. Screening for TB against microbiological reference at Grade 1, CD4 < 50.	87
Test 37. Screening for TB against microbiological reference at Grade 2, CD4 > 100.	88
Test 38. Screening for TB against microbiological reference at Grade 2, CD4 < 100.	88
Test 39. Screening for TB against microbiological reference at Grade 1: outpatients.	89
Test 40. Screening for TB with LAM at Grade 2 (among studies comparing with smear microscopy).	89
Test 41. Screening for TB with smear microscopy (among studies comparing with LAM).	90
Test 42. Screening for TB with the combination of LAM (Grade 2) and smear microscopy.	90
ADDITIONAL TABLES	90
APPENDICES	92
WHAT'S NEW	107
CONTRIBUTIONS OF AUTHORS	107
DECLARATIONS OF INTEREST	107
SOURCES OF SUPPORT	108
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	108
INDEX TERMS	109

[Diagnostic Test Accuracy Review]

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

Maunank Shah¹, Colleen Hanrahan², Zhuo Yu Wang³, Nandini Dendukuri³, Stephen D Lawn⁴, Claudia M Denkinger⁵, Karen R Steingart⁶

¹Department of Medicine, Division of Infectious Diseases, John Hopkins University School of Medicine, Baltimore, Maryland, USA. ²Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Maryland, USA. ³Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada. ⁴Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. ⁵FIND, Geneva, Switzerland. ⁶Cochrane Infectious Diseases Group, Liverpool School of Tropical Medicine, Liverpool, UK

Contact address: Maunank Shah, Department of Medicine, Division of Infectious Diseases, John Hopkins University School of Medicine, Baltimore, Maryland, USA. mshah28@jhmi.edu.

Editorial group: Cochrane Infectious Diseases Group.

Publication status and date: Edited (no change to conclusions), published in Issue 5, 2016.

Citation: Shah M, Hanrahan C, Wang ZY, Dendukuri N, Lawn SD, Denkinger CM, Steingart KR. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database of Systematic Reviews* 2016, Issue 5. Art. No.: CD011420. DOI: 10.1002/14651858.CD011420.pub2.

Copyright © 2016 The Authors. Cochrane Database of Systematic Reviews published by John Wiley & Sons, Ltd. on behalf of The Cochrane Collaboration. This is an open access article under the terms of the [Creative Commons Attribution-Non-Commercial](#) Licence, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

ABSTRACT

Background

Rapid detection of tuberculosis (TB) among people living with human immunodeficiency virus (HIV) is a global health priority. HIV-associated TB may have different clinical presentations and is challenging to diagnose. Conventional sputum tests have reduced sensitivity in HIV-positive individuals, who have higher rates of extrapulmonary TB compared with HIV-negative individuals. The lateral flow urine lipoarabinomannan assay (LF-LAM) is a new, commercially available point-of-care test that detects lipoarabinomannan (LAM), a lipopolysaccharide present in mycobacterial cell walls, in people with active TB disease.

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).

Search methods

We searched the following databases without language restriction on 5 February 2015: the Cochrane Infectious Diseases Group Specialized Register; MEDLINE (PubMed, 1966); EMBASE (OVID, from 1980); Science Citation Index Expanded (SCI-EXPANDED, from 1900), Conference Proceedings Citation Index-Science (CPCI-S, from 1900), and BIOSIS Previews (from 1926) (all three using the Web of Science platform); MEDION; LILACS (BIREME, from 1982); SCOPUS (from 1995); the metaRegister of Controlled Trials (mRCT); the search portal of the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP); and ProQuest Dissertations & Theses A&I (from 1861).

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults (Review)

Copyright © 2016 The Authors. Cochrane Database of Systematic Reviews published by John Wiley & Sons, Ltd. on behalf of The Cochrane Collaboration.

1

Selection criteria

Eligible study types 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). A higher quality reference standard was one in which two or more specimen types were evaluated for TB, and a lower quality reference standard was one in which only one specimen type was evaluated for TB. Participants were HIV-positive people aged 15 years and older.

Data collection and analysis

Two review authors independently extracted data from each included study using a standardized form. We appraised the quality of studies using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. We evaluated the test at two different cut-offs: (grade 1 or 2, based on the reference card scale of five intensity bands). Most analyses used grade 2, the manufacturer's currently recommended cut-off for positivity. We carried out meta-analyses to estimate pooled sensitivity and specificity using a bivariate random-effects model and estimated the models using a Bayesian approach. We determined accuracy of LF-LAM combined with sputum microscopy or Xpert® MTB/RIF. In addition, we explored the influence of CD4 count on the accuracy estimates. We assessed the quality of the evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

Main results

We included 12 studies: six studies evaluated LF-LAM for TB diagnosis and six studies evaluated the test for TB screening. All studies were cross-sectional or cohort studies. Studies for TB diagnosis were largely conducted among inpatients (median CD4 range 71 to 210 cells per μL) and studies for TB screening were largely conducted among outpatients (median CD4 range 127 to 437 cells per μL). All studies were conducted in low- or middle-income countries. Only two studies for TB diagnosis (33%) and one study for TB screening (17%) used a higher quality reference standard.

LF-LAM for TB diagnosis (grade 2 cut-off): meta-analyses showed median pooled sensitivity and specificity (95% credible interval (CrI)) of 45% (29% to 63%) and 92% (80% to 97%), (five studies, 2313 participants, 35% with TB, *low quality evidence*). The pooled sensitivity of a combination of LF-LAM and sputum microscopy (either test positive) was 59% (47% to 70%), which represented a 19% (4% to 36%) increase over sputum microscopy alone, while the pooled specificity was 92% (73% to 97%), which represented a 6% (1% to 24%) decrease from sputum microscopy alone (four studies, 1876 participants, 38% with TB). The pooled sensitivity of a combination of LF-LAM and sputum Xpert® MTB/RIF (either test positive) was 75% (61% to 87%) and represented a 13% (1% to 37%) increase over Xpert® MTB/RIF alone. The pooled specificity was 93% (81% to 97%) and represented a 4% (1% to 16%) decrease from Xpert® MTB/RIF alone (three studies, 909 participants, 36% with TB). Pooled sensitivity and specificity of LF-LAM were 56% (41% to 70%) and 90% (81% to 95%) in participants with a CD4 count of less than or equal to 100 cells per μL (five studies, 859 participants, 47% with TB) versus 26% (16% to 46%) and 92% (78% to 97%) in participants with a CD4 count greater than 100 cells per μL (five studies, 1410 participants, 30% with TB).

LF-LAM for TB screening (grade 2 cut-off): for individual studies, sensitivity estimates (95% CrI) were 44% (30% to 58%), 28% (16% to 42%), and 0% (0% to 71%) and corresponding specificity estimates were 95% (92% to 97%), 94% (90% to 97%), and 95% (92% to 97%) (three studies, 1055 participants, 11% with TB, *very low quality evidence*). There were limited data for additional analyses.

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. The results should, therefore, be interpreted with caution.

Authors' conclusions

We found that LF-LAM has low sensitivity to detect TB in adults living with HIV whether the test is used for diagnosis or screening. For TB diagnosis, the combination of LF-LAM with sputum microscopy suggests an increase in sensitivity for TB compared to either test alone, but with a decrease in specificity. In HIV-positive individuals with low CD4 counts who are seriously ill, LF-LAM may help with the diagnosis of TB.

PLAIN LANGUAGE SUMMARY

The lateral flow urine lipoarabinomannan (LF-LAM) test for diagnosis of tuberculosis in people living with human immunodeficiency virus (HIV)

Background

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults (Review)

Copyright © 2016 The Authors. Cochrane Database of Systematic Reviews published by John Wiley & Sons, Ltd. on behalf of The Cochrane Collaboration.

Tuberculosis (TB) is a common cause of death in people with human immunodeficiency virus (HIV) infection, but diagnosis is difficult, and depends on testing for TB in the sputum and other sites, which may take weeks to give results. A rapid and accurate point-of-care test could reduce delays in diagnosis, allow treatment to start promptly, and improve linkage between diagnosis and treatment.

Test evaluated by this review

The lateral flow urine lipoarabinomannan assay (LF-LAM, Alere DetermineTM TB LAM Ag, Alere Inc, Waltham, MA, USA) is a commercially available point-of-care test for active TB (pulmonary and extrapulmonary TB). The test detects lipoarabinomannan (LAM), a component of the bacterial cell walls, which is present in some people with active TB. The test is performed by placing urine on one end of a test strip, with results appearing as a line (that is, a band) on the strip if TB is present. The test is simple, requires no special equipment, and shows results in 25 minutes. During the period we conducted the review, the manufacturer issued new recommendations for defining a positive test. We collected data based on both the original and the new recommendations

Objectives

We aimed to see how accurately LF-LAM diagnosed TB in people living with HIV with TB symptoms, and how accurately LF-LAM diagnosed TB in people living with HIV being screened for TB whether or not they had TB symptoms.

Main results

We examined evidence up to 5 February 2015 and included 12 studies: six studies evaluated LF-LAM for TB diagnosis and six studies evaluated the test for TB screening. All studies were conducted in low- or middle-income countries.

Quality of the evidence

We assessed quality by describing how participants were selected for the studies, details of the test and reference standards (the benchmark test), and study flow and timing, using the standard QUADAS-2 approach. Few studies used multiple types of specimens for the reference standard (higher quality standard) and most relied on sputum culture alone (lower quality standard), which may have affected results.

What do the results mean?

In a population of 1000 HIV-positive individuals with TB symptoms, where 300 actually have TB, the test will correctly identify 135 people as having TB, but miss the remaining 165 people; for the 700 people who do not have TB, the test will correctly identify 644 people as not having TB, but will misclassify 56 as having TB.

The sensitivity of the test is higher in people living with HIV with low CD4 cell counts who are at risk of life-threatening illnesses. In patients with a CD4 \leq 100 cells per μ L, LF-LAM sensitivity was 56% (41% to 70%) versus 26% (16% to 46%) in patients with a CD4 count $>$ 100 cells per μ L.

If the test is used in screening HIV-positive people for TB, in a population of 1000 where 10 actually have TB, LF-LAM will correctly identify none of the 10, or up to four of the 10; on the other hand, the test will miss six to 10 people with TB; in the remaining 990 who do not have TB, the test will correctly identify 931 to 941 people as not having TB while misclassifying 49 to 59 as having TB.

Limitations

The main limitations of the review were the use of a lower quality reference standard in most included studies, and small number of studies and participants included in the analyses. The results should, therefore, be interpreted with caution.

Conclusions

In this Cochrane review, we found that LF-LAM, whether the test is used for diagnosis or screening, has low sensitivity to detect TB. However, in HIV-positive people with low CD4 counts who are seriously ill, LF-LAM may help with the diagnosis of TB.

BACKGROUND

Target condition being diagnosed

Tuberculosis (TB) is caused by the mycobacterium *Mycobacterium tuberculosis* and is transmitted through respiratory aerosols produced by people with active TB disease. Infection can be asymptomatic for many years (referred to as latent TB infection), with bacteria persisting in a viable but minimally replicating state. In a small percentage of people, latent infection may reactivate at a later time in a process that is commonly characterized by an increase in the replication and number of bacteria, which leads to symptomatic 'active TB disease'. TB predominantly affects the lungs (pulmonary TB), but can also affect other parts of the body, such as the brain or spine (extrapulmonary TB). Human immunodeficiency virus (HIV) infection is the most potent risk factor for progression from latent TB infection to active TB disease (Kwan 2011). In this systematic review we are interested in the diagnosis of active TB disease in people living with HIV.

Worldwide, in 2014, an estimated 1.2 million (12%) of the 9.6 million people who developed TB were HIV-positive; 74% of these HIV-positive TB cases were in the African Region (WHO Global Tuberculosis Report 2015). In 2014, TB killed 1.5 million people. The number of people dying from HIV-associated TB has been falling since 2004 (570,000 deaths); however in 2014 an estimated 390,000 people died from HIV-associated TB (WHO Global Tuberculosis Report 2015). TB is the leading cause of death among people living with HIV, estimated to account for around 33% of all HIV-related deaths globally (WHO Global Tuberculosis Report 2015). However, a recent systematic review of the prevalence of TB identified at autopsy suggests that, in resource-limited settings, TB is responsible for an even higher percentage (around 37%) of all HIV-related deaths (Gupta 2015). Most deaths from TB are preventable. It is estimated that 43 million lives were saved between 2000 and 2014 through effective diagnosis and treatment (WHO Global Tuberculosis Report 2015). TB and HIV are linked by geography and biology, and their intersection has been called a "synergy from hell" (Bartlett 2007). Geographically, HIV and TB are often concentrated in areas of poverty with limited resources for diagnosis, treatment, and prevention of TB. Much of the global burden of TB is concentrated in sub-Saharan Africa, and HIV coinfection represents a major driver of the epidemic in many areas. Linked by biology, each disease speeds the progress of the other. The risk of developing TB is much higher in people living with HIV, estimated to be 20 to 37 times higher in HIV-positive individuals than in HIV-negative individuals (Getahun 2010). TB occurs early in the course of HIV infection and shortens survival (Havlr 2008; Whalen 1995). Many HIV-positive people in developing countries develop TB as the first manifestation of AIDS (WHO Global Tuberculosis Report 2015).

Signs and symptoms of TB in people living with HIV vary, which makes it challenging to determine when to consider a diagnosis of TB. Fever, weight loss, and fatigue are often the only symptoms of HIV-associated TB and are non-specific. As a result, testing may be delayed while other causes are evaluated. A cough for longer than two weeks is a common distinguishing feature that prompts diagnostic testing for TB in people who are HIV-negative, but is present in less than a third of people with TB who are HIV positive (Cain 2010). Similarly, radiographic features of TB in people living with HIV can be misleading or atypical. Whereas upper lobe cavitory lesions are often seen in HIV-negative TB patients, such lesions are less common in HIV-positive TB patients. In comparison with HIV-negative people, HIV-positive people also have higher rates of extrapulmonary TB or mycobacteraemia (TB bloodstream infection). Extrapulmonary TB presentation varies depending on the body site affected, and can mimic other diseases such as cancer and bacterial and fungal infections. Identifying people who warrant further testing for TB may therefore be challenging in HIV-positive people.

Even when TB is suspected, current diagnostic approaches are inadequate. Sputum smear microscopy is the most widely utilized TB diagnostic test worldwide. It has low sensitivity for TB (50% to 60% on average) and identifies less than 50% of people with HIV-related TB (Siddiqi 2003). Sputum smear microscopy does not, by definition, identify smear-negative pulmonary or extrapulmonary TB, which are more common in people who are HIV-positive than HIV-negative (Getahun 2007; Perkins 2007; Steingart 2006). Mycobacterial culture, the reference standard for TB, is costly, requires significant laboratory infrastructure, takes weeks to provide results, and is not widely available. Detection of extrapulmonary TB is dependent on intensive evaluation of sites of disease and culture of clinical specimens other than sputum, in conjunction with clinical and radiographic evaluations. Hence, many people with TB remain undiagnosed if co-infected with HIV.

Recently, a new test has become available for TB detection. The Xpert® MTB/RIF assay (Cepheid, Sunnyvale, USA) is a fully automated, nucleic acid amplification test (NAAT) for TB and drug-resistant TB and is endorsed by the World Health Organization (WHO) (WHO Xpert® Policy Update 2013). A recent Cochrane Review found that Xpert® MTB/RIF was sensitive and specific for both TB detection and rifampicin resistance detection and, compared with smear microscopy, substantially increased TB detection among culture-confirmed cases (Steingart 2014). Although sputum testing with Xpert® MTB/RIF has high sensitivity for smear-positive pulmonary TB (98%), sensitivity is lower for smear-negative pulmonary TB (67%) (Steingart 2014). This finding has bearing for HIV-associated TB where smear-negative TB and extrapulmonary TB are disproportionately higher.

Index test(s)

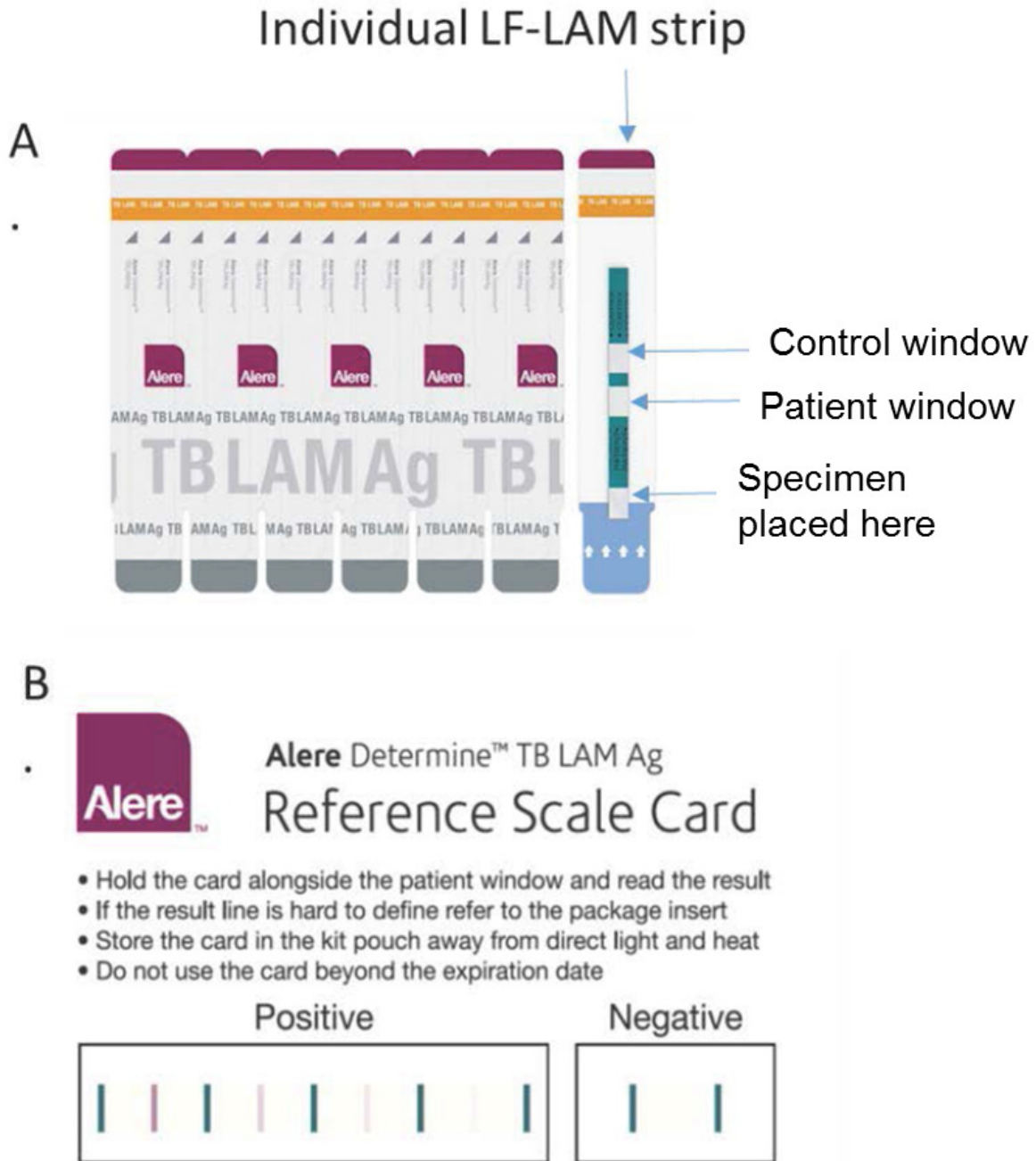
The lateral flow urine lipoarabinomannan assay (LF-LAM) is a commercially available point-of-care test for active TB (Alere DetermineTM TB LAM Ag, Alere Inc, Waltham, MA, USA). The test detects lipoarabinomannan (LAM), a lipopolysaccharide present in mycobacterial cell walls, which is released from metabolically active or degenerating bacterial cells and appears to be present only in people with active TB disease.

LAM is detectable in the urine of people with active TB (Peter 2010). Urine-based testing has advantages over sputum-based testing because urine is easy to collect and store, and lacks the infection control risks associated with sputum collection. Several studies (Dheda 2010; Lawn 2009; Lawn 2012a; Lawn 2012b; Peter 2012; Shah 2009) and a meta-analysis of an earlier generation LAM-ELISA test (Minion 2011) have found that the accuracy of urinary LAM detection may be improved among people living with HIV with advanced immunosuppression. Several hypotheses may explain the higher sensitivity of urine LAM detection in HIV-positive people including higher bacillary burden and antigen load, greater likelihood of genitourinary tract TB involve-

ment, and greater glomerular permeability to allow increased antigen levels in urine (Minion 2011).

LF-LAM is performed manually by applying 60 µL of urine to the DetermineTM TB LAM Ag test strip and incubating at room temperature for 25 minutes (Alere 2014). The strip is then inspected by eye. The intensity of any visible band on the test strip is graded by comparing it with the intensities of the bands on a manufacturer-supplied reference card. Prior to January 2014, this reference card included five bands (grade 1 representing a very low intensity band to grade 5 representing a high/dark intensity band). Some studies prior to January 2014 utilized grade 1 as the threshold for test positivity, while other studies utilized grade 2 as the positivity threshold. After January 2014, the manufacturer revised the reference card to have four reference bands, such that the band intensity for the new grade 1 corresponded to the band intensity for the previous grade 2. Under manufacturer recommendations (using the new reference card), bands that are grade 1 or higher are considered positive (Figure 1).

Figure 1. (A) Alere Determine™ TB LAM Ag tests. To the sample pad (white pad marked by the arrow symbols) 60 µL of urine is applied and visualized bands are read 25 minutes later. (B) Reference card accompanying test strips to 'grade' the test result and determine positivity (33). Copyright © [2014] [Alere Inc]: reproduced with permission.



Clinical pathway

The WHO guidelines on intensified case-finding in people living with HIV recommend that, in resource-constrained settings, people living with HIV who report “any one of the symptoms of current cough, fever, weight loss, or night sweats should be evaluated for TB and other diseases” (Getahun 2011; WHO Tuberculosis Screening 2013). Further diagnostic work-up for people with a positive symptom screen includes sputum smear microscopy, and, in some settings, sputum Xpert® MTB/RIF or mycobacterial culture (WHO Tuberculosis Screening 2013). Importantly, some HIV-positive people may have difficulty producing any sputum at all, or may have extrapulmonary forms of TB. When extrapulmonary TB is suspected, it is standard practice to obtain “appropriate specimens from the suspected sites of involvement for microscopy, culture, and histopathological examination” (TB CARE I 2014). Evaluation for extrapulmonary TB often requires invasive diagnostic procedures that may have low yield even in people with advanced disease.

People living with HIV are at increased risk of TB but may be asymptomatic. As such, urine LF-LAM could also be used to screen for TB in people living with HIV, irrespective of symptoms. The WHO defines screening for active TB as “the systematic identification of people with suspected active TB, in a predetermined target group, using tests, examinations or other procedures that can be rapidly applied” (WHO Tuberculosis Screening 2013). LF-LAM could be used as either (1) a replacement test for an existing test (that is, sputum smear microscopy or sputum Xpert® MTB/RIF), or (2) a parallel test, i.e. a ‘new test’ intended to be used concurrently in combination with an existing test.

Alternative test(s)

We evaluated the performance of LF-LAM in comparison to or in combination with sputum smear microscopy or sputum Xpert® MTB/RIF. Sputum smear microscopy involves the microscopic evaluation of sputum specimens using stains for acid-fast bacilli that identify mycobacteria (but do not distinguish *M. tuberculosis* from nontuberculous mycobacteria). Xpert® MTB/RIF is a semi-automated integrated specimen processing and NAAT for detection of TB and rifampicin resistance.

Rationale

New tests and strategies for detection of TB are urgently needed to curb the ongoing HIV-TB co-epidemic. Among the key priorities identified by the WHO, healthcare providers, patients, and advocacy groups is development of a point-of-care test for TB (Batz 2011; Pai 2012; Weyer 2011). LF-LAM, if sufficiently accurate, would satisfy many of the established minimum specifications for a point-of-care test for TB (Appendix 1; Batz 2011). Furthermore,

the test could provide obvious benefits for HIV-positive people who suffer the highest morbidity and mortality, by earlier detection of pulmonary TB that may be missed by sputum smear microscopy and sputum Xpert® MTB/RIF and extrapulmonary TB that may be missed by sputum-based testing. The draft of this systematic review informed the WHO policy recommendations on the use of LF-LAM for the diagnosis and screening of active TB in people living with HIV (WHO Lipoarabinomannan Policy Guidance 2015).

OBJECTIVES

Primary objectives

- To assess the accuracy of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis (TB) disease in HIV-positive adults who have signs and symptoms suggestive of TB.
- To assess the accuracy of the LF-LAM as a screening test for active TB disease in HIV-positive adults irrespective of signs and symptoms suggestive of TB.

Secondary objectives

- To compare the 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. Possible sources of heterogeneity include CD4 count and clinical setting (inpatient versus outpatient setting).

METHODS

Criteria for considering studies for this review

Types of studies

We included primary studies that evaluated the accuracy of LF-LAM for diagnosis of or screening for active TB in people living with HIV and compared the index test with a defined reference

standard. We also included studies that assessed the diagnostic accuracy of one of the alternative tests (sputum microscopy or sputum Xpert® MTB/RIF) in addition to LF-LAM. Eligible study types included randomized controlled trials, cross-sectional studies, and observational cohort studies. We included abstracts with sufficient data. We excluded case-control studies.

We included studies that provided data from which we could extract true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values, based on one of the reference standards defined below.

Participants

Participants were adults (15 years and older is considered 'adult' for purpose of TB surveillance) and who were HIV positive. We included participants in whom there was a suspicion of TB based on the presence of signs and symptoms compatible with TB, as well as participants who may or may not have had signs and symptoms compatible with TB and had not been previously evaluated for TB. Signs and symptoms of TB include cough, fever, weight loss, and night sweats.

Index tests

The index test was Alere Determine™ TB LAM Ag test (LF-LAM, Alere Inc., Waltham, MA, USA), the only commercial lateral flow urine LAM assay available as of December 2015. We evaluated the test at two different cut-off values for positivity (grade 1 and grade 2) based on the original manufacturer reference card. Grade 2 (corresponding to grade 1 on the new manufacturer reference card) is the currently recommended threshold for positivity.

Target conditions

The target condition was active TB disease, which includes pulmonary and extrapulmonary TB.

Reference standards

We required studies to diagnose TB using at least one of the following two reference standards.

- Microbiological reference standard:
 - we defined 'TB' as a positive *M. tuberculosis* culture or nucleic acid amplification test (NAAT);
 - we defined 'not TB' as a negative *M. tuberculosis* culture and NAAT (if performed).
- Composite reference standard that included *M. tuberculosis* culture together with at least one of the following components: NAAT, smear, or clinical findings:
 - we defined 'TB' as (1) a positive culture, or (2) a positive NAAT, or (3) a positive smear, or (4) a clinical decision to start TB treatment, and, after at least one month of follow-up, the participant was diagnosed as having TB;

- we defined 'not TB' as a negative culture and NAAT (if performed), no TB treatment given, and resolution of signs and symptoms at follow-up.

NAATs included: Enhanced Amplified Mycobacterium Tuberculosis Direct Test (E-MTD, Gen-Probe, San Diego, USA); Amplicor *Mycobacterium tuberculosis* Test (Amplicor, Roche Diagnostics, Basel, Switzerland); COBAS® TaqMan® MTB Test (Roche Diagnostics); GenoType MTBDR^{plus} (HAIN Lifesciences, Nehren, Germany); and Xpert® MTB/RIF assay (Cepheid, Sunnyvale, USA).

For a microbiological reference standard, we considered a higher quality reference standard to be one in which two or more specimen types were evaluated for TB diagnosis and a lower quality reference standard to be one in which only one specimen type was evaluated for TB diagnosis. For a composite reference standard, we did not require all components to be provided on all participants. Of note, we excluded sputum smear from a composite reference standard and sputum Xpert® MTB/RIF from both a microbiological and a composite reference standard for the purpose of performing analyses of LF-LAM in combination with either sputum smear or sputum Xpert® MTB/RIF.

We excluded participants as 'unclassifiable' if we could not classify them as either 'TB' or 'not TB' based on these reference standard definitions.

We consider that there are strengths and limitations to each reference standard. A microbiological reference standard, primarily culture, is considered the best reference standard. We expected all studies to obtain sputum specimens and some studies to obtain additional specimens for culture. However, the primary concern with relying on sputum culture alone is that people with TB disease may be missed for the following reasons: HIV-positive people may not be able to provide sputum specimens of sufficient quality; sputum bacillary load is typically low in people living with HIV; and a substantial proportion of people with HIV-associated TB cannot produce sputum at all (Lawn 2013a) or have extrapulmonary TB without pulmonary TB. This means that index test TPs may be misclassified as FPs by sputum culture. Therefore, when evaluating LF-LAM with respect to sputum culture, the number of FPs (classified as positive by the index test and negative by the reference test) may be increased and LF-LAM specificity may be underestimated (Lawn 2015). This misclassification may also lead to underestimation of sensitivity. Increasing the sensitivity of the reference standard by evaluating multiple specimens, including evaluating specimens from sites of disease for extrapulmonary TB, may reduce the number of cases of TB disease incorrectly classified as 'not TB' by culture.

In contrast, a composite reference standard that includes microbiological or clinical components may correctly classify index test results as TPs (instead of as FPs with respect to culture), especially in people with paucibacillary disease in whom culture may be negative. However, because of the uncertainties that surround a clinical diagnosis of TB, a reference standard that uses clinical TB

(in culture-negative people) is considered a lower quality reference standard and may incorrectly classify people without TB as having TB.

Search methods for identification of studies

Electronic searches

To identify all relevant studies, on 5 February 2015, Vittoria Lutje (VL), the Information Specialist for the Cochrane Infectious Diseases Group (CIDG), performed literature searches without language restrictions in the following databases using the search terms we have reported in [Appendix 2](#): the Cochrane Infectious Diseases Group Specialized Register; MEDLINE (PubMed, from 1966); EMBASE (COVID, from 1980); Science Citation Index Expanded (SCI-EXPANDED, from 1900), Conference Proceedings Citation Index- Science (CPCI-S, from 1900), and BIOSIS Previews (from 1926), all three using the Web of Science platform; MEDION (<http://www.mediondatabase.nl/>); LILACS (BIREME, from 1982); and SCOPUS (from 1995). She also searched the metaRegister of Controlled Trials (mRCT) and the search portal of the WHO International Clinical Trials Registry Platform (WHO ICTRP, www.who.int/trialsearch) to identify ongoing trials, and ProQuest Dissertations & Theses A&I (from 1861) to identify relevant dissertations.

Searching other resources

We examined reference lists of relevant reviews and studies, searched the WHO websites, and contacted Alere Inc. and researchers in the field to identify ongoing studies.

Data collection and analysis

Selection of studies

Two review authors (MS and CH) first independently examined all titles and abstracts identified from the electronic search to determine potentially eligible studies. We then obtained the full-text articles of these potentially eligible studies and the same two review authors independently assessed inclusion based on predefined inclusion and exclusion criteria. We resolved disagreements through discussion and, if necessary, consulted a third review author (KRS).

Data extraction and management

We developed a standardized data extraction form and piloted the form on two of the included studies. Based upon the results of the pilot, we finalized the form. Then two review authors (MS and CH) independently extracted data from each included study on the following characteristics.

- Author, publication year, study design, country(ies), clinical setting (outpatient or inpatient).
- Participants: age, gender, CD4 count, clinical status (asymptomatic, symptomatic).
- Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) items.
 - Cut-off used for determining a positive index test result, grade 1 or 2.
 - Reference standard(s).
 - Number of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) values and for the index test, sputum microscopy, and sputum Xpert® MTB/RIF.
 - Missing or unavailable test results.
 - Intra-reader and inter-reader variability.
 - Participant health outcomes: time to treatment initiation (time from specimen collection until time the participant starts treatment), morbidity, and mortality.

We assigned country income status (high-income, or low- and middle-income) as classified by the World Bank List of Economies ([World Bank 2014](#)).

For one study that reported results using the new reference card ([LaCourse 2015](#)), we converted results to the original manufacturer scale to allow consistent comparisons.

For a composite reference standard, we found that these data were largely unavailable in the publications and interpreted differently by study authors. Therefore, we sent our definition of a composite reference standard to study authors and requested that they provide TP, FP, FN, and TP values using our definition. Furthermore, we excluded participants assigned to 'TB' and 'not TB' if study authors considered clinical findings alone and there was not at least one month of follow-up. Participants may be lost during follow-up for unknown reasons or death. We excluded these participants because we could not accurately classify their TB status.

If a study applied an approach of testing all participants in the inpatient setting, we assigned the purpose of testing as 'diagnosis', rather than screening. We made this assignment because we considered it likely that most inpatients had TB symptoms even though they were not enrolled in the study on the basis of specific TB symptoms.

We contacted study authors for missing data and clarifications and entered all data into Microsoft® Excel.

The data extraction form is in [Appendix 3](#).

Assessment of methodological quality

We used the QUADAS-2 tool to assess the quality of the included studies (Whiting 2011; Appendix 4). QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing (flow and timing domain includes differential verification of TB status for study participants). We assessed all domains for risk of bias and the first three domains for concerns regarding applicability. As recommended, we first developed guidance on how to appraise each question and interpreted this information tailored to this review. Then, one review author (MS) piloted the tool with two of the included studies. Based on the experience we gained from the pilot, we finalized the tool. Two review authors (MS and CH) independently completed QUADAS-2. We resolved disagreements through discussion or consulted a third review author (KRS). We presented the results of the quality assessment in the text, the 'Characteristics of included studies' table, and graphs.

Statistical analysis and data synthesis

We performed descriptive analyses of the characteristics of the included studies using Stata 13 (StataCorp 2011), and presented key study characteristics in the 'Characteristics of included studies' table. We used the number of TPs, FPs, FNs, and TNs to calculate the individual study estimates of sensitivity and specificity and their 95% confidence intervals (CI). We presented individual study results graphically by plotting the estimates of sensitivity and specificity (and their 95% CIs) in forest plots and receiver operating characteristic (ROC) space using Review Manager (RevMan) (Review Manager). We presented the results from all studies using the original manufacturer scale for test interpretation, with band intensities graded on a scale of 1 to 5 (from lightest to darkest). The newer manufacturer reference card is graded on a scale of 1 to 4, with lightest band intensity (grade 1) corresponding to the prior grade 2. To allow consistent comparisons, we converted results from studies utilizing the newer manufacturer reference card to the original manufacturer scale (that is, 'grade 1' results using the new scale are treated in this review as 'grade 2').

As mentioned above, during the period in which we conducted the review, the manufacturer issued new recommendations for defining a positive test. Although, we determined accuracy estimates separately at the original (grade 1) and currently recommended (grade 2) cut-offs and present the estimates in forest plots and tables, we focused our attention in the main text on the diagnostic accuracy of grade 2.

We used a microbiological reference standard for most analyses. We grouped studies according to the purpose of testing, diagnosis, or screening. Then, when we considered the studies to be comparable, we carried out meta-analyses to estimate LF-LAM pooled sensitivity and specificity with a bivariate random-effects model (Chu 2009; Reitsma 2005). This approach allowed us to calculate pooled sensitivity and specificity while dealing with potential sources of variation caused by: (1) imprecision of sensitivity and

specificity estimates within individual studies; (2) correlation between sensitivity and specificity across studies; and (3) variation in sensitivity and specificity between studies.

We compared the sensitivity and specificity of LF-LAM when performed alone, with the sensitivity and specificity of sputum smear microscopy or sputum Xpert® MTB/RIF in direct comparisons in which both tests were evaluated in the same participants. We additionally determined the value of LF-LAM when used as a parallel test in combination with sputum microscopy or sputum Xpert® MTB/RIF based on data from studies that reported results of LF-LAM and either or both of these tests. To illustrate, for LF-LAM used as a parallel test, we determined the increase in sensitivity of a combination of LF-LAM and (1) sputum microscopy or (2) sputum Xpert® MTB/RIF compared with the sensitivity of sputum smear microscopy or sputum Xpert® MTB/RIF alone. We determined the potential decrease in specificity of a combination of LF-LAM and (1) sputum microscopy or (2) sputum Xpert® MTB/RIF compared with the specificity of sputum microscopy or sputum Xpert® MTB/RIF alone. In these meta-analyses, we accounted for conditional dependence between the accuracy estimates of LF-LAM and the comparator test (sputum microscopy, or sputum Xpert® MTB/RIF, or a combination of LF-LAM with sputum microscopy or sputum Xpert® MTB/RIF) by extending the standard bivariate model to include separate study-level covariance terms (Novielli 2013; Vacek 1985) between the sensitivities of the two tests, for example, Xpert® MTB/RIF and LF-LAM. We did not add covariance terms between the specificities, which were generally very high for all tests.

We estimated all models using a Bayesian approach implemented using WinBUGS (Lunn 2000). Under the Bayesian approach, all unknown parameters must be provided a prior distribution that defines the range of possible values of the parameter and the weight of each of those values, based on information external to the data. In order to let the observed data dominate the final results, we chose to use low-information prior distributions. We defined prior distributions on the log-odds scale over the pooled sensitivity and specificity parameters, their corresponding between-study standard deviations (SDs) and the correlation between the sensitivities and specificities across studies. For the pooled log odds of the sensitivity or log odds of the specificity, we used a normal prior distribution with mean 0 and a wide variance of 4 (or a precision of 0.25). This corresponds to a roughly uniform distribution over the pooled sensitivity and pooled specificity on the probability scale. For the between-study precision we used a gamma distribution with a shape parameter of two and rate parameter of 0.5. This corresponds to a 95% prior credible interval (CrI) for the between-study SD in the log odds of sensitivity or log odds of specificity ranging from roughly 0.29 to 1.44, corresponding to moderate to high values of between-study heterogeneity. Covariance terms followed a uniform prior distribution whose upper and lower limits were determined by the sensitivity of the two tests. We have summarized the models we used (including the prior distri-

butions) and the WinBUGS programs we used to estimate them in [Appendix 5](#).

We had two instances where there were only three included studies in the meta-analysis, which led the model to be just identified and highly dependent on the prior information. Nonetheless, we chose not to simplify the bivariate to a univariate model in these cases as we wished to estimate the incremental value of one test over another by taking the difference between parameters in these two meta-analysis models. Bivariate models are preferable for such comparisons [Trikalinos 2014](#).

Meta-analysis models based on few studies are sensitive to the choice of prior distributions, particularly the prior distributions over the between-study SD parameters. To study the sensitivity of all our results to the choice of prior distributions given above, we considered alternative prior distributions that were less informative, which allowed a wider range of possible values. We increased the variance of the normal distributions over the pooled log odds of the sensitivity or specificity to 100. We used a uniform prior distribution ranging from zero to three over the between-study SD on the log odds scale. We found that the pooled estimates remained roughly the same with these alternative priors, though the posterior CrIs were wider, as expected.

We combined information from the prior distribution with the likelihood of the observed data, in accordance with Bayes' theorem in the WinBUGS program, which resulted in a sample from the posterior distribution of each unknown parameter. Using this sample, we calculated various descriptive statistics of interest. We estimated the median pooled sensitivity and specificity and their 95% CrI. The median or the 50% quantile is the value below which 50% of the posterior sample lies. We reported the median because the posterior distributions of some parameters may be skewed and the median would be considered a better point estimate of the unknown parameter than the mean in such cases. The 95% CrI is the Bayesian equivalent of the classical (frequentist) 95% CI (we indicated 95% CI for individual study estimates and 95% CrI for pooled study estimates as appropriate). The 95% CrI may be interpreted as an interval that has a 95% probability of capturing the true value of the unknown parameter given the observed data and the prior information.

To compare tests, we first defined the difference in the pooled sensitivity and the difference in the pooled specificity. Once the posterior distribution of the difference in sensitivities between two tests is available, we can estimate the probability that this difference exceeds zero.

Approach to uninterpretable LF-LAM results

We excluded uninterpretable test results from the analyses for determination of sensitivity and specificity. We summarized uninterpretable results as reported by the included studies.

Investigations of heterogeneity

Initially, we investigated heterogeneity through visual examination of forest plots of sensitivities and specificities and through visual examination of the ROC plot of the raw data. When data were sufficient, we performed subgroup analyses with the following categorical covariates: CD4 count (> 200 versus ≤ 200 cells per μL ; > 100 versus ≤ 100 cells per μL ; and > 50 versus ≤ 50 cells per μL) and clinical setting (inpatient versus outpatient). We generated the plots depicting the pooled results within CD4 count categories using R ([R Statistical Computing 2015](#)).

Sensitivity analyses

For our primary analysis using a microbiological reference standard, we performed sensitivity analyses by limiting inclusion in the meta-analysis to the following.

- Studies that avoided inappropriate exclusions, for example, studies that included participants who could not produce sputum. For this analysis we included studies that we scored as 'yes' for the QUADAS-2 question, "Did the study avoid inappropriate exclusions?" (low risk of bias for participant selection).
- Studies in which two or more specimen types were evaluated for TB diagnosis. For this analysis, we included studies that we scored as 'yes' for the QUADAS-2 question, "Is the reference standard likely to correctly classify the target condition" (low risk of bias for the reference standard).

Additional analyses

We extracted data on intra- and inter-reader variability. Within the included studies, we recorded whether there was an association between LF-LAM results and disease severity or mortality. We had intended to describe the effect of LF-LAM implementation (that is, LF-LAM used versus LF-LAM not used) on patient important outcomes; however, there were limited data to do this.

Assessment of reporting bias

We did not carry out a formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been helpful for diagnostic test accuracy studies ([Macaskill 2010](#)). However, with our extensive outreach to researchers, we believe reporting bias to be minimal.

Assessment of the quality of the evidence

We assessed the quality of the evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach ([Schünemann 2008](#); [Schünemann 2013](#)). As recommended, we rated the quality of evidence as either high (not downgraded), moderate (downgraded by one level), low (downgraded by two levels), or very low (downgraded by more than two

levels) based on five domains: risk of bias, indirectness, inconsistency, imprecision, and publication bias. For each outcome, the quality of evidence started as high when there were high quality observational studies (cross-sectional or cohort studies) that enrolled participants with diagnostic uncertainty. If we found a reason for downgrading, we used our judgement to classify the reason as either serious (downgraded by one level) or very serious (downgraded by two levels).

Four review authors (MS, CH, ND, and KRS) discussed judgements and applied GRADE in the following way:

- Risk of bias: we used QUADAS-2 to assess risk of bias.
- Indirectness: we considered indirectness from the perspective of test accuracy. We used QUADAS-2 for concerns of applicability and looked for important differences between the populations studied (for example, in the spectrum of disease), the setting, and the review questions.
 - Inconsistency: GRADE recommends downgrading for unexplained inconsistency in sensitivity and specificity estimates. We carried out pre-specified analyses to investigate potential sources of heterogeneity and did not downgrade when we felt we could explain inconsistency in the accuracy estimates.
 - Imprecision: we considered a precise estimate to be one that would allow a clinically meaningful decision. We considered the width of the CrI, and asked ourselves, “Would we make a different decision if the lower or upper boundary of the CrI represented the truth?” In addition, we worked out projected ranges for TP, FN, TN, and FP for a given prevalence of TB and made judgements on imprecision from these calculations.
 - Publication bias: we rated publication bias as undetected

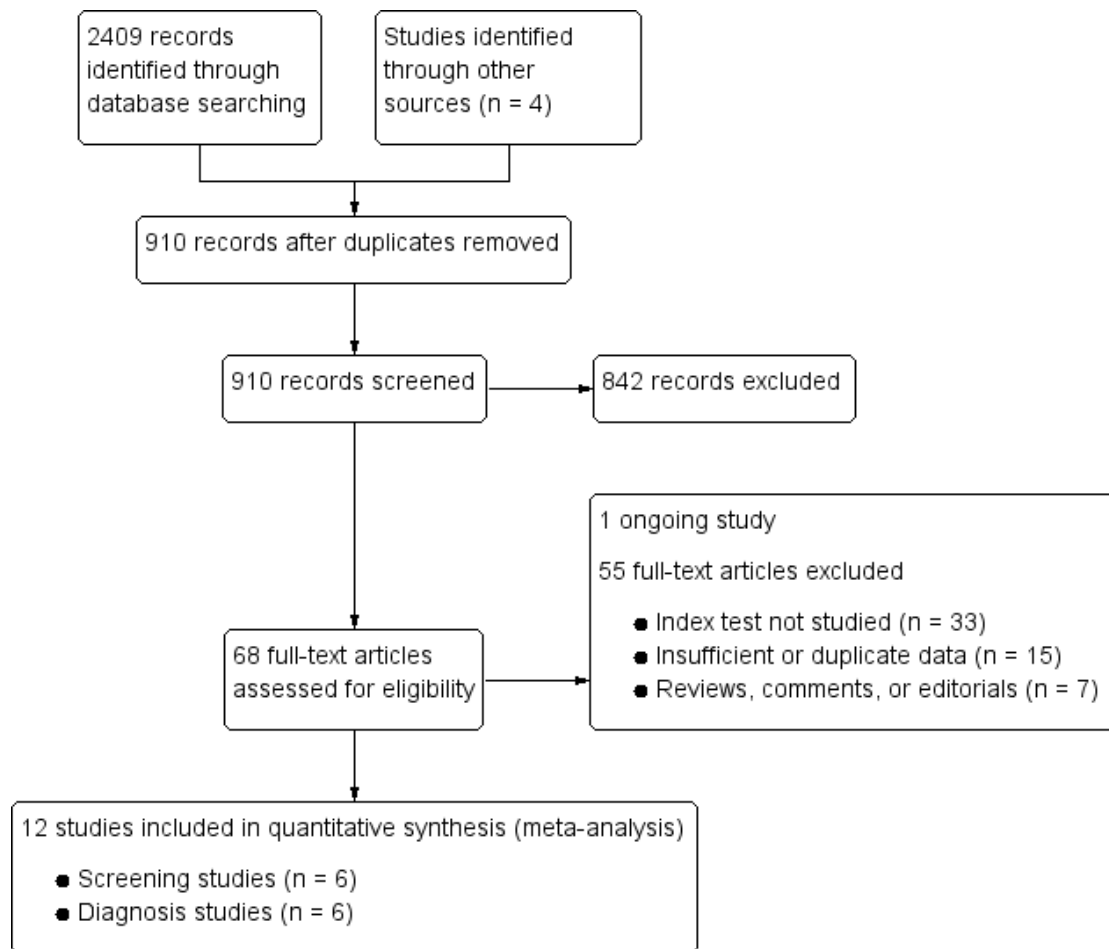
(not serious) for several reasons including the comprehensiveness of the literature search, communication with Alere, the test manufacturer, and extensive outreach to TB researchers to identify studies.

RESULTS

Results of the search

Our search yielded 2409 records. We identified four additional studies through contact with experts. After we removed duplicates, we identified 910 records. We excluded 842 records based on a review of title, or abstract, or both. We retrieved 68 full-text articles, of which we excluded 55 studies for the following reasons: index test not evaluated (33 studies); insufficient data (lack of sufficient data to derive 2 x 2 tables or assess test performance in participants with and without TB) or duplicate data (15 studies); and editorials, reviews, or comments (seven studies). In addition, we excluded one ongoing study ([Grant ongoing](#)) which remained unpublished as of December 2015. Therefore, we included 12 unique studies in this review ([Andrews 2014](#); [Balcha 2014](#); [Bjerrum 2015](#); [Drain 2014b](#); [Drain 2014a](#); [Drain 2014c](#); [LaCourse 2015](#); [Lawn 2012a](#); [Lawn 2014a](#); [Nakiyingi 2014](#); [Peter 2012a](#); [Peter 2015](#)). We listed the excluded studies and reasons for their exclusion in the ‘[Characteristics of excluded studies](#)’ section. [Figure 2](#) shows the flow of studies in the review.

Figure 2. Study flow diagram.






All included studies were performed in low- or middle-income countries. We noted substantial differences in the included studies for the following characteristics: purpose for which lateral flow urine lipoarabinomannan assay (LF-LAM) was applied (diagnosis versus screening); setting (inpatient versus outpatient); threshold used to define LF-LAM positivity (grade 1 versus grade 2); inclusion and exclusion of participants based on whether or not they could produce sputum and whether or not they received an evaluation for extrapulmonary TB; and type of reference standard (microbiological versus composite reference standard) (see the 'Characteristics of included studies' section).

Methodological quality of included studies

For the purpose of TB diagnosis, six studies contributed data (Figure 3). In the patient selection domain, we considered two studies (33%) to be at high risk of bias because: (1) the study excluded all smear-positive participants (Drain 2014c); (2) the study excluded participants who were unable to expectorate sputum (Peter 2015). All studies were cross-sectional or cohort studies. Regarding applicability (i.e. Are there concerns that the included patients do not match the review question?), we judged that all studies included the appropriate participants and settings.

Figure 3. LF-LAM for TB diagnosis. 'Risk of bias' and applicability concerns summary: review authors' judgements about each domain for each included study.

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Andrews 2014	+	+	-	+	+	+	+
Drain 2014c	-	+	-	+	+	+	+
Lawn 2014a	+	+	+	+	+	+	+
Nakiyingi 2014	+	+	+	+	+	+	+
Peter 2012a	+	+	-	-	+	+	-
Peter 2015	-	+	-	+	+	+	+

 High	 Unclear	 Low
---	--	--

In the index test domain, we considered all studies to be at low risk of bias, as all studies used LF-LAM, pre-specified the grade used for positivity, and interpreted the test without knowledge of the results of the reference standard. In addition, we considered the conduct and interpretation of the index test to be of low concern for applicability in all included studies.

In the reference standard domain, we considered four studies (66%) to be at high risk of bias because: (1) the study only conducted mycobacterial blood culture as the reference standard without testing of respiratory specimens (Andrews 2014); (2) the studies did not include testing of extrapulmonary specimens (Drain 2014c; Peter 2015); and (3) health providers selected the sites for




testing based on their own clinical suspicion (Peter 2012a). In terms of applicability, we deemed one study at high concern because of a lack of study directed testing (Peter 2012a). In this study health providers selected the sites for testing based on their own clinical suspicion, and it was unclear if their choice of reference standard would correctly classify TB.

In the flow and timing domain, we considered one study (17%) to be at high risk of bias because not all participants received the same reference standard (Peter 2012a). We judged the remaining studies to be at low risk of bias because all participants received the index test and the same reference standard and no participants were excluded from the 2 x 2 table.

For the purpose of TB screening, six studies contributed data (Figure 4). In the patient selection domain, we considered two studies (33%) to be at high risk of bias because these studies excluded participants who were unable to expectorate sputum and made no attempts at sputum induction (Balcha 2014; Bjerrum 2015). All studies were cross-sectional or cohort studies. Regarding applicability, we judged that all studies (100%) included the appropriate participants and settings.

Figure 4. LF-LAM for TB screening. 'Risk of bias' and applicability concerns summary: review authors' judgements about each domain for each included study.

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Balcha 2014	-	+	+	-	+	+	+
Bjerrum 2015	-	+	-	+	+	+	+
Drain 2014a	+	+	-	+	+	+	+
Drain 2014b	+	+	-	+	+	+	+
LaCourse 2015	+	+	-	+	+	+	+
Lawn 2012a	+	+	-	+	+	+	+

 High	 Unclear	 Low
--	---	---

In the index test domain, we considered all studies at low risk of bias as all studies used LF-LAM, pre-specified the grade used for positivity, and interpreted the test without knowledge of the results of the reference standard. We considered the test conduct and interpretation in all studies (100%) to be applicable.

In the reference standard domain, we considered five studies (83%) to be at high risk of bias because these studies did not conduct microbiological testing on extrapulmonary specimens (Bjerrum 2015; Drain 2014a; Drain 2014b; LaCourse 2015; Lawn 2012a). We judged these studies to be of low concern in terms of applicability.

In the flow and timing domain, we considered one study (17%) to be at high risk of bias because not all participants received the same reference standard (Balcha 2014). We considered five studies (83%) to be of low risk of bias because all participants received the index test and the same reference standard and no participants were excluded from the 2 x 2 table.

Findings

We present pooled sensitivity and specificity of LF-LAM (grade 1 and grade 2) for TB diagnosis in Table 1.

LF-LAM for TB diagnosis among participants with signs or symptoms of TB

Six studies evaluated the accuracy of LF-LAM for TB diagnosis in

participants with signs and symptoms suggestive of TB (Andrews 2014; Drain 2014c; Lawn 2014a; Nakiyingi 2014; Peter 2012a; Peter 2015). All studies provided data at grade 1 (2402 participants; 876 with TB) and five (83%) provided data at grade 2 (2313 participants; 819 with TB). The median CD4 cell count in these studies ranged from 71 to 210 cells per μL . Three studies (50%) were conducted exclusively in an inpatient setting (Andrews 2014; Lawn 2014a; Peter 2012a), one study was conducted exclusively in an outpatient setting (Peter 2015), and two studies were conducted in both inpatient and outpatient settings (Nakiyingi 2014; Drain 2014c).

LF-LAM (grade 2), microbiological reference standard

In the analysis of LF-LAM (grade 2) for TB diagnosis with respect to a microbiological reference standard, we included five studies that included 2313 HIV-positive participants, 819 (35%) with TB (Andrews 2014; Lawn 2014a; Nakiyingi 2014; Peter 2012a; Peter 2015). Sensitivity estimates ranged from 23% to 84%, and specificity estimates from 75% to 99% (Figure 5). Sensitivity was lowest in Peter 2015. Differences between this study and the other studies in this analysis included setting (outpatient only), focus on pulmonary TB (no extrapulmonary samples were taken), and exclusion of participants unable to produce sputum. The pooled sensitivity and specificity (95% CrI) were 45% (29% to 63%) and 92% (80% to 97%).

Figure 5. Forest plots of LF-LAM (Grade 1 and 2) sensitivity and specificity for TB against a microbiological reference standard (TB diagnosis). TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets are the 95% confidence interval (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).

Diagnosis of TB against microbiological reference at Grade 1: all participants

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Andrews 2014	19	37	0	37	1.00 [0.82, 1.00]	0.50 [0.38, 0.62]		
Drain 2014c	24	5	33	28	0.42 [0.29, 0.56]	0.85 [0.68, 0.95]		
Lawn 2014a	64	13	72	264	0.47 [0.38, 0.56]	0.95 [0.92, 0.97]		
Nakiyingi 2014	226	141	141	488	0.62 [0.56, 0.67]	0.78 [0.74, 0.81]		
Peter 2012a	77	43	39	82	0.66 [0.57, 0.75]	0.66 [0.57, 0.74]		
Peter 2015	68	81	113	307	0.38 [0.30, 0.45]	0.79 [0.75, 0.83]		

Diagnosis of TB against microbiological reference at Grade 2: all participants

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Andrews 2014	16	14	3	60	0.84 [0.60, 0.97]	0.81 [0.70, 0.89]		
Lawn 2014a	53	3	83	274	0.39 [0.31, 0.48]	0.99 [0.97, 1.00]		
Nakiyingi 2014	136	21	231	609	0.37 [0.32, 0.42]	0.97 [0.95, 0.98]		
Peter 2012a	58	31	58	94	0.50 [0.41, 0.59]	0.75 [0.67, 0.82]		
Peter 2015	41	27	140	361	0.23 [0.17, 0.29]	0.93 [0.90, 0.95]		

LF-LAM (grade 1), microbiological reference standard

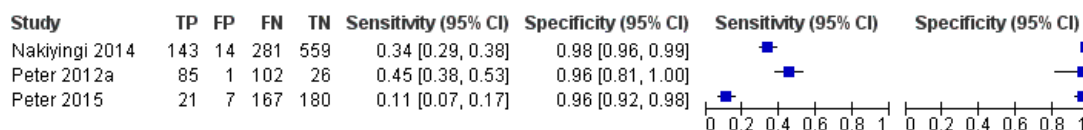
In the analysis of LF-LAM (grade 1) for TB diagnosis, with respect to a microbiological reference standard, we included six studies involving 2402 HIV-positive patients, 876 (36%) with TB (Andrews 2014; Drain 2014c; Lawn 2014a; Nakiyingi 2014; Peter 2012a; Peter 2015). Sensitivity estimates ranged from 38% to 100% and specificity estimates, from 50% to 95% (Figure 5). Sensitivity was lowest in Peter 2015 as discussed above for LF-LAM (grade 2). Specificity was lowest in Andrews 2014, which was notable for the lack of microbiological testing beyond mycobacterial blood

culture and may have led to misclassification of some participants with TB as 'not TB'. The pooled sensitivity and specificity (95% CrI) were 59% (43% to 77%) and 78% (64% to 88%).

LF-LAM (grade 2), composite reference standard

Three studies evaluated LF-LAM (grade 2) for TB diagnosis, with respect to a composite reference standard (Nakiyingi 2014; Peter 2012a; Peter 2015). In these studies, we considered 270 participants to be 'unclassifiable' (54 LF-LAM-positive; 216 LF-LAM-negative). Sensitivity estimates ranged from 11% to 45% and specificity estimates ranged from 96% to 98% (Figure 6).

Figure 6. Forest plots of LF-LAM (grade 2) sensitivity and specificity for TB against a composite reference standard (TB diagnosis). TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets are the 95% confidence interval (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).

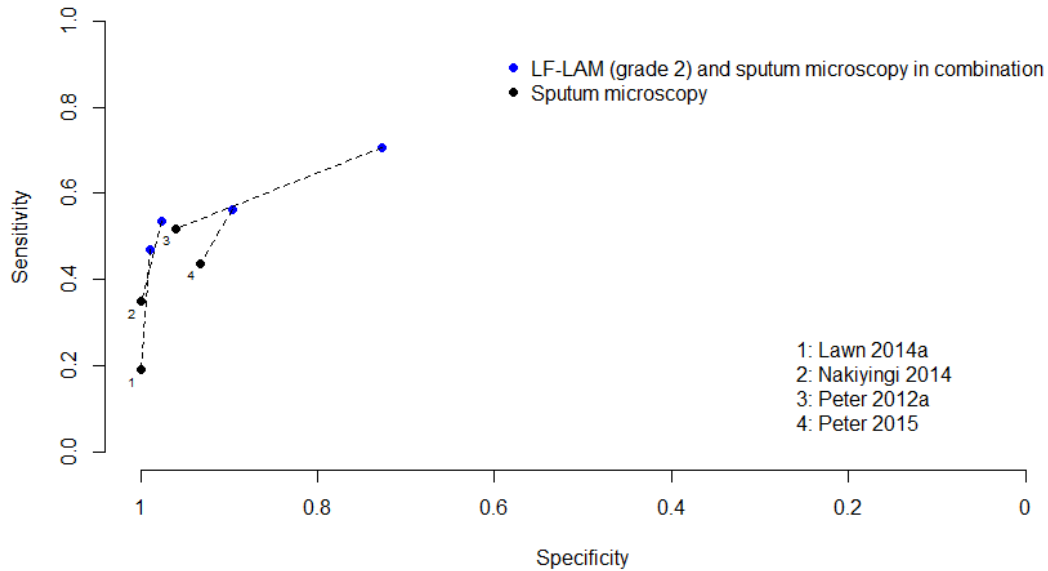


LF-LAM (grade 2) and existing tests

LF-LAM versus sputum microscopy

Four studies including 1876 participants (708 (38%) with TB), directly compared LF-LAM and sputum microscopy in the same patients (Lawn 2014a; Nakiyingi 2014; Peter 2012a; Peter 2015). (We classified participants unable to produce sputum as negative by sputum testing). Pooled sensitivities were similar, 38% (34% to 42%) for LF-LAM and 40% (27% to 54%) for sputum microscopy. Pooled specificity was 95% (94% to 97%) for LF-LAM and 98% (93% to 100%) for sputum microscopy (Figure 7).

Figure 7. Accuracy of smear microscopy (black circle) and smear microscopy in combination with LF-LAM (Grade 2) (blue circle) plotted in receiver operating characteristic (ROC) space. The dashed lines are only for illustration to show how sensitivity and specificity changed with the addition of the LF-LAM test.



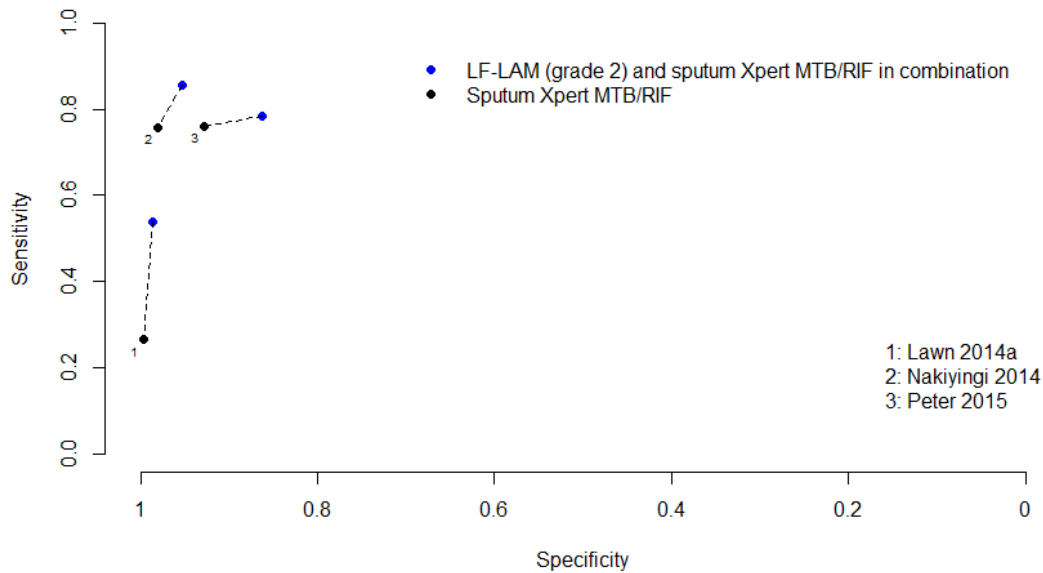
LF-LAM in combination with sputum microscopy

The pooled sensitivity of a combination of LF-LAM and sputum microscopy (either test positive) was 59% (47% to 70%), representing a 19% (4% to 36%) increase over sputum microscopy alone, while the pooled specificity was 92% (73% to 97%), representing a 6% (1% to 24%) decrease from sputum microscopy alone (Figure 7).

LF-LAM versus sputum Xpert® MTB/RIF

Three studies including 909 participants (327 (36%) with TB), directly compared LF-LAM and sputum Xpert® MTB/RIF in the same patients (Lawn 2014a; Nakiyingi 2014; Peter 2015). (We classified participants unable to produce sputum as negative by sputum testing). The pooled sensitivity of LF-LAM was 36% (31% to 42%), lower than the pooled sensitivity of Xpert® MTB/RIF of 61% (39% to 77%). Pooled specificities were similar, 96% (94% to 98%) for LF-LAM and 97% (94% to 99%) for Xpert® MTB/RIF (Figure 8).

Figure 8. Accuracy of Xpert® MTB/RIF (black circle) and Xpert® MTB/RIF in combination with LF-LAM (grade 2) (blue circle) plotted in receiver operating characteristic (ROC) space. The dashed lines are only for illustration to show how sensitivity and specificity changed with the addition of the LF-LAM test.



LF-LAM in combination with sputum Xpert® MTB/RIF

The pooled sensitivity of a combination of LF-LAM and sputum Xpert® MTB/RIF (either test positive) was 75% (61% to 87%) representing a 13% (1% to 37%) increase over Xpert® MTB/RIF alone, while the pooled specificity was 93% (81% to 97%), representing a 4% (1% to 16%) decrease from Xpert® MTB/RIF alone (Figure 8).

results. Nakiyingi 2014 reported that a valid LF-LAM result was obtained on the first attempt for all tests (997/997, 100%). Peter 2015 reported that fewer than 1% of LF-LAM strip tests failed on the first attempt and required usage of a second strip to produce valid results.

Uninterpretable index test results

Studies reported few uninterpretable test results. Peter 2012a reported that 1% to 2% of 423 tests remained 'indeterminate' after repeat testing. Lawn 2014a reported no invalid or uninterpretable

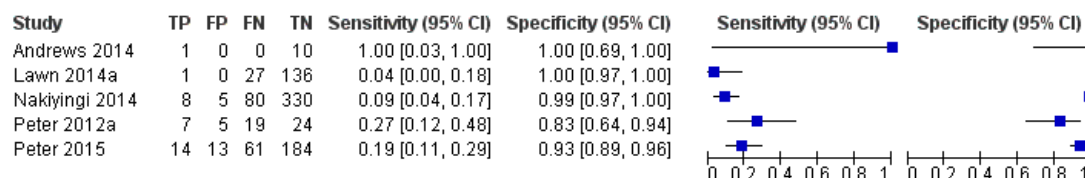
Investigations of heterogeneity

LF-LAM (grade 2), stratified by CD4 count

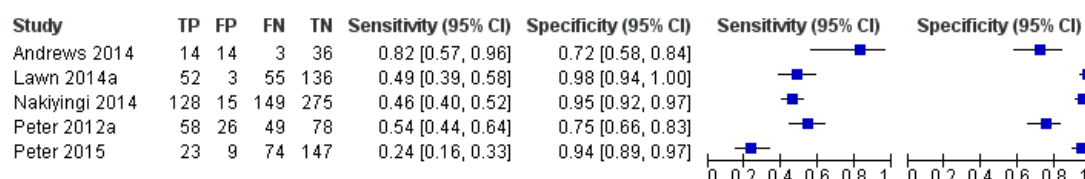
We included five studies in the analyses by CD4 count (Andrews 2014; Lawn 2014a; Nakiyingi 2014; Peter 2012a; Peter 2015). See Figure 9 and Figure 10.

Figure 9. Forest plots of LF-LAM (grade 2) sensitivity and specificity for TB against a microbiological reference standard, by CD4 strata (TB Diagnosis). TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets are the 95% confidence interval (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).

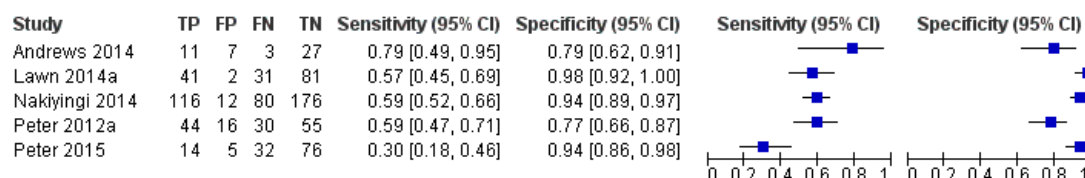
Diagnosis of TB against microbiological reference at Grade 2, CD4 > 200



Diagnosis of TB against microbiological reference at Grade 2, CD4 < 200



Diagnosis of TB against microbiological reference at Grade 1, CD4 < 100



Diagnosis of TB against microbiological reference at Grade 1, CD4 < 50

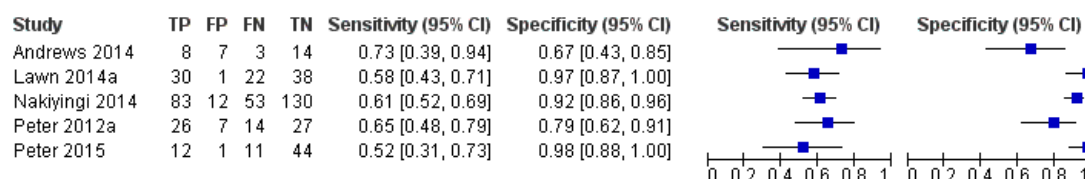
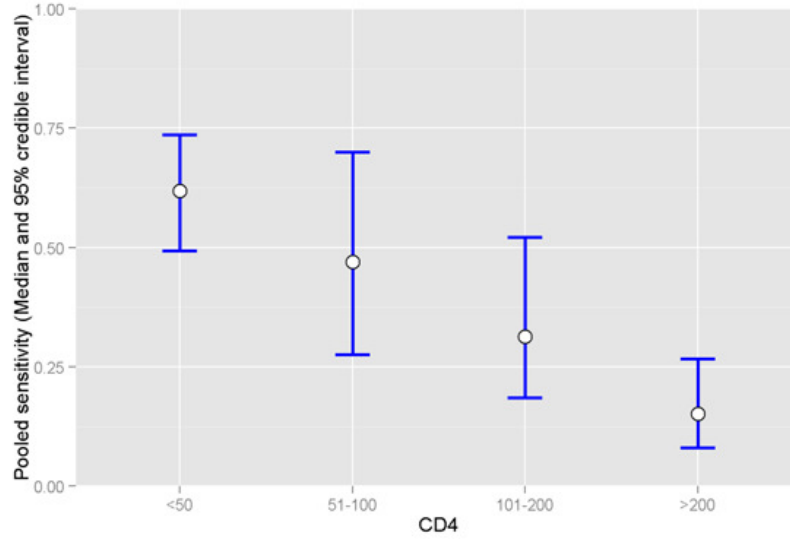
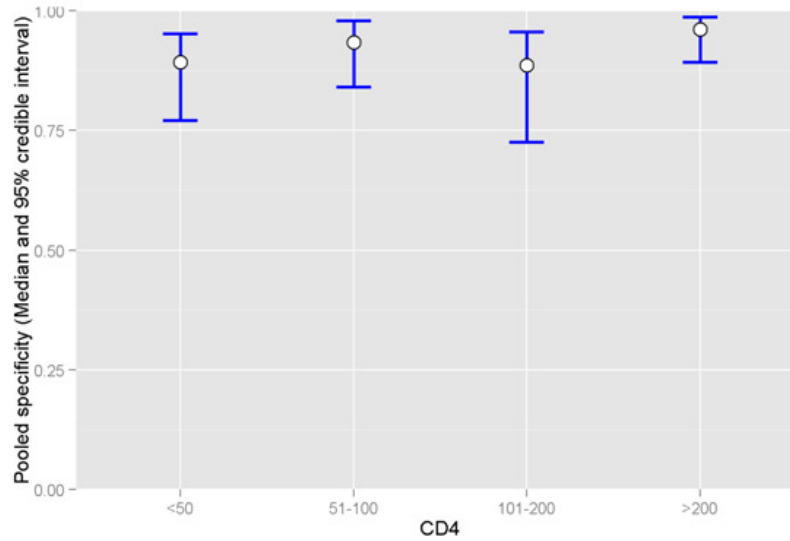


Figure 10. Plots of (A) sensitivity and (B) specificity of LF-LAM (grade 2) for TB diagnosis stratified by CD4 count. Plots were derived using a sample from the posterior distribution of the pooled sensitivity (A) and specificity (B). Circle represents the pooled estimates (median), with bars representing 95% credible intervals.

A. Sensitivity



B. Specificity



CD4 > 200 cells per μ L

The five studies that evaluated participants with CD4 > 200 cells per μ L included 925 HIV-positive participants, 218 (24%) with TB. Sensitivity estimates ranged from 4% to 100% and specificity estimates ranged from 83% to 100%. The pooled sensitivity and specificity (95% CrI) were 15% (8% to 27%) and 96% (89% to 99%). When we limited the analysis to studies involving inpatients (excluding Peter 2015), the pooled sensitivity and specificity were 14% (6% to 33%) and 97% (88% to 99%).

CD4 \leq 200 cells per μ L

The five studies that evaluated participants with CD4 \leq 200 cells per μ L included 1344 HIV-positive participants, 605 (45%) with TB. Sensitivity estimates ranged from 24% to 82% and specificity estimates ranged from 72% to 98%. The pooled sensitivity and specificity (95% CrI) were 50% (95% CrI 35% to 67%) and 90% (95% CrI 78% to 95%). When we limited the analysis to studies involving inpatients (excluding Peter 2015), the pooled sensitivity and specificity were 56% (95% CrI 42% to 71%) and 88% (95% CrI 70% to 95%). The probability that LAM sensitivity and specificity are higher among participants with CD4 \leq 200 compared to those with CD4 > 200 were 100% and 6%, respectively.

CD4 \leq 100 cells per μ L

The five studies that evaluated participants with CD4 \leq 100 cells per μ L included 859 HIV-positive participants, 402 (47%) with TB. Sensitivity estimates ranged from 30% to 79% and specificity estimates ranged from 79% to 98%. The pooled sensitivity and specificity (95% CrI) were 56% (41% to 70%) and 90% (81% to 95%). When we limited the analysis to studies involving inpatients

(excluding Peter 2015), the pooled sensitivity and specificity were 62% (48% to 75%) and 89% (75% to 95%). The probability that LAM sensitivity and specificity are higher among participants with CD4 \leq 100 compared to those with CD4 > 100 were 99% and 33%, respectively.

CD4 \leq 50 cells per μ L

The five studies that evaluated participants with CD4 \leq 50 cells per μ L included 543 HIV-positive participants, 262 (48%) with TB. Sensitivity estimates ranged from 52% to 73% and specificity estimates ranged from 67% to 98%. The pooled sensitivity and specificity (95% CrI) were 62% (49% to 73%) and 89% (77% to 95%). When we limited the analysis to studies involving inpatients (excluding Peter 2015), the pooled sensitivity and specificity were 63% (49% to 76%) and 86% (71% to 94%). The probability that LAM sensitivity and specificity are higher among participants with CD4 \leq 50 compared to those with CD4 > 50 were 98% and 25%, respectively.

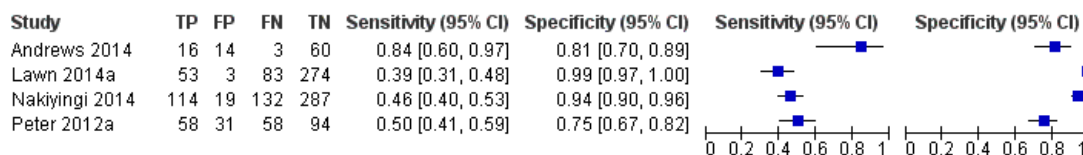
LF-LAM (grade 2), stratified by clinical setting

Inpatients

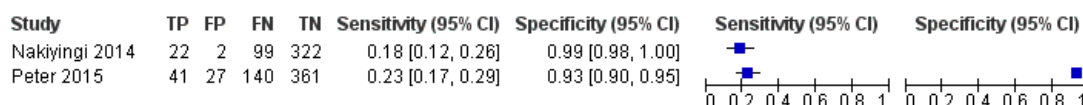
In the analysis of LF-LAM (grade 2) for TB diagnosis among inpatients, we included four studies involving 1299 HIV-positive inpatients, 517 (40%) with TB (Andrews 2014; Lawn 2014a; Nakiyingi 2014; Peter 2012a). Sensitivity estimates ranged from 39% to 84% and specificity estimates ranged from 75% to 99% (Figure 11). The pooled sensitivity and specificity (95% CrI) were 53% (38% to 70%) and 90% (73% to 96%).

Figure 11. Forest plots of LF-LAM (grade 2) sensitivity and specificity for TB against a microbiological reference standard, by health care setting (TB Diagnosis). TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets are the 95% confidence interval (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).

Diagnosis of TB against microbiological reference at Grade 2: inpatients



Diagnosis of TB against microbiological reference at Grade 2: outpatients



Outpatients

We identified two studies that evaluated LF-LAM (grade 2) for TB diagnosis among outpatients (Nakiyingi 2014; Peter 2015). Sensitivity estimates ranged from 18% to 23% and specificity estimates ranged from 93% to 99% (Figure 11).

Sensitivity analyses

LF-LAM (grade 2) for TB diagnosis

In the analysis of LF-LAM (grade 2) for TB diagnosis with respect to a microbiological reference standard, when we limited the studies to those at low risk of bias for patient selection (excluding Peter 2015; four studies, 1744 HIV-positive participants, 638 (36%) with TB) pooled sensitivity and specificity were 50% (35% to 69%) and 91% (75% to 97%). There were insufficient data to carry out a sensitivity analysis to assess the robustness of meta-analyses based on the quality of the reference standard.

LF-LAM for TB screening among participants irrespective of signs and symptoms for TB

Six studies evaluated the accuracy of LF-LAM for screening in a target population irrespective of TB symptoms (Balcha 2014;

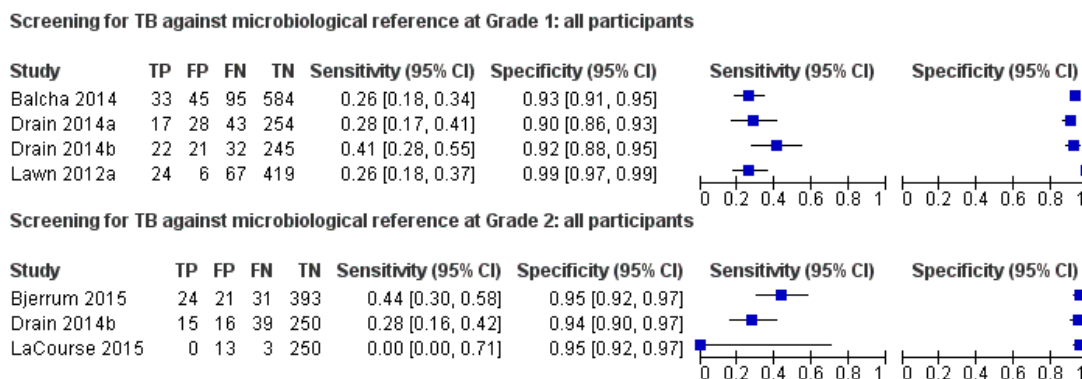
Bjerrum 2015; Drain 2014a; Drain 2014b; LaCourse 2015; Lawn 2012a). Four (67%) studies provided data at grade 1 (1935 participants; 333 (17%) with TB) (Balcha 2014; Drain 2014a; Drain 2014b; Lawn 2012a), and three (43%) studies provided data at grade 2 (1055 participants; 112 (11%) with TB) (Bjerrum 2015; Drain 2014b; LaCourse 2015). The median CD4 in these studies ranged from 127 to 437 cells per μL . All studies were carried out exclusively or largely in an outpatient setting; a single study, Bjerrum 2015, included a minority of participants from an inpatient setting.

We had limited data for LF-LAM for TB screening as described below.

LF-LAM (grade 2), microbiological reference standard

Three studies evaluated LF-LAM (grade 2) for TB screening (1055 HIV-positive participants, 112 (11%) with TB), with respect to a microbiological reference standard (Bjerrum 2015; Drain 2014b; LaCourse 2015). Sensitivity estimates ranged from 0% to 44%, and specificity estimates from 94% to 95% (Figure 12). Sensitivity was lowest in LaCourse 2015. Differences between this study and others included low TB prevalence (1%), a study population consisting exclusively of pregnant women, and high median CD4 count.

Figure 12. Forest plots of LF-LAM (grade 1 and 2) sensitivity and specificity for TB against a microbiological reference standard (TB Screening). TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets are the 95% confidence interval (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).



LF-LAM (grade 1), microbiological reference standard

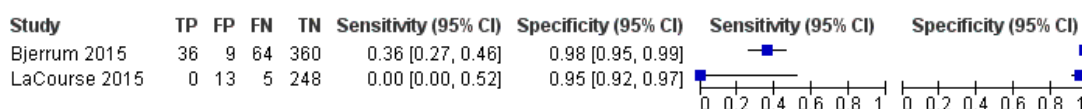
In the analysis of LF-LAM (grade 1) for TB screening with respect to a microbiological reference standard, we included four studies (different from those studies contributing data for LF-LAM grade 2) involving 1935 HIV-positive participants, 333 (17%) with TB (Balcha 2014; Drain 2014a; Drain 2014b; Lawn 2012a). Both sensitivity and specificity estimates were variable; sensitivity estimates ranged from 26% to 41%, and specificity estimates from

90% to 99% (Figure 12). The pooled sensitivity and specificity (95% CrI) were 30% (20% to 43%) and 94% (86% to 97%).

LF-LAM (grade 2), composite reference standard

Two studies evaluated LF-LAM for TB screening, with respect to a composite reference standard (Bjerrum 2015; LaCourse 2015). Sensitivity estimates ranged from 0% to 36%, and specificity estimates ranged from 95% to 98% (Figure 13).

Figure 13. Forest plots of LF-LAM (grade 2) sensitivity and specificity for TB against a composite reference standard (TB Screening). TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets are the 95% confidence interval (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).

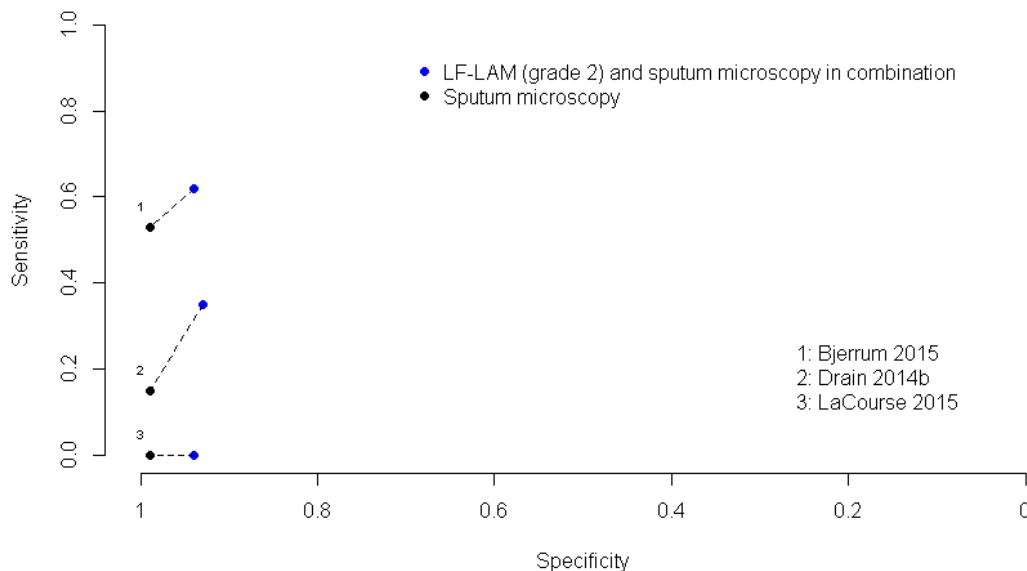


LF-LAM (grade 2) and existing tests

LF-LAM versus sputum microscopy

Three studies directly compared LF-LAM and sputum microscopy in the same participants (Bjerrum 2015; Drain 2014b; LaCourse 2015). The sensitivity of LF-LAM compared to smear in these studies was 44% (30% to 58%) versus 53% (39% to 66%) for Bjerrum 2015; 28% (16% to 42%) versus 15% (7% to 27%) for Drain 2014b; and 0% versus 0%, for LaCourse 2015 (Figure 14).

Figure 14. Accuracy of smear microscopy (black circle) and smear microscopy in combination with LF-LAM (grade 2) (blue circle) plotted in receiver operating characteristic (ROC) space. The dashed lines are only for illustration to show how sensitivity and specificity changed with the addition of the LF-LAM test.



LF-LAM in combination with sputum microscopy

Three studies directly compared LF-LAM and sputum microscopy in the same participants (Bjerrum 2015; Drain 2014b; LaCourse 2015). The sensitivity of LF-LAM combined with smear in these studies (compared to smear alone) was 62% (48% to 75%) versus 53% (39% to 66%) for Bjerrum 2015; 35% (23% to 49%) versus 15% (7% to 27%) for Drain 2014b; and 0% versus 0%, for LaCourse 2015 (Figure 14).

LF-LAM and sputum Xpert® MTB/RIF

We did not have sufficient data to evaluate LF-LAM and sputum Xpert® MTB/RIF (one study, three participants with TB; LaCourse 2015).

Uninterpretable index test results

The included studies reported high levels of test interpretability. Bjerrum 2015 reported that all (100%) 469 LAM tests in their study yielded interpretable results; similarly, LaCourse 2015 reported no uninterpretable results.

There were limited data to evaluate LF-LAM (grade 2) by CD4 thresholds.

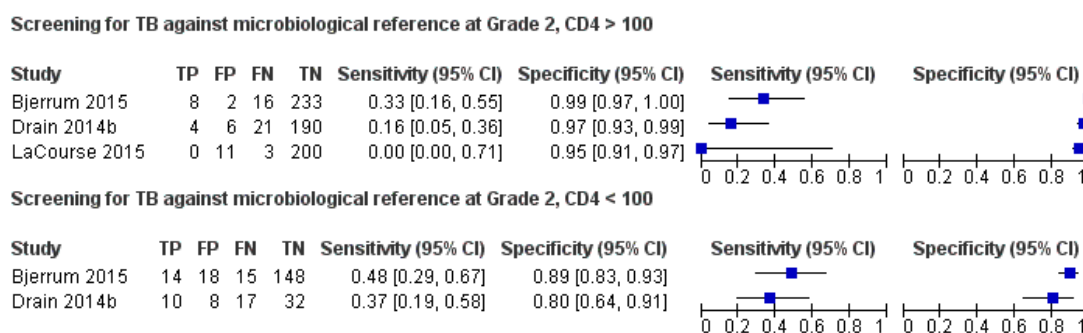
Investigations of heterogeneity

LF-LAM (grade 2), stratified by CD4 count

CD4 > 100 cells per μ L

Three studies evaluated LF-LAM in patients with CD4 > 100 cells per μ L (Bjerrum 2015; Drain 2014b; LaCourse 2015). Sensitivity estimates ranged from 0% to 16% and specificity estimates ranged from 89% to 95% (Figure 15).

Figure 15. Forest plots of LF-LAM (grade 2) sensitivity and specificity for TB (TB Screening). TP = True Positive; FP = stratified by CD4 count. False Positive; FN = False Negative; TN = True Negative. Between brackets are the 95% confidence interval (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).



CD4 ≤ 100 cells per μ L

Two studies evaluated LF-LAM in patients with CD4 ≤ 100 cells per μ L (Bjerrum 2015; Drain 2014b). Sensitivity estimates ranged from 37% to 48% and specificity estimates ranged from 80% to 100% (Figure 15).

Sensitivity analyses, LF-LAM (grade 2) for TB screening

For TB screening, there were insufficient data to carry out sensitivity analyses to assess the robustness of the meta-analyses based on methodological quality items.

Other analyses

Concerning inter-reader variability, there was a high degree of agreement; the included studies reported on inter-reader variability in the form of a Kappa statistic or percent concordance between

multiple readers. Four studies of TB diagnosis reported Kappa statistics ranging from 0.78 to 0.97 (Lawn 2014a; Nakiyingi 2014; Peter 2012a; Peter 2015), and most reported values > 0.92. Two studies of TB screening reported Kappa values of 0.92 to 0.97 (Bjerrum 2015; Lawn 2012a), and one study reported 100% concordance between readers (Balcha 2014). There were limited data on intra-reader agreement; Peter 2012a reported a Kappa statistic of 0.92 to 0.96, which indicated very good agreement.

We acknowledge that patient outcomes are clearly important to patients, decision makers, and the wider TB community. However, we could not systematically address outcomes in addition to diagnostic accuracy as they would have required a different methodology. Nonetheless, we looked for and summarized the data on patient outcomes from the included studies, including data from secondary analyses. Six included studies provided data on the association of LF-LAM and mortality in publications and related reports or in unpublished data (Balcha 2014; Bjerrum 2015; Drain

2014c; Lawn 2012a; Nakiyingi 2014; Peter 2015; see Table 2). These studies were conducted in both inpatient and outpatient settings. Data on patient outcomes were largely restricted to post-hoc analyses. Nonetheless, available data consistently suggested higher disease severity among LF-LAM positive TB participants than LF-LAM negative TB participants. All six studies showed a consistent finding of increasing mortality with LF-LAM positivity, despite considerable variability in the length of follow-up, method of TB diagnosis, and provision of treatment. We noted that the investigators in these studies did not use the results of LF-LAM to decide TB treatment initiation.

Balcha 2014 reported higher mortality (20% versus 3%, $P < 0.001$) in LF-LAM positive than LF-LAM negative participants. Similarly Manabe 2014 reported higher mortality (28% versus 13%, $P = 0.035$) in LF-LAM positive than LF-LAM negative TB participants (secondary analysis, Nakiyingi 2014). Manabe 2014 additionally found higher mortality in LF-LAM positive participants without microbiological evidence of TB (34% versus 19% for LF-LAM positive and LF-LAM negative participants, respectively). Lawn 2012c found that among 23 TB participants who were LF-LAM positive, five people died (22%) compared to zero deaths (0/36) among TB participants who were LF-LAM negative (secondary analysis, Lawn 2012a). In another secondary analysis (Lawn 2012a), Lawn 2013b reported that LF-LAM sensitivity was 100% among TB participants who died compared to 25% among TB participants who were alive at 90 days ($P = 0.002$). Peter 2015 reported mortality of 25% and 11% in LF-LAM positive and LF-LAM negative participants, respectively. Finally, Bjerrum 2015 reported that among all participants, 49% of those who were LF-LAM-positive died versus 14% of those who were LF-LAM negative ($P < 0.001$). Among TB participants, 54% of LF-LAM-positive participants died compared to only 16% of those who were

LF-LAM negative ($P = 0.003$).

Bjerrum 2015 and Drain 2015 (related report, Drain 2014c) included treatment data in their studies. Bjerrum 2015 reported that among TB participants who received TB treatment, 31% of those who were LF-LAM positive died compared to only 4% of those who were LF-LAM negative. Among TB participants who did not receive treatment at the time of assessment in the study, 100% of those who were LF-LAM positive died compared to 33% of those who were LF-LAM negative. In each of these studies, LF-LAM results were unavailable to clinicians. In another post-hoc analysis, Peter 2013 reported that among inpatients, LF-LAM-positive TB participants missed by empirical early treatment had lower CD4 counts and higher median illness severity scores, compared to participants who received early treatment based on clinical decision making. Drain 2015 reported LF-LAM responses over time. They reported that among participants receiving TB therapy, having a positive LF-LAM test at the two-month visit was associated with an adjusted hazard ratio (HR) of 5.58 for mortality (median follow-up time of 49 months) compared to participants with a negative LF-LAM test at the two-month visit. Participants with a positive LF-LAM at six months had an adjusted HR of 42.1 for mortality during study follow-up. They found no difference (adjusted HR 1.41, $P = 0.49$) in mortality comparing baseline LF-LAM results. Concerning the impact of LF-LAM implementation on patient outcomes, Peter 2016 found that, in HIV-positive inpatients, LF-LAM in combination with routine diagnostic tests to guide the rapid initiation of TB treatment was associated with a relative risk reduction of 17% (95% CI 4% to 28%) in eight-week mortality compared with routine diagnostic tests alone (no LF-LAM). We are also aware of two additional studies with data on the impact of LF-LAM on patient outcomes expected in 2016 (Grant ongoing; NCT01990274).

Summary of findings

Review question: what is the diagnostic accuracy of LF-LAM (grade 2) for diagnosing tuberculosis (TB) in adults living with HIV?

Participants/population: HIV-positive adults with symptoms of TB

Index test: LF-LAM (grade 2)

Role: a replacement test or additional test along with sputum smear microscopy or sputum Xpert® MTB/RIF

Reference standard: microbiological (mainly mycobacterial culture)

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: 0.45 (95% CrI: 0.29 to 0.63); **pooled specificity:** 0.92 (95% CrI: 0.80 to 0.97)

Test result	Number of results per 1000 participants tested (95% CrI)		Number of participants (studies)	Quality of the evidence (GRADE)
	Prevalence 10%	Prevalence 30%		
True positives (patients correctly classified as having TB)	45 (29 to 63)	135 (87 to 189)	819 (5)	⊕⊕⊕○ moderate ^{1,2,3}
False negatives (patients incorrectly classified as not having TB)	55 (37 to 71)	165 (111 to 213)		
True negatives (patients correctly classified as not having TB)	828 (720 to 873)	644 (560 to 679)	1494 (5)	⊕⊕○○ low ^{2,4,5}
False positives (patients incorrectly classified as having TB)	72 (27 to 180)	56 (21 to 140)		
Inconclusive	Infrequent			
Complications	LF-LAM is a non-invasive urine test with no known complications			

Abbreviations: CrI: credible interval; LF-LAM: lateral flow urine lipoarabinomannan assay; GRADE: Grading of Recommendations Assessment, Development and Evaluation; HIV: human immunodeficiency virus; TB: tuberculosis

GRADE quality of evidence ([GRADEpro 2015](#); [Schünemann 2013](#))

High quality: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate quality: we are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low quality: our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.

Very low quality: we have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

The table displays normalised frequencies within a hypothetical cohort of 1000 patients at two different TB prevalences (pre-test probabilities): 10% and 30%. Credible limits were estimated based on those around the point estimates for pooled sensitivity and specificity

Note: the results on this table should not be interpreted in isolation from the results of the individual included studies contributing to each summary test accuracy measure. These are reported in the main body of the text of the review

¹We used QUADAS-2 to assess risk of bias. We considered one study to be at high risk of bias because this study excluded patients who were unable to produce sputum ([Peter 2015](#)). We did not downgrade the evidence.

²Four included studies were performed in inpatient settings ([Andrews 2014](#); [Lawn 2014a](#); [Nakiyingi 2014](#); [Peter 2012a](#)).

³The wide 95% CrI for true positives and false negatives may lead to different decisions depending on which credible limits are assumed. We downgraded the evidence by one level. Of note, in a subgroup analysis of patients with CD4 count ≤ 100 cells per μL , the pooled sensitivity increased to 56% (95% CrI 41% to 70%) while the pooled specificity decreased slightly to 90% (95% CrI 81% to 95%; five studies, 859 participants, 402 with confirmed TB).

⁴We used QUADAS-2 to assess risk of bias. We considered three studies to be at high risk of bias because they used a lower quality reference standard ([Andrews 2014](#); [Peter 2012a](#); [Peter 2015](#)). We downgraded the evidence by one level.

⁵The wide 95% CrI for true negatives and false positives may lead to different decisions depending on which credible limits are assumed. We downgraded the evidence by one level.

<p>Review question: what is the accuracy of LF-LAM (grade 2) for screening for tuberculosis (TB) in adults living with HIV? Patients/population: HIV-positive adults regardless of signs and symptoms of TB Index test: LF-LAM (grade 2) Role: a replacement test or additional test along with sputum smear microscopy or sputum Xpert® MTB/RIF Reference standard: microbiological (mainly mycobacterial culture) Studies: cross-sectional Setting: outpatient Limitations: the main limitations of the review were the use of a lower quality reference standard in all included studies, and the small number of studies and participants included in the analyses Sensitivity range: 0% to 44%; specificity range: 94% to 95%</p>				
Test result	Number of results per 1000 participants tested		Number of participants (studies)	Quality of the evidence (GRADE)
	Prevalence 1%	Prevalence 10%		
True positives (patients correctly classified as having TB)	0 to 4	1 to 44	112 (3)	⊕○○○ very low ^{1,2}
False negatives (patients incorrectly classified as not having TB)	6 to 10	56 to 99		
True negatives (patients correctly classified as not having TB)	931 to 941	846 to 855	943 (3)	⊕⊕○○ low ¹
False positives (patients incorrectly classified as having TB)	49 to 59	45 to 54		
Inconclusive	Infrequent			
Complications	LF-LAM is a non-invasive urine test with no known complications			

Abbreviations: LF-LAM: lateral flow urine lipoarabinomannan assay; HIV: human immunodeficiency virus; GRADE: Grading of Recommendations Assessment, Development and Evaluation; TB: tuberculosis

GRADE quality of evidence ([GRADEpro 2015](#); [Schünemann 2013](#))

High quality: We are very confident that the true effect lies close to that of the estimate of the effect.

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

The table displays normalised frequencies within a hypothetical cohort of 1000 patients at two different TB prevalences (pre-test probabilities): 1% and 10%. Credible limits were estimated based on those around the point estimates for pooled sensitivity and specificity

Note: the results on this table should not be interpreted in isolation from the results of the individual included studies contributing to each summary test accuracy measure. These are reported in the main body of the text of the review

LF-LAM for TB screening was often performed at grade 1, which is no longer recommended. Nonetheless, we found pooled sensitivity and pooled specificity of 30% (95% CrI 20% to 43%) and 94% (95% CrI 86% to 97%), four studies, 1935 participants, 333 (17%) TB.

¹We used QUADAS-2 to assess risk of bias. We considered one study to be at high risk of bias because the study excluded patients who were unable to produce sputum. All studies used a lower quality reference standard. We downgraded the evidence by two levels.

²Sensitivity estimates were variable. We downgraded the evidence by one level.

DISCUSSION

This systematic review summarizes the current literature and includes 12 unique studies on the accuracy of lateral flow urine lipoarabinomannan assay (LF-LAM) for tuberculosis (TB) in people with human immunodeficiency virus (HIV). Six studies used LF-LAM for TB diagnosis in participants with signs and symptoms of TB. The studies of TB diagnosis were largely conducted in inpatient settings and had high TB prevalence. Six studies used LF-LAM for systematic screening of participants for TB regardless of their symptoms. The studies of TB screening were conducted exclusively or predominantly in outpatient settings and, compared to studies of TB diagnosis, had lower TB prevalence and involved patients with higher CD4 counts. All studies were conducted in low- and middle-income countries.

LF-LAM for TB diagnosis

See 'Summary of findings' table 1 ([Summary of findings 1](#)). For TB diagnosis, the pooled sensitivity of LF-LAM (grade 2) was low (45%) with relatively high pooled specificity (92%). We explored an alternative threshold (grade 1) for test positivity and found that both sensitivity (59%) and specificity (78%) were low. We found higher sensitivity (59%) when LF-LAM was combined with microscopy (either test positive) which represented an increase of 19% over sputum microscopy alone, although pooled specificity decreased 6% from microscopy alone (four studies). In planned investigations, we observed an inverse correlation between LF-LAM sensitivity and CD4 count, with increasing sensitivity as patient CD4 count decreased (increased from 15% in patients with CD4 cell count > 200 cells per μL to 62% in patients with CD4 \leq 50 cells per μL). This inverse correlation may reflect increased LF-LAM sensitivity among those with advanced disease, disseminated disease, or higher bacillary burden. Pooled specificity was similar across CD4 strata. These findings suggest that LF-LAM may aid in the diagnosis of TB in seriously ill patients with low CD4 counts. While the test does not identify all TB cases, it may be of particular value in diagnosing seriously ill patients who are unable to produce sputum or cannot be diagnosed with other TB diagnostic tests. We a priori planned to investigate and expected to find higher accuracy in patients with lower CD4 counts and patients who were hospitalized. In subgroup analyses, we found LF-LAM (grade 2) sensitivity was generally higher among inpatients (53%) than among outpatients (18% to 23%). Although subgroup comparisons in diagnostic accuracy reviews are observational and suffer from the same limitations as all observational findings (for example, confounding between characteristics), there is a scientific rationale for this finding in that inpatients are likely to have higher disease severity or higher bacillary burden.

LF-LAM for screening for TB

See 'Summary of findings' table 2 ([Summary of findings 2](#)). Few studies evaluated LF-LAM for TB screening among individuals irrespective of symptoms at the currently recommended grade 2 threshold for positivity. Sensitivity ranged from 0% to 44%, while specificity ranged from 94% to 95%. We explored an alternative threshold (grade 1) for positivity and found low pooled sensitivity (30%) and reasonably high pooled specificity (94%).

Summary of main results

We have summarized the main results in the 'Summary of findings' tables ([Summary of findings 1](#); [Summary of findings 2](#)).

- For TB diagnosis in HIV-positive people with signs and symptoms suggestive of TB, the pooled sensitivity and specificity (95% credible interval (CrI)) of LF-LAM (grade 2) were 45% (29% to 63%) and 92% (80% to 97%).
- For TB screening in HIV-positive people regardless of signs and symptoms of TB, LF-LAM (grade 2) sensitivity estimates ranged from 0% to 44% and specificity estimates from 94% to 95%.
- Few studies used a higher quality reference standard with more than one type of clinical specimen.

We note that the band intensity corresponding to grade 2 in this review corresponds to the current manufacturer threshold for positivity (that is, grade 1 on the new manufacturer reference card).

Application of the meta-analysis to a hypothetical cohort

LF-LAM for TB diagnosis

See 'Summary of findings' table 1 ([Summary of findings 1](#)). If the pooled sensitivity and specificity estimates for LF-LAM (grade 2) are applied to a hypothetical cohort of 1000 HIV-positive individuals where 300 (30%) of those with symptoms actually have TB, LF-LAM would correctly identify 135 people as having TB while missing 165 people with TB. In this same population of 1000 people, where 700 do not have TB, the test will correctly classify 644 people as not having TB while misclassifying 56 as having TB.

LF-LAM for TB screening

See 'Summary of findings' table 2 ([Summary of findings 2](#)). If the ranges in sensitivity and specificity for LF-LAM (grade 2) are applied to a hypothetical cohort of 1000 HIV-positive individuals where 10 (1%) of those being screened for TB actually have TB, LF-LAM would correctly identify zero to four people as having TB while missing six to 10 people with TB. In this same population of 1000 people, where 990 do not have TB, the test will correctly

classify 931 to 941 people as not having TB while misclassifying 49 to 59 as having TB.

Strengths and weaknesses of the review

The findings in this review are based on comprehensive searching, strict inclusion criteria, and standardized data extraction. The strength of our review is that it enabled an assessment of the accuracy of LF-LAM in HIV-positive people for two purposes, diagnosis and screening. We determined test accuracy at two different positivity cut-offs and for LF-LAM combined with existing sputum-based tests. We also investigated heterogeneity with respect to the CD4 count. 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. The results should, therefore, be interpreted with caution.

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. However, as diagnostic accuracy studies are poorly indexed, we acknowledge that we may have missed some studies despite the comprehensive search.

Accuracy of the reference standards used

In a diagnostic test accuracy systematic review, the reference standard is the best available test to determine the presence or absence of the target condition. A microbiological reference standard, primarily culture, is considered the best reference standard for TB. 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. The higher quality reference standard is better at classifying which patients have and do not have TB. A lower quality reference standard may miss some TB cases and classify some TB patients as not having TB. This may make a truly positive LF-LAM result seem like an FP leading to an underestimation of specificity. However, in the review, only two studies (33%) for TB diagnosis and one study (17%) for TB screening used a higher quality reference standard.

Compared with a microbiological reference standard, a composite reference standard using clinical information may also be less likely to misclassify TB patients as not having TB, especially in patients

with paucibacillary pulmonary disease in whom sputum culture may be negative. We found LF-LAM specificity ranged from 96% to 99% for TB diagnosis with respect to a composite reference standard. However, we had limited data and these findings should be interpreted with caution.

Quality and quality of reporting of the included studies

There were few studies that included extrapulmonary specimens in addition to sputum, and few studies that included participants unable to expectorate sputum. We had limited data to address these quality items in sensitivity analyses and acknowledge that these features may have contributed to risk of bias in the accuracy estimates. For TB diagnosis, in the patient selection domain, we considered two studies (33%) to be at high risk of bias, and in the reference standard domain, we considered four studies (66%) to be at high risk of bias. For TB screening, in the patient selection domain, we considered two studies (33%) to be at high risk of bias and in the reference standard domain, we considered five studies (83%) to be at high risk of bias. Using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, we judged the evidence for diagnostic accuracy of LF-LAM to be of low or very low quality. This means that our confidence in the effect estimate is limited and the true effect may be substantially different from the estimate of the effect. In general, studies were fairly well reported, though we corresponded with almost all study authors for additional data and missing information.

Applicability of findings to the review question

We had low concern about the applicability of the included studies to our review question as assessed by the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. We considered the degree to which setting, patient spectrum, index test, and reference standard characteristics could affect test accuracy estimates. Although all studies were performed in low- or middle-income countries, we think, given the underlying pathophysiology of HIV-associated TB, LF-LAM would perform similarly in high-income countries. This review evaluated LF-LAM at the current manufacturer-recommended cut-off for positivity, as well as an alternative threshold. Therefore the findings in this review should be considered applicable to the test. However, it is important to note that this review assessed sensitivity and specificity in applied research settings. Although the participant characteristics and settings matched our review question in most cases, as studies were carried out under research conditions, it is possible that the accuracy of LF-LAM may be lower in routine practice settings.

AUTHORS' CONCLUSIONS

Implications for practice

Among HIV-positive people who present with symptoms suggestive of TB, this Cochrane review found low LF-LAM sensitivity, which suggests that LF-LAM cannot be relied on alone for the diagnosis of TB. For TB diagnosis, the combination of LF-LAM with sputum microscopy suggests increased sensitivity for TB compared to either test alone, but with a decrease in specificity. The findings also suggest increased sensitivity at lower CD4 counts compared to higher CD4 counts. However, we had limited data and these findings should be interpreted with caution. Nonetheless, in HIV-positive people with low CD4 counts who are seriously ill, LF-LAM may help with the diagnosis of TB. As a simple point-of-care test that does not depend upon sputum evaluation, there may also be situations where LF-LAM testing offers advantages compared to existing sputum-based tests.

Clinicians must consider the need for additional testing when interpreting negative LF-LAM results. The consequences of false negative results are increased risk of morbidity and mortality, delayed treatment initiation, and the continued risk of TB transmission. The consequences of false positive results are likelihood of anxiety and morbidity caused by additional testing, unnecessary treatment, and possible adverse events; possible stigma associated with a diagnosis of TB; and the chance that a false positive may halt further diagnostic evaluation. As LF-LAM does not offer information about drug-resistance, some patients with unidentified drug-resistant TB may be inappropriately treated with a regimen appropriate only for drug-sensitive disease.

For TB screening among HIV-positive people regardless of TB symptoms, LF-LAM yielded low sensitivity and relatively high specificity. These findings do not support the use of LF-LAM as screening test for TB.

A systematic review of economic evaluations of LF-LAM for TB, [Dowdy 2015 \[pers comm\]](#), identified two studies that evaluated the addition of LF-LAM to existing diagnostic strategies including smear, Xpert® MTB/RIF, and clinical diagnosis ([Shah 2013](#); [Sun 2013](#)). Both studies evaluated LF-LAM in the African setting; one study was restricted to inpatients with CD4 < 100 cells per µL ([Sun 2013](#)). The two studies, using similar methodologies, found that incorporation of LF-LAM into TB diagnostic algorithms for HIV-positive people is highly cost-effective. This result was robust to wide sensitivity analyses that explored uncertainties in input parameter estimates. Given the low cost of LF-LAM, the findings

suggest that the cost-effectiveness of LF-LAM ultimately converges to the cost-effectiveness of HIV and TB treatment because the incremental diagnostic costs for LF-LAM are dwarfed by treatment costs for those who test positive. Key parameters that influenced the degree of LF-LAM cost-effectiveness included the specificity of LF-LAM, prevalence of TB in the target population, the life expectancy of HIV-infected TB survivors, and costs of treating TB and HIV.

Implications for research

Future studies that evaluate the diagnostic accuracy of non-sputum-based tests for TB, such as LF-LAM, in HIV-positive people should use a reference standard that includes at least two specimen types or extrapulmonary specimens in addition to sputum. Moreover, future studies should include patients unable to expectorate sputum in the analysis. These features of study design may decrease the risk of bias in the accuracy estimates. Further research on effective implementation of LF-LAM within routine clinical practice is needed because the test can only influence clinical practice if the results are believed and acted upon.

ACKNOWLEDGEMENTS

MS is supported through an National Institute of Health (NIH) K23 grant (K23AI089259).

The editorial base for the Cochrane Infectious Diseases Group (CIDG) is funded by the UK Department for International Development (DFID) in a grant related to evidence synthesis for the benefit of developing countries (Grant: 5242). The views expressed in this review do not necessarily reflect UK government policy.

We thank all authors of the included studies for answering our questions and providing data. We are grateful to Vittoria Lutje, the Information Retrieval Specialist with the CIDG, and Hacs Horvath, from the University of California, San Francisco, for their help with the search strategy. In addition, we thank Samuel Schumacher for his helpful comments on the manuscript. We also thank Rachel O'Shea, Alere, Inc, and Amy Yorston, formerly with Alere, Inc, for their comments on the index test and help in identifying studies. We are grateful to the peer reviewers for their helpful comments and suggestions.

REFERENCES

References to studies included in this review

Andrews 2014 *{unpublished data only}*

Andrews B, Muchemwa L, Lakhi S, Bwalya M, Mabula C, Chipii G, Heimburger DC, Bernard GR. Validation of a clinical prediction score and performance of urine lipoarabinomannan test for detecting tuberculosis bacteremia in HIV-positive patients with severe sepsis. *American Thoracic Society*. 2014:A2559.

Balcha 2014 *{published and unpublished data}*

Balcha TT, Winqvist N, Sturegård E, Skogmar S, Reepalu A, Jemal ZH, et al. Detection of lipoarabinomannan in urine for identification of active tuberculosis among HIV-positive adults in Ethiopian health centres. *Tropical Medicine & International Health* 2014;**19**(6):734–42.

Bjerrum 2015 *{published and unpublished data}*

Bjerrum S, Kenu E, Lartey M, Newman MJ, Addo KK, Andersen AB, et al. Diagnostic accuracy of the rapid urine lipoarabinomannan test for pulmonary tuberculosis among HIV-infected adults in Ghana- findings from the DETECT HIV-TB study. *BMC Infectious Diseases* 2015;**15**:407.

Drain 2014a *{published and unpublished data}*

Drain PK, Losina E, Coleman SM, Giddy J, Ross D, Katz JN, et al. Diagnostic accuracy of a point-of-care urine test for tuberculosis screening among newly-diagnosed HIV-infected adults: a prospective, clinic-based study. *BMC Infectious Diseases* 2014;**14**:110.

Drain 2014b *{published and unpublished data}*

Drain PK, Losina E, Coleman SM, Giddy J, Ross D, Katz JN, et al. Value of urine lipoarabinomannan grade and second test for optimising clinic-based screening for HIV-associated pulmonary tuberculosis. *Journal of Acquired Immune Deficiency Syndromes* 2015;**68**(3):274–80.

Drain 2014c *{unpublished data only}*

Drain P, Grobler A, Gounder L, Sahid F, Wilson D, Bassett I, et al. Rapid urine lipoarabinomannan testing after two months of tuberculosis treatment independently predicts mortality in a resource-limited setting. 44th World Conference on Lung Health of the International Union Against Tuberculosis and Lung Disease. Paris, 2013; Vol. 17 (Suppl 2):S264.

LaCourse 2015 *{published and unpublished data}*

LaCourse SM, Cranmer LM, Matemo D, Richardson B, Kinuthia J, John-Stewart G, et al. Active tuberculosis case finding in HIV-infected pregnant women in Kenya reveals poor performance of symptom screening and rapid diagnostic tests. *Journal of Acquired Immune Deficiency Syndromes* 2016;**71**(2):219–27.

Lawn 2012a *{published and unpublished data}*

Lawn SD, Kerkhoff AD, Vogt M, Wood R. Diagnostic accuracy of a low-cost, urine antigen, point-of-care screening assay for HIV-associated pulmonary tuberculosis before antiretroviral therapy: a descriptive study. *Lancet Infectious Diseases* 2012;**12**(3):201–9.

Lawn 2014a *{unpublished data only}*

Lawn SD, Kerkhoff A, Burton R, Schutz C, van Wyk G, Vogt M, et al. Massive diagnostic yield of HIV-associated tuberculosis using rapid urine assays in South Africa. Conference on Retroviruses and Opportunistic Infections (CROI). Boston (MA), 2014.

Nakiyingi 2014 *{published and unpublished data}*

Nakiyingi L, Moodley VM, Manabe YC, Nicol MP, Holshouser M, Armstrong DT, et al. Diagnostic accuracy of a rapid urine lipoarabinomannan test for tuberculosis in HIV-infected adults. *Journal of Acquired Immune Deficiency Syndromes* 2014;**66**(3):270–9.

Peter 2012a *{published and unpublished data}*

Peter JG, Theron G, van Zyl-Smit R, Haripersad A, Mottay L, Kraus S, et al. Diagnostic accuracy of a urine lipoarabinomannan strip-test for TB detection in HIV-infected hospitalised patients. *European Respiratory Journal* 2012;**40**(5):1211–20.

Peter 2015 *{published and unpublished data}*

Peter J, Theron G, Chanda D, Clowes P, Rachow A, Lesosky M, et al. Test characteristics and potential impact of the urine LAM lateral flow assay in HIV-infected outpatients under investigation for TB and able to self-expectorate sputum for diagnostic testing. *BMC Infectious Diseases* 2015;**15**:262.

References to studies excluded from this review

Achkar 2011 *{published data only}*

Achkar JM, Lawn SD, Moosa MY, Wright CA, Kasprovicz VO. Adjunctive tests for diagnosis of tuberculosis: serology, ELISPOT for site-specific lymphocytes, urinary lipoarabinomannan, string test, and fine needle aspiration. *Journal of Infectious Diseases* 2011;**204**(Suppl 4):S1130–41.

Agha 2013 *{published data only}*

Agha MA, El-Helbawy RH, El-Helbawy NG, El-Sheak NM. Utility of quantitative analysis of urine lipoarabinomannan in the diagnosis of tuberculosis. *Egyptian Journal of Chest Diseases and Tuberculosis* 2013;**62**(3):401–7.

Amoudy 1997 *{published data only}*

Amoudy HA, Al-Asmer ABH, Abul AT, Mustafa AS. Evaluation of complex and defined antigens of *Mycobacterium tuberculosis* in an IgG-specific ELISA for the diagnosis of tuberculosis. *Medical Principles and Practice* 1997;**6**(2):103–9.

Boehme 2005 *{published data only}*

Boehme C, Molokova E, Minja F, Geis S, Loscher T, Maboko L, et al. Detection of mycobacterial lipoarabinomannan with an antigen-capture ELISA in unprocessed urine of Tanzanian patients with suspected tuberculosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2005;**99**(12):893–900.

Boyles 2012 *{published data only}*

Boyles TH. The pros and cons of urinary lipoarabinomannan testing. *AIDS* 2012;**26**(17):2263–4; author reply 2264–5.

- Chan 2000** *{published data only}*
Chan ED, Reves R, Belisle JT, Brennan PJ, Hahn WE. Diagnosis of tuberculosis by a visually detectable immunoassay for lipoarabinomannan. *American Journal of Respiratory and Critical Care Medicine* 2000;**161**(5):1713–9.
- Cho 1997** *{published data only}*
Cho S-N, Lee J-H, Lee H-Y, Won H-J, Chong Y, Chang J, et al. Detection of Mycobacterium tuberculosis antigens in sputum for the diagnosis of pulmonary tuberculosis. *Journal of the Korean Society for Microbiology* 1997;**32**(3):285–91.
- Conesa-Botella 2007** *{published data only}*
Conesa-Botella A, Loembé MM, Manabe YC, Worodria W, Mazakpwe D, Luzinda K, et al. Urinary lipoarabinomannan as predictor for the tuberculosis immune reconstitution inflammatory syndrome. *Journal of Acquired Immune Deficiency Syndromes* 2011;**58**(5):463–8.
- Daley 2009** *{published data only}*
Daley P, Michael JS, Hmar P, Latha A, Chordia P, Mathai D, et al. Blinded evaluation of commercial urinary lipoarabinomannan for active tuberculosis: a pilot study. *International Journal of Tuberculosis and Lung Disease* 2009;**13**(8):989–95.
- Deng 2011** *{published data only}*
Deng S, Yuan T, Xia J, Huang H, Cheng X, Chen M. Clinical utility of a combination of lipoarabinomannan, 38-kDa, and 16-kDa antigens as a diagnosis tool for tuberculosis. *Diagnostic Microbiology and Infectious Disease* 2011;**71**(1):46–50.
- Dheda 2009** *{published data only}*
Dheda K, Van-Zyl Smit RN, Sechi LA, Badri M, Meldau R, Symons G, et al. Clinical diagnostic utility of IP-10 and LAM antigen levels for the diagnosis of tuberculous pleural effusions in a high burden setting. *PLoS One* 2009;**4**(3): e4689.
- Dheda 2010** *{published data only}*
Dheda K, Davids V, Lenders L, Roberts T, Meldau R, Ling D, et al. Clinical utility of a commercial LAM-ELISA assay for TB diagnosis in HIV-infected patients using urine and sputum samples. *PLoS One* 2010;**5**(3):e9848.
- Elsawy 2012** *{published data only}*
Elsawy A, Redwan EM. Urine lipoarabinomannan as initial markers for active pulmonary tuberculosis. *Australian Journal of Basic and Applied Sciences* 2012;**6**(3):751–5.
- Gounder 2011** *{published data only}*
Gounder CR, Kufa T, Wada NI, Mngomezulu V, Charalambous S, Hanifa Y, et al. Diagnostic accuracy of a urine lipoarabinomannan enzyme-linked immunosorbent assay for screening ambulatory HIV-infected persons for tuberculosis. *Journal of Acquired Immune Deficiency Syndromes* 2011;**58**(2):219–23.
- Hamasur 2001** *{published data only}*
Hamasur B, Bruchfeld J, Haile M, Pawlowski A, Bjorvatn B, Källenius G, et al. Rapid diagnosis of tuberculosis by detection of mycobacterial lipoarabinomannan in urine. *Journal of Microbiological Methods* 2001;**45**(1):41–52.
- Hanrahan 2012** *{published data only}*
Hanrahan CF, Van Rie A. Urine antigen test for diagnosis of HIV-associated tuberculosis. *The Lancet Infectious Diseases* 2012;**12**(11):826.
- Kashino 2008** *{published data only}*
Kashino SS, Pollock N, Napolitano DR, Rodrigues V Jr, Campos-Neto A. Identification and characterization of Mycobacterium tuberculosis antigens in urine of patients with active pulmonary tuberculosis: an innovative and alternative approach of antigen discovery of useful microbial molecules. *Clinical and Experimental Immunology* 2008;**153**(1):56–62.
- Kerkhoff 2014a** *{published data only}*
Kerkhoff AD, Wood R, Vogt M, Lawn SD. Predictive value of anaemia for tuberculosis in HIV-infected patients in sub-Saharan Africa: an indication for routine microbiological investigation using new rapid assays. *Journal of Acquired Immune Deficiency Syndromes* 2014;**66**(1):33–40.
- Kerkhoff 2014b** *{published data only}*
Kerkhoff AD, Wood R, Vogt M, Lawn SD. Prognostic value of a quantitative analysis of lipoarabinomannan in urine from patients with HIV-associated tuberculosis. *PLoS One* 2014;**9**(7):e103285.
- Koulchin 2007** *{published data only}*
Koulchin VA, Boehme C, Molokova E, Turecheck C, Hoelscher M, Perkins M, et al. Evaluation of the detection of urinary lipoarabinomannan as a tool for diagnosis of mycobacterial infections. *Abstracts of the General Meeting of the American Society for Microbiology* 2007;**107**:174–5.
- Kroidi 2014** *{published data only}*
Kroidl I, Clowes P, Mwakyelu J, Maboko L, Kiangi A, Rachow A, et al. Reasons for false-positive lipoarabinomannan ELISA results in a Tanzanian population. *Scandinavian Journal of Infectious Diseases* 2014;**46**(2):144–8.
- Lawn 2009** *{published data only}*
Lawn SD, Edwards DJ, Kranzer K, Vogt M, Bekker LG, Wood R. Urine lipoarabinomannan assay for tuberculosis screening before antiretroviral therapy diagnostic yield and association with immune reconstitution disease. *AIDS* 2009;**23**(14):1875–80.
- Lawn 2012b** *{published data only}*
Lawn SD. Point-of-care detection of lipoarabinomannan (LAM) in urine for diagnosis of HIV-associated tuberculosis: a state of the art review. *BMC Infectious Diseases* 2012;**12**: 103.
- Lawn 2012c** *{published data only}*
Lawn SD, Kerkhoff AD, Vogt M, Wood R. Clinical significance of lipoarabinomannan detection in urine using a low-cost point-of-care diagnostic assay for HIV-associated tuberculosis. *AIDS* 2012;**26**(13):1635–43.
- Lawn 2012d** *{published data only}*
Lawn SD, Wood R. Point-of-care urine antigen screening tests for tuberculosis and cryptococcosis: potential for mortality reduction in antiretroviral treatment programs in Africa. *Clinical Infectious Diseases* 2012;**54**(5):739–40.

- Lawn 2013b** *{published data only}*
Lawn SD, Kerkhoff AD, Vogt M, Wood R. HIV-associated tuberculosis: relationship between disease severity and the sensitivity of new sputum-based and urine-based diagnostic assays. *BMC Medicine* 2013;**11**:231.
- Lawn 2014b** *{published data only}*
Lawn SD, Kerkhoff AD, Burton R, Meintjes G. Underestimation of the incremental diagnostic yield of HIV-associated tuberculosis in studies of the Determine TB-LAM Ag urine assay. *AIDS* 2014;**28**(12):1846–8.
- Manabe 2014** *{published data only}*
Manabe YC, Nonyane BA, Nakiyingi L, Mbabazi O, Lubega G, Shah M, et al. Point-of-care lateral flow assays for tuberculosis and cryptococcal antigenuria predict death in HIV infected adults in Uganda. *PLoS One* 2014;**9**(7): e101459.
- Mukundan 2012** *{published data only}*
Mukundan H, Kumar S, Price DN, Ray SM, Lee YJ, Min S, et al. Rapid detection of Mycobacterium tuberculosis biomarkers in a sandwich immunoassay format using a waveguide-based optical biosensor. *Tuberculosis (Edinburgh, Scotland)* 2012;**92**(5):407–16.
- Mutetwa 2009** *{published data only}*
Mutetwa R, Boehme C, Dimairo M, Bandason T, Munyati SS, Mangwanya D, et al. Diagnostic accuracy of commercial urinary lipoarabinomannan detection in African tuberculosis suspects and patients. *International Journal of Tuberculosis and Lung Disease* 2009;**13**(10): 1253–9.
- Patel 2009** *{published data only}*
Patel VB, Bhigjee AI, Paruk HF, Singh R, Meldau R, Connolly C, et al. Utility of a novel lipoarabinomannan assay for the diagnosis of tuberculous meningitis in a resource-poor high-HIV prevalence setting. *Cerebrospinal Fluid Research* 2009;**6**:13.
- Patel 2010** *{published data only}*
Patel VB, Singh R, Connolly C, Kasprovicz V, Zumla A, Ndungu T, et al. Comparison of a clinical prediction rule and a LAM antigen-detection assay for the rapid diagnosis of TBM in a high HIV prevalence setting. *PLoS One* 2010;**5**(12):e15664.
- Peter 2011** *{published data only}*
Peter JG, Haripesad A, Mottay L, Kraus S, Meldau R, Dheda K. The clinical utility of urine lipoarabinomannan and the novel point-of-care lateral flow strip test (Determine TB) for the diagnosis of tuberculosis in hospitalised patients with HIV-related advanced immunosuppression. *American Journal of Respiratory and Critical Care Medicine Conference: American Thoracic Society International Conference, ATS* 2011;**183**:A5313.
- Peter 2012b** *{published data only}*
Peter JG, Cashmore TJ, Meldau R, Theron G, van Zyl-Smit R, Dheda K. Diagnostic accuracy of induced sputum LAM ELISA for tuberculosis diagnosis in sputum-scarce patients. *International Journal of Tuberculosis and Lung Disease* 2012;**16**(8):1108–12.
- Peter 2012c** *{published data only}*
Peter JG, Theron G, Muchinga TE, Govender U, Dheda K. The diagnostic accuracy of urine-based Xpert MTB/RIF in HIV-infected hospitalised patients who are smear-negative or sputum scarce. *PLoS One* 2012;**7**(7):e39966.
- Peter 2013** *{published data only}*
Peter JG, Theron G, Dheda K. Can point-of-care urine LAM strip testing for tuberculosis add value to clinical decision making in hospitalised HIV-infected persons?. *PLoS One* 2013;**8**(2):e54875.
- Reid 2015** *{unpublished data only}*
Reid S, Kancheya N, Kapata N, Kruuner A, Henostroza G, Harris J. Detection of lipoarabinomannan in urine for the diagnosis of tuberculosis in ART clinic enrollees in Lusaka, Zambia. <http://www.cidrz.org/past-studies/> (accessed 4 September 2015).
- Reither 2009** *{published data only}*
Reither K, Saathoff E, Jung J, Minja LT, Kroidl I, Saad E, et al. Low sensitivity of a urine LAM-ELISA in the diagnosis of pulmonary tuberculosis. *BMC Infect Diseases* 2009;**9**: 141.
- Sada 1992** *{published data only}*
Sada E, Aguilar D, Torres M, Herrera T. Detection of lipoarabinomannan as a diagnostic test for tuberculosis. *Journal of Clinical Microbiology* 1992;**30**(9):2415–8.
- Savolainen 2013** *{published data only}*
Savolainen L, Kantele A, Sandboge B, Sirén M, Valleala H, Tuompo R, et al. Modification of Clearview tuberculosis (TB) enzyme-linked immunosorbent assay for TB patients not infected with HIV. *Clinical and Vaccine Immunology* 2013;**20**(9):1479–82.
- Schmidt 2011** *{published data only}*
Schmidt R, Jacak J, Schirwitz C, Stadler V, Michel G, Marmé N, et al. Single-molecule detection on a protein-array assay platform for the exposure of a tuberculosis antigen. *Journal of Proteome Research* 2011;**10**(3):1316–22.
- Shah 2009** *{published data only}*
Shah M, Variava E, Holmes CB, Coppin A, Golub JE, McCallum J, et al. Diagnostic accuracy of a urine lipoarabinomannan test for tuberculosis in hospitalised patients in a high HIV prevalence setting. *Journal of Acquired Immune Deficiency Syndromes* 2009;**52**(2):145–51.
- Shah 2010** *{published data only}*
Shah M, Martinson NA, Chaisson RE, Martin DJ, Variava E, Dorman SE. Quantitative analysis of a urine-based assay for detection of lipoarabinomannan in patients with tuberculosis. *Journal of Clinical Microbiology* 2010;**48**(8): 2972–4.
- Shah 2013** *{published data only}*
Shah M, Dowdy D, Joloba M, Ssengooba W, Manabe YC, Ellner J, et al. Cost-effectiveness of novel algorithms for rapid diagnosis of tuberculosis in HIV-infected individuals in Uganda. *AIDS* 2013;**27**(18):2883–92.
- Shah 2014** *{published data only}*
Shah M, Ssengooba W, Armstrong D, Nakiyingi L, Holshouser M, Ellner JJ, et al. Comparative performance of

urinary lipoarabinomannan assays and Xpert MTB/RIF in HIV-infected individuals. *AIDS* 2014;**28**(9):1307–14.

Singh 2011 {published data only}

Singh L, Grover N. Detection of TB antigen by rapid test kit. *Medical Journal Armed Forces India* 2011;**67**(2):196–7.

Sun 2013 {published data only}

Sun D, Dorman S, Shah M, Manabe YC, Moodley VM, Nicol MP, et al. Cost utility of lateral-flow urine lipoarabinomannan for tuberculosis diagnosis in HIV-infected African adults. *International Journal of Tuberculosis and Lung Disease* 2013;**17**(4):552–8.

Swaminathan 2012 {published data only}

Swaminathan S, Rekha VV. Antigen detection as a point-of-care test for TB: the case of lipoarabinomannan. *Future Microbiology* 2012;**7**(5):559–64.

Talbot 2012 {published data only}

Talbot E, Munseri P, Teixeira P, Matee M, Bakari M, Lahey T, et al. Test characteristics of urinary lipoarabinomannan and predictors of mortality among hospitalised HIV-infected tuberculosis suspects in Tanzania. *PLoS One* 2012;**7**(3):e32876.

Tessema 2001 {published data only}

Tessema TA, Hamasur B, Bjun G, Svenson S, Bjorvatn B. Diagnostic evaluation of urinary lipoarabinomannan at an Ethiopian tuberculosis centre. *Scandinavian Journal of Infectious Diseases* 2001;**33**(4):279–84.

Tessema 2002a {published data only}

Tessema TA, Bjune G, Assefa G, Svenson S, Hamasur B, Bjorvatn B. Clinical and radiological features in relation to urinary excretion of lipoarabinomannan in Ethiopian tuberculosis patients. *Scandinavian Journal of Infectious Diseases* 2002;**34**(3):167–71.

Tessema 2002b {published data only}

Tessema TA, Bjune G, Hamasur B, Svenson S, Syre H, Bjorvatn B. Circulating antibodies to lipoarabinomannan in relation to sputum microscopy, clinical features and urinary anti-lipoarabinomannan detection in pulmonary tuberculosis. *Scandinavian Journal of Infectious Diseases* 2002;**34**(2):97–103.

Tlali 2014 {unpublished data only}

Tlali M, Fielding K, Charalambous S, Karat A, Hoffmann C, Johnson S, et al. Sensitivity of the TB Determine LAM test compared to sputum culture gold standard TB in ambulant HIV positive participants enrolled in the TB fast track study in South Africa. 4th South African TB Conference. Durban, 2014.

Van Rie 2013 {unpublished data only}

Van Rie A, Jong E, Mkhwanazi M, Sanne I. Diagnosing TB in those hardest to diagnose: urine lipoarabinomannan for suspects of disseminated and extrapulmonary TB. Conference on Retroviruses and Opportunistic Infections (CROI). Atlanta (GA), 2013:443.

Wood 2012 {published data only}

Wood R, Racow K, Bekker LG, Middelkoop K, Vogt M, Kreiswirth BN, et al. Lipoarabinomannan in urine during

tuberculosis treatment: association with host and pathogen factors and mycobacteriuria. *BMC Infect Diseases* 2012;**12**:47.

References to studies awaiting assessment

NCT01990274 {published data only}

NCT01990274. The utility of intensified case finding combined with a package of novel TB diagnostics performed at community-based-clinics in South Africa - a multi-centric prospective cohort study (XACT study). <https://clinicaltrials.gov/ct2/show/NCT01990274?term=xact&rank=2> (accessed 04 September 2015).

References to ongoing studies

Grant ongoing {unpublished data only}

Grant A. Xpert MTB/Rif for people attending HIV care: an interventional cohort study to guide rational implementation (XPHACTOR). http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=10&ved=0ahUKewj5tLmS-JvMAhUKRCYKHRClA2sQFghBMak&url=http%3A%2F%2Fwww.auruminstitute.org%2Findex.php%3Foption%3Dcom_k2%2Fusg=AFQjCNF7uy6NidwmpPa-FbGGDfjEmMpvQ (accessed 4 April 2016).

Additional references

Alere 2014

Alere. Alere DetermineTM TB LAM Ag Product Information. <http://www.alere.com/ww/en/product-details/determine-tb-lam.html> (accessed 16 November 2014).

Bartlett 2007

Bartlett JG. Tuberculosis and HIV infection: partners in human tragedy. *Journal of Infectious Diseases* 2007;**196** (Suppl 1):S124–5.

Batz 2011

Batz HG, Cooke GS, Reid SD. Towards lab-free tuberculosis diagnosis. August 2011. http://www.msface.org/sites/default/files/MSF_assets/TB/Docs/TB_Report_TowardsLabFreeTBDX_2011_ENG.pdf. London: Imperial College, (accessed 16 November 2014).

Cain 2010

Cain KP, McCarthy KD, Heilig CM, Monkongdee P, Tasaneeyapan T, Kanara N, et al. An algorithm for tuberculosis screening and diagnosis in people with HIV. *New England Journal of Medicine* 2010;**362**(8):707–16.

Chu 2009

Chu H, Chen S, Louis TA. Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. *Journal of the American Statistical Association* 2009;**104**(486):512–23.

Dowdy 2015 [pers comm]

Dowdy D. A review of cost-effectiveness analyses of LF-LAM for TB diagnosis among people living with HIV [personal communication]. Email to: M Shah. 4 September 2015.

Drain 2015

Drain PK, Gounder L, Grobler A, Sahid F, Bassett IV, Moosa MY. Urine lipoarabinomannan to monitor antituberculosis therapy response and predict mortality in an HIV-endemic region: a prospective cohort study. *BMJ Open* 2015;**5**(4): e006833.

Getahun 2007

Getahun H, Harrington M, O'Brien R, Nunn P. Diagnosis of smear-negative pulmonary tuberculosis in people with HIV infection or AIDS in resource-constrained settings: informing urgent policy changes. *Lancet* 2007;**369**(9578): 2042–9.

Getahun 2010

Getahun H, Gunneberg C, Granich R, Nunn P. HIV infection-associated tuberculosis: the epidemiology and the response. *Clinical Infectious Diseases* 2010;**50**(Suppl 3): S201–7.

Getahun 2011

Getahun H, Kittikraisak W, Heilig CM, Corbett EL, Ayles H, Cain KP, et al. Development of a standardized screening rule for tuberculosis in people living with HIV in resource-constrained settings: individual participant data meta-analysis of observational studies. *PLoS Medicine* 2011;**8**(1): e1000391.

GRADEpro 2015 [Computer program]

McMaster University (developed by Evidence Prime, Inc.). GRADEpro GDT: GRADEpro Guideline Development Tool. McMaster University (developed by Evidence Prime, Inc.), 2015.

Gupta 2015

Gupta RK, Lucas SB, Fielding KL, Lawn SD. Prevalence of tuberculosis in post-mortem studies of HIV-infected adults and children in resource-limited settings: a systematic review and meta-analysis. *AIDS* 2015;**29**(15):1987–2002.

Havlir 2008

Havlir DV, Getahun H, Sanne I, Nunn P. Opportunities and challenges for HIV care in overlapping HIV and TB epidemics. *Journal of the American Medical Association* 2008;**300**(4):423–30.

Kwan 2011

Kwan CK, Ernst JD. HIV and tuberculosis: a deadly human syndemic. *Clinical Microbiology Reviews* 2011;**24**(2):351–76.

Lawn 2013a

Lawn SD, Dheda K, Kerkhoff AD, Peter JG, Dorman S, Boehme CC, et al. Determine TB-LAM lateral flow urine antigen assay for HIV-associated tuberculosis: recommendations on the design and reporting of clinical studies. *BMC Infectious Diseases* 2013;**13**:407.

Lawn 2015

Lawn SD, Kerkhoff AD, Nicol PM, Meintjes G. Underestimation of the true specificity of the urine lipoarabinomannan (LAM) point-of-care diagnostic assay for HIV-associated tuberculosis. *Journal of Acquired Immune Deficiency Syndromes* 2015;**69**(4):e144–6.

Lunn 2000

Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000;**10**:325–37.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0. The Cochrane Collaboration, 2010. Available from <http://srda.cochrane.org/>.

Minion 2011

Minion J, Leung E, Talbot E, Dheda K, Pai M, Menzies D. Diagnosing tuberculosis with urine lipoarabinomannan: systematic review and meta-analysis. *European Respiratory Journal* 2011;**38**(6):1398–405.

Novielli 2013

Novielli N, Sutton AJ, Cooper NJ. Meta-analysis of the accuracy of two diagnostic tests used in combination: application to the d-dimer test and the wells score for the diagnosis of deep vein thrombosis. *Value Health* 2013;**16**(4):619–28.

Pai 2012

Pai NP, Pai M. Point-of-care diagnostics for HIV and tuberculosis: landscape, pipeline, and unmet needs. *Discovery Medicine* 2012;**13**(68):35–45.

Perkins 2007

Perkins MD, Cunningham J. Facing the crisis: improving the diagnosis of tuberculosis in the HIV era. *Journal of Infectious Diseases* 2007;**196**(Suppl 1):S15–27.

Peter 2010

Peter J, Green C, Hoelscher M, Mwaba P, Zumla A, Dheda K. Urine for the diagnosis of tuberculosis: current approaches, clinical applicability, and new developments. *Current Opinion of Pulmonary Medicine* 2010;**16**(3): 262–70.

Peter 2012

Peter JG, Theron G, van Zyl-Smit R, Haripersad A, Mottay L, Kraus S, et al. Diagnostic accuracy of a urine lipoarabinomannan strip-test for TB detection in HIV-infected hospitalised patients. *European Respiratory Journal* 2012;**40**(5):1211–20.

Peter 2016

Peter JG, Zijenah LS, Chanda D, Clowes P, Lesosky M, Gina P, et al. Effect on mortality of point-of-care, urine-based lipoarabinomannan testing to guide tuberculosis treatment initiation in HIV-positive hospital inpatients: a pragmatic, parallel-group, multicountry, open-label, randomised controlled trial. *Lancet* 2016;**387**(10024): 1187–97.

R Statistical Computing 2015 [Computer program]

R Development Core Team. R: A language and environment for statistical computing. Version 3.2.1. Vienna: R Foundation for Statistical Computing, 2015.

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–90.

Review Manager [Computer program]

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Schünemann 2008

Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; **336**(7653):1106–11.

Schünemann 2013

Schünemann H, Brozek E, Guyatt G, Oxman A, editors. The GRADE Working Group. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. Available from http://gdt.guidelinedevelopment.org/central_prod/_design/client/handbook/handbook.html (accessed 12 April 2016).

Siddiqi 2003

Siddiqi K, Lambert ML, Walley J. Clinical diagnosis of smear-negative pulmonary tuberculosis in low-income countries: the current evidence. *Lancet Infectious Diseases* 2003; **3**(5):288–96.

StataCorp 2011 [Computer program]

StataCorp. Stata Statistical Software. Version 13. Texas: StataCorp LP, 2011.

Steingart 2006

Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infectious Diseases* 2006; **6**(9):570–81.

Steingart 2014

Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* 2014, Issue 1. [DOI: 10.1002/14651858.CD009593.pub3]

TB CARE I 2014

TB CARE I. International Standards for Tuberculosis Care, Edition 3. TB Care 1, The Hague. http://www.who.int/tb/publications/ISTC_3rdEd.pdf?ua=1 (accessed 16 Nov 2014):1–87.

Trikalinos 2014

Trikalinos TA, Hoaglin DC, Schmid CH. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. *Statistics in Medicine* 2014; **33**(9):1441–59.

Vacek 1985

Vacek PM. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* 1985; **41**(4):959–68.

Weyer 2011

Weyer K, Carai S, Nunn P. Viewpoint TB diagnostics: what does the world really need?. *Journal of Infectious Diseases* 2011; **204**(Suppl 4):S1196–202.

Whalen 1995

Whalen C, Horsburgh CR, Hom D, Lahart C, Simberkoff M, Ellner J. Accelerated course of human immunodeficiency virus infection after tuberculosis. *American Journal of Respiratory and Critical Care Medicine* 1995; **151**(1):129–35.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011; **155**(8):529–36.

WHO Global Tuberculosis Report 2015

World Health Organization. Global tuberculosis report 2015. WHO/HTM/TB/2015.22. *Global tuberculosis report 2015. WHO/HTM/TB/2015.22*. Geneva: World Health Organization, 2015.

WHO Lipoarabinomannan Policy Guidance 2015

World Health Organization. *The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV. Policy guidance. WHO/HTM/TB/2015.25*. Geneva: World Health Organization, 2015.

WHO Tuberculosis Screening 2013

World Health Organization. Systematic screening for active tuberculosis: principles and recommendations. WHO/HTM/TB/2013.04. *Systematic screening for active tuberculosis: principles and recommendations. WHO/HTM/TB/2013.04*. Geneva: World Health Organization, 2013.

WHO Xpert® Policy Update 2013

World Health Organization. Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF system for the diagnosis of pulmonary and extrapulmonary TB in adults and children: policy update. WHO/HTM/TB/2013.16. *Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF system for the diagnosis of pulmonary and extrapulmonary TB in adults and children: policy update. WHO/HTM/TB/2013.16*. Geneva: World Health Organization, 2013.

World Bank 2014

World Bank. World Bank List of Economies. <http://data.worldbank.org/country>. Washington (District of Columbia): World Bank, (accessed 10 February 2014).

References to other published versions of this review

Shah 2014

Shah M, Hanrahan C, Wang ZY, Steingart KR, Lawn SD, Denkinger C, et al. Urine lateral flow lipoarabinomannan assay for diagnosing active tuberculosis in adults living with HIV. *Cochrane Database of Systematic Reviews* 2014, Issue 12. [DOI: 10.1002/14651858.CD011420

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Andrews 2014

Study characteristics			
Patient sampling	Cross-sectional, consecutive enrolment		
Patient characteristics and setting	Presenting signs and symptoms: Signs and symptoms of sepsis Age: median 36 (interquartile range (IQR) 29 to 45) Sex, female: 39% HIV infection: 100% History of TB: not stated Sample size: 94 Clinical setting: inpatient Country: Zambia TB incidence rate: 410 per 100,000 Number (proportion) of TB cases in the study: 19 (20%)		
Index tests	LF-LAM		
Target condition and reference standard(s)	Target condition: mycobacteraemia; reference standard: mycobacterial blood culture (BacTec)		
Flow and timing	The study authors obtained all specimens at the same time during study enrolment		
Comparative			
Notes	The study authors performed mycobacterial blood cultures on all participants and did not exclude participants based on lack of sputum		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low

DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Balcha 2014

Study characteristics	
Patient sampling	Cross-sectional, consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: Irrespective of symptoms, patients with CD4 < 350 or WHO stage 4 who could produce sputum Age: median 33.5 (IQR 26.5 to 40)

	Sex, female: 44% HIV infection: 100% History of TB: not stated Sample size: 757 Clinical setting: outpatient Country: Ethiopia TB incidence rate: 224 per 100,000 Number (proportion) of TB cases in the study: 128 (17%)		
Index tests	LF-LAM		
Target condition and reference standard(s)	HIV-associated TB; expectorated sputum mycobacterial culture (liquid) and some lymph node mycobacterial culture (liquid)		
Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors excluded individuals who were unable to expectorate		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
		High	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
		High	

Bjerrum 2015

Study characteristics	
Patient sampling	Cross-sectional, consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: irrespective of symptoms, patients who could produce sputum Age: mean 39 (standard deviation (SD) 10) Sex, female: 64% HIV infection: 100% History of TB: not stated Sample size: 469 Clinical setting: outpatient Country: Ghana TB incidence rate: 66 per 100,000 Number (proportion) of TB cases in the study: 55 (12%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target: pulmonary TB; reference standard: sputum mycobacterial culture (solid and liquid); no extrapulmonary specimens tested

Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors excluded individuals without some respiratory specimen available and did not attempt to identify extrapulmonary TB. When using a composite reference standard, the study authors considered 5 LAM-positive and 11 LAM-negative (grade 2) patients 'unclassifiable' regarding their TB status		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		High	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Drain 2014a

Study characteristics	
Patient sampling	Cross-sectional, consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: not specified Age: mean 35.6 (SD 9.8) Sex, female: 45% HIV infection: 100% History of TB: not stated Sample size: 342 Clinical setting: outpatient Country: South Africa TB incidence rate: 860 per 100,000 Number (proportion) of TB cases in the study: 60 (18%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary TB; reference standard: mycobacterial (solid and liquid) culture of sputum without evaluation of extrapulmonary TB
Flow and timing	All patients were included in the analysis
Comparative	
Notes	The study authors performed sputum induction for individuals who could not expectorate. They excluded individuals without some respiratory specimen available and did not attempt to identify extrapulmonary TB. When using a composite reference standard, the study authors did not consider any patients 'unclassifiable' regarding their TB status

Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		

Drain 2014a (Continued)

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Drain 2014b

Study characteristics			
Patient sampling	Cross-sectional, consecutive enrolment		
Patient characteristics and setting	Presenting signs and symptoms: not specified Age: not specified Sex, female: not specified HIV infection: 100% History of TB: not stated Sample size: 320 Clinical setting: outpatient Country: South Africa TB incidence rate: 860 per 100,000 Number (proportion) of TB cases in the study: 54 (17%)		
Index tests	LF-LAM		
Target condition and reference standard(s)	Target condition: pulmonary TB; reference standard: mycobacterial (solid and liquid) culture of sputum without testing of extrapulmonary TB		
Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors performed sputum induction for individuals who could not expectorate. They excluded individuals without some respiratory specimen available and did not attempt to identify extrapulmonary TB. When using a composite reference standard, The study authors did not consider any patients 'unclassifiable' regarding their TB status		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Drain 2014b (Continued)

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Drain 2014c

Study characteristics			
Patient sampling	Cross-sectional, patient selection by convenience		
Patient characteristics and setting	Presenting signs and symptoms: signs and symptoms suggestive of TB, but not specified Age: not specified Sex, female: not specified HIV infection: 100% History of TB: not stated Sample size: 90 Clinical setting: primary care clinic and tertiary hospital Country: South Africa TB incidence rate: 860 per 100,000 Number (proportion) of TB cases in the study: 57 (63%)		
Index tests	LF-LAM		
Target condition and reference standard(s)	Target condition: pulmonary TB; reference standard: mycobacterial (solid and liquid) culture of sputum without testing of extrapulmonary samples		
Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors performed sputum induction for individuals who could not expectorate. They excluded individuals without some respiratory specimen available and did not attempt to identify extrapulmonary TB. All included individuals were smear-negative. When using a composite reference standard, the study authors did not consider any patients 'unclassifiable' regarding their TB status. Using a composite reference standard, the study authors categorized all participants as having TB		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
		High	Low
DOMAIN 2: Index Test All tests			

Drain 2014c (Continued)

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

LaCourse 2015

Study characteristics	
Patient sampling	Cross-sectional, consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: irrespective of symptoms Other characteristics: pregnant women Age: median 25 (IQR 22 to 30) Sex, female: 100% HIV infection: 100%

	History of TB: not stated Sample size: 283 Clinical setting: outpatient Country: Kenya TB incidence rate: 268 per 100,000 Number (proportion) of TB cases in the study: 3 (1%)		
Index tests	LF-LAM		
Target condition and reference standard(s)	Target condition: pulmonary TB; reference standard: expectorated sputum mycobacterial culture (liquid)		
Flow and timing	All patients were included in the analysis		
Comparative			
Notes	This study included 14 patients without sputum. The study authors did not perform sputum induction or test extrapulmonary specimens. When using a composite reference standard, the study authors considered 0 LAM-positive and 17 LAM-negative (grade 2) patients 'unclassifiable' regarding their TB status		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Lawn 2012a

Study characteristics

Patient sampling	Cross-sectional, consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: irrespective of symptoms, patients referred for antiretroviral therapy Age: median 34.1 (IQR 28.6 to 41.3) Sex, female: 64% HIV infection: 100% History of TB: not stated Sample size: 516 Clinical setting: outpatient Country: South Africa TB incidence rate: 860 per 100,000 Number (proportion) of TB cases in the study: 91 (18%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: HIV-associated TB; expectorated sputum mycobacterial (liquid) culture

Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors performed sputum induction for all individuals. They excluded individuals without some respiratory specimen available and did not attempt to identify extrapulmonary TB. When using a composite reference standard, the study authors considered 6 LAM-positive and 418 LAM-negative (grade 1) patients 'unclassifiable' regarding their TB status		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Lawn 2012a (Continued)

		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Lawn 2014a

Study characteristics	
Patient sampling	Cross-sectional, consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: not specified Age: Median 36.4 (IQR 29.0 to 42.4) Sex, female: 61% HIV infection: 100% History of TB: not stated Sample size: 413 Clinical setting: inpatient Country: South Africa TB incidence rate: 860 per 100,000 Number (proportion) of TB cases in the study: 136 (33%)
Index tests	LF-LAM
Target condition and reference standard(s)	HIV-associated TB; sputum mycobacterial culture (liquid), sputum Xpert® MTB/RIF, mycobacterial blood culture, 2 urine Xpert® MTB/RIF, and additional extrapulmonary culture and nucleic acid amplification test (NAAT)
Flow and timing	All patients were included in the analysis
Comparative	
Notes	The study authors performed sputum induction for all individuals and included those without sputum. They also performed multiple tests on extrapulmonary specimens. When using a composite reference standard, the study authors considered 3 LAM-positive and 270 LAM-negative (grade 2) patients 'unclassifiable' regarding their TB status

Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Nakiyingi 2014

Study characteristics			
Patient sampling	Cross-sectional, consecutive enrolment		
Patient characteristics and setting	Presenting signs and symptoms: signs and symptoms suggestive of TB, but not specified Age: median 33 (IQR 29 to 37) Sex, female: 63% HIV infection: 100% History of TB: not stated Sample size: 997 Clinical setting: inpatient and outpatient Country: South Africa and Uganda TB incidence rate: 860 per 100,000 (South Africa); 166 per 100,000 (Uganda) Number (proportion) of TB cases in the study: 367 (37%)		
Index tests	LF-LAM		
Target condition and reference standard(s)	HIV-associated TB; sputum mycobacterial culture (liquid and solid), mycobacterial blood cultures; sputum induction for those unable to expectorate		
Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors excluded patients who were unable to expectorate and those without a specimen following induction. When using a composite reference standard, the study authors did not consider any patients 'unclassifiable' regarding their TB status		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Nakiyingi 2014 (Continued)

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Study characteristics			
Patient sampling	Cross-sectional, consecutive enrolment		
Patient characteristics and setting	Presenting signs and symptoms: signs and symptoms suggestive of TB, but not specified Age: median 35 (IQR 29 to 40) Sex, female: 63% HIV infection: 100% History of TB: not stated Sample size: 241 Clinical setting: inpatient Country: South Africa TB incidence rate: 860 per 100,000 Number (proportion) of TB cases in the study: 116 (48%)		
Index tests	LF-LAM		
Target condition and reference standard(s)	There was no protocol to ensure a minimum standard of testing		
Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors included all patients with symptoms of TB irrespective of their ability to produce sputum; clinicians chose the reference standard rather than it being directed by the study protocol. The reference standard included testing of pulmonary and extrapulmonary sites by mycobacterial culture. When using a composite reference standard, the study authors considered 12 LAM positive and 54 LAM-negative (grade 2) patients 'unclassifiable' regarding their TB status		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			

Peter 2012a (Continued)

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		High	High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
		High	

Peter 2015

Study characteristics	
Patient sampling	Cross-sectional, consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: signs and symptoms suggestive of TB, but not specified Age: median 36 (IQR 30 to 41) Sex, female: 46% HIV infection: 100% History of TB: not stated

	<p>Sample size: 569 Clinical setting: outpatient Country: South Africa, Tanzania, Zambia TB incidence rate: 860 per 100,000 (South Africa); 164 per 100,000 (Tanzania); 410 per 100,000 (Zambia) Number (proportion) of TB cases in the study: 181 (32%)</p>		
Index tests	LF-LAM		
Target condition and reference standard(s)	HIV-associated TB; sputum mycobacterial (liquid) culture		
Flow and timing	All patients were included in the analysis		
Comparative			
Notes	The study authors excluded individuals unable to expectorate and did not test extrapulmonary specimens. When using a composite reference standard, the study authors considered 42 LAM-positive and 162 LAM-negative patients as 'unclassifiable' regarding their TB status		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
		High	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Unclear		
		Low	Low

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Achkar 2011	Review/editorial/commentary
Agha 2013	Index test not studied
Amoudy 1997	Index test not studied
Boehme 2005	Index test not studied
Boyles 2012	Index test not studied
Chan 2000	Index test not studied
Cho 1997	Index test not studied
Conesa-Botella 2007	Index test not studied

(Continued)

Daley 2009	Index test not studied
Deng 2011	Index test not studied
Dheda 2009	Index test not studied
Dheda 2010	Index test not studied
Elsawy 2012	Index test not studied
Gounder 2011	Index test not studied
Hamasur 2001	Index test not studied
Hanrahan 2012	Review/editorial/comment
Kashino 2008	Index test not studied
Kerkhoff 2014a	Duplicate data
Kerkhoff 2014b	Duplicate data
Koulchin 2007	Index test not studied
Kroidi 2014	Index test not studied
Lawn 2009	Index test not studied
Lawn 2012b	Review/editorial/comment
Lawn 2012c	Insufficient data
Lawn 2012d	Review/editorial/comment
Lawn 2013b	Insufficient data
Lawn 2014b	Review/editorial/comment
Manabe 2014	Duplicate data
Mukundan 2012	Index test not studied
Mutetwa 2009	Index test not studied
Patel 2009	Index test not studied
Patel 2010	Index test not studied
Peter 2011	Duplicate data

(Continued)

Peter 2012b	Duplicate data
Peter 2012c	Duplicate data
Peter 2013	Duplicate data
Reid 2015	Insufficient data
Reither 2009	Index test not studied
Sada 1992	Index test not studied
Savolainen 2013	Index test not studied
Schmidt 2011	Index test not studied
Shah 2009	Index test not studied
Shah 2010	Index test not studied
Shah 2013	Duplicate data
Shah 2014	Duplicate data
Singh 2011	Index test not studied
Sun 2013	Insufficient data
Swaminathan 2012	Review/editorial/comment
Talbot 2012	Review/editorial/comment
Tessema 2001	Index test not studied
Tessema 2002a	Index test not studied
Tessema 2002b	Index test not studied
Tlali 2014	Insufficient data
Van Rie 2013	Insufficient data
Wood 2012	Index test not studied

Characteristics of studies awaiting classification *[ordered by study ID]*

NCT01990274

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	Target condition: active TB; reference standard: MGIT liquid TB culture
Flow and timing	
Comparative	
Notes	A standard diagnostics package consisting of smear microscopy and culture will be compared with a novel diagnostics package involving point-of-care sputum GeneXpert MTB/RIF performed at a mobile or conventional clinic, sputum culture, and lateral flow urinary lipoarabinomannan (LAM) testing

Characteristics of ongoing studies *[ordered by study ID]*

Grant ongoing

Trial name or title	XPHACTOR
Target condition and reference standard(s)	Target condition: TB (both pulmonary and extrapulmonary); reference standard: liquid culture
Index and comparator tests	Index test: Determine TB-LAM; Comparator tests: Xpert® MTB/RIF
Starting date	Unknown
Contact information	Allison Grant, Aurum Institute
Notes	

DATA

Presented below are all the data for all of the tests entered into the review.

Tests. Data tables by test

Test	No. of studies	No. of participants
1 Diagnosis of TB against microbiological reference at Grade 1: all participants	6	2402
2 Diagnosis of TB against microbiological reference at Grade 2: all participants	5	2313
3 Diagnosis of TB against composite reference at Grade 1: all participants	3	1585
4 Diagnosis of TB against composite reference at Grade 2: all participants	3	1586
5 Diagnosis of TB against microbiological reference at Grade 1, CD4 > 200	4	870
6 Diagnosis of TB against microbiological reference at Grade 1, CD4 < 200	4	1131
7 Diagnosis of TB against microbiological reference at Grade 1, CD4 > 100	4	1289
8 Diagnosis of TB against microbiological reference at Grade 1, CD4 < 100	4	712
9 Diagnosis of TB against microbiological reference at Grade 1, CD4 > 50	4	1534
10 Diagnosis of TB against microbiological reference at Grade 1, CD4 < 50	4	467
11 Diagnosis of TB against microbiological reference at Grade 2, CD4 > 200	5	925
12 Diagnosis of TB against microbiological reference at Grade 2, CD4 < 200	5	1344
13 Diagnosis of TB against microbiological reference at Grade 1, CD4 > 100	5	1410

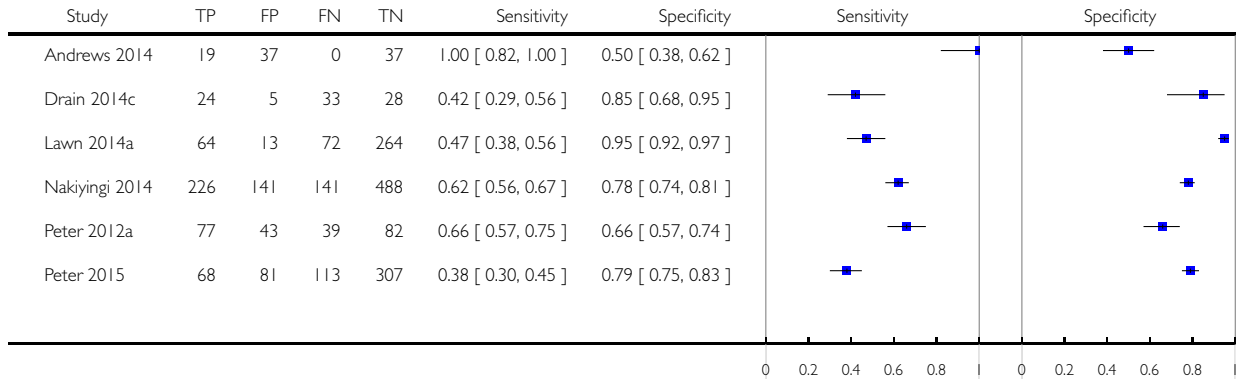
14	Diagnosis of TB against microbiological reference at Grade 1, CD4 < 100	5	859
15	Diagnosis of TB against microbiological reference at Grade 1, CD4 > 50	5	1726
16	Diagnosis of TB against microbiological reference at Grade 1, CD4 < 50	5	543
17	Diagnosis of TB against microbiological reference at Grade 1: inpatients	4	1298
18	Diagnosis of TB against microbiological reference at Grade 2: inpatients	4	1299
19	Diagnosis of TB against microbiological reference at Grade 1: outpatients	2	1014
20	Diagnosis of TB against microbiological reference at Grade 2: outpatients	2	1014
21	Diagnosis of TB with LAM at Grade 2 (among studies comparing with smear)	4	1875
22	Diagnosis of TB with smear (among studies comparing with LAM)	4	1875
23	Diagnosis of TB with combination of LAM at Grade 2 with smear	4	1876
24	Diagnosis of TB with LAM at Grade 2 (among studies comparing with Xpert®)	3	909
25	Diagnosis of TB with Xpert® (among studies comparing with LAM)	3	909
26	Diagnosis of TB with combination of LAM at Grade 2 with Xpert®	3	909
27	Screening for TB against microbiological reference at Grade 1: all participants	4	1935
28	Screening for TB against microbiological reference at Grade 2: all participants	3	1055
29	Screening for TB against composite reference at Grade 1: all participants	2	662
30	Screening for TB against composite reference at Grade 2: all participants	2	735

31 Screening for TB against microbiological reference at Grade 1, CD4 > 200	2	594
32 Screening for TB against microbiological reference at Grade 1, CD4 < 200	2	678
33 Screening for TB against microbiological reference at Grade 1, CD4 > 100	3	1221
34 Screening for TB against microbiological reference at Grade 1, CD4 < 100	3	579
35 Screening for TB against microbiological reference at Grade 1, CD4 > 50	2	1152
36 Screening for TB against microbiological reference at Grade 1, CD4 < 50	2	120
37 Screening for TB against microbiological reference at Grade 2, CD4 > 100	3	694
38 Screening for TB against microbiological reference at Grade 2, CD4 < 100	2	262
39 Screening for TB against microbiological reference at Grade 1: outpatients	4	1935
40 Screening for TB with LAM at Grade 2 (among studies comparing with smear microscopy)	3	1057
41 Screening for TB with smear microscopy (among studies comparing with LAM)	3	1059
42 Screening for TB with the combination of LAM (Grade 2) and smear microscopy	3	1075

Test 1. Diagnosis of TB against microbiological reference at Grade 1: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

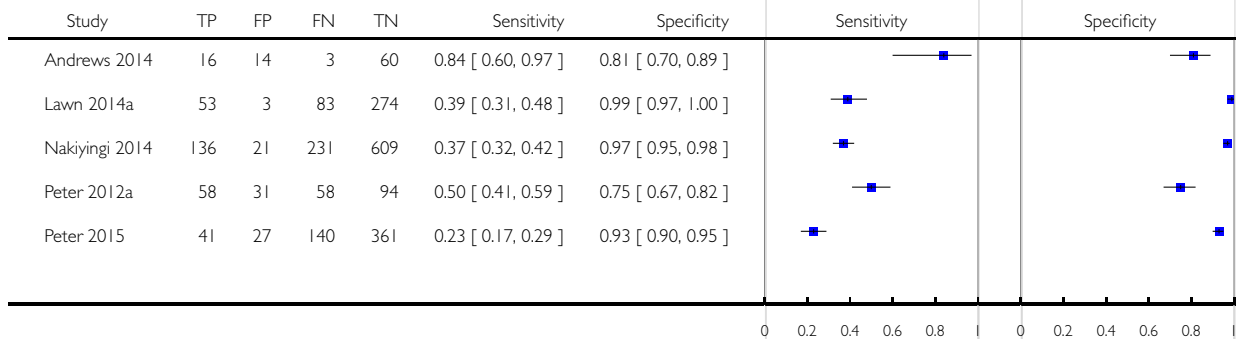
Test: 1 Diagnosis of TB against microbiological reference at Grade 1: all participants



Test 2. Diagnosis of TB against microbiological reference at Grade 2: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

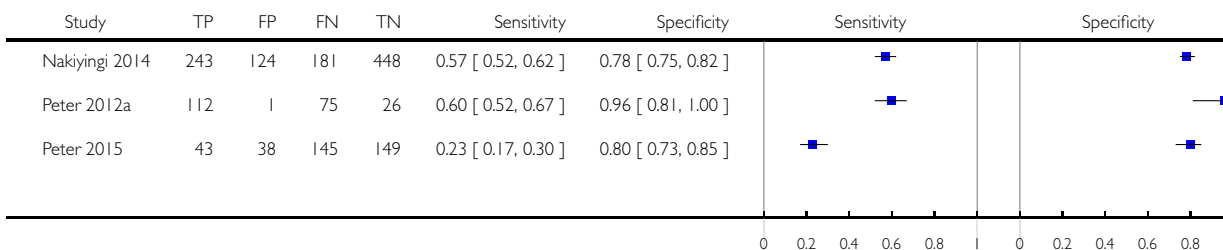
Test: 2 Diagnosis of TB against microbiological reference at Grade 2: all participants



Test 3. Diagnosis of TB against composite reference at Grade 1: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

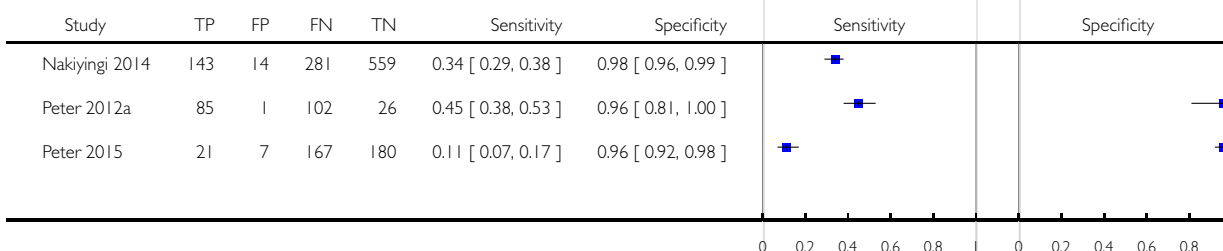
Test: 3 Diagnosis of TB against composite reference at Grade 1: all participants



Test 4. Diagnosis of TB against composite reference at Grade 2: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

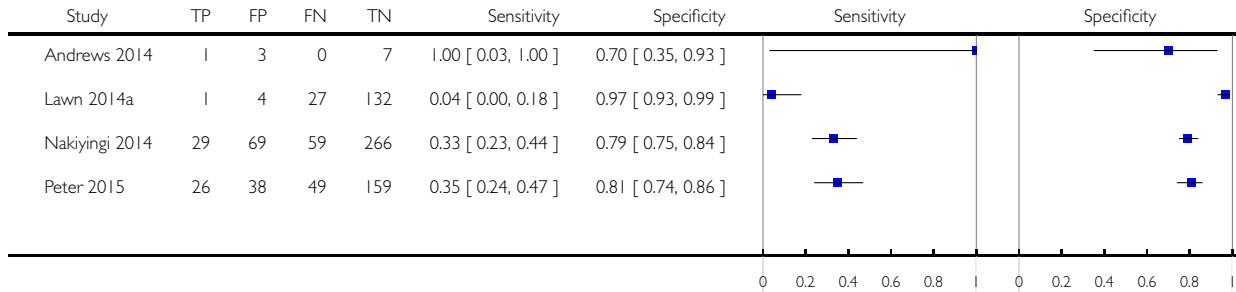
Test: 4 Diagnosis of TB against composite reference at Grade 2: all participants



Test 5. Diagnosis of TB against microbiological reference at Grade I, CD4 > 200.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

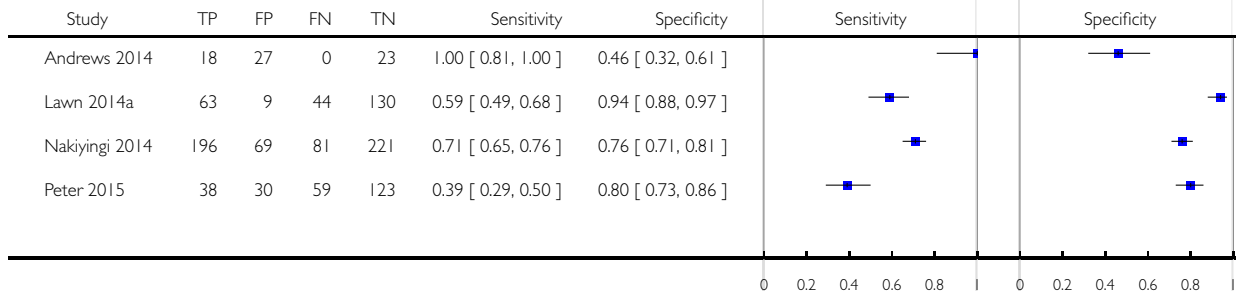
Test: 5 Diagnosis of TB against microbiological reference at Grade I, CD4 > 200



Test 6. Diagnosis of TB against microbiological reference at Grade I, CD4 < 200.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

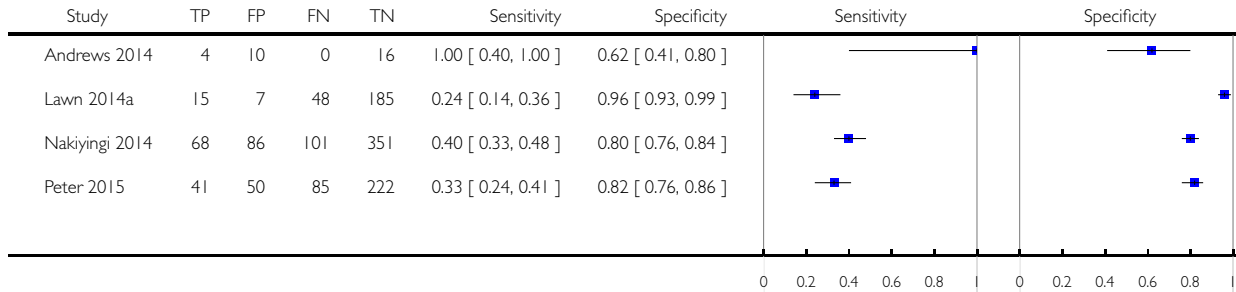
Test: 6 Diagnosis of TB against microbiological reference at Grade I, CD4 < 200



Test 7. Diagnosis of TB against microbiological reference at Grade I, CD4 > 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

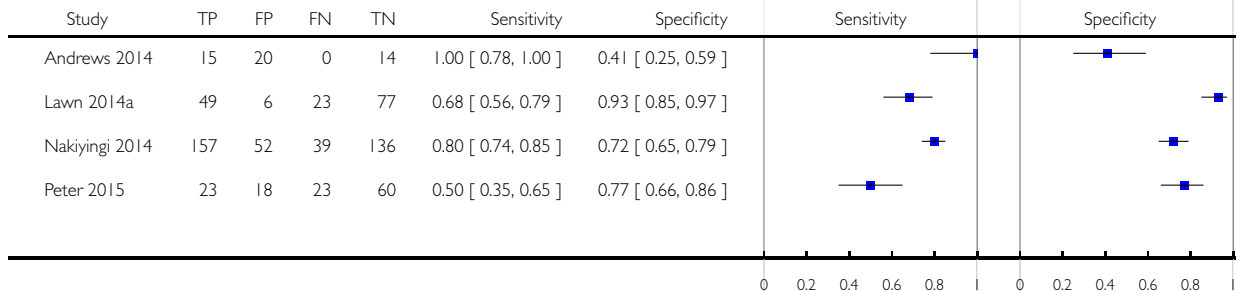
Test: 7 Diagnosis of TB against microbiological reference at Grade I, CD4 > 100



Test 8. Diagnosis of TB against microbiological reference at Grade I, CD4 < 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

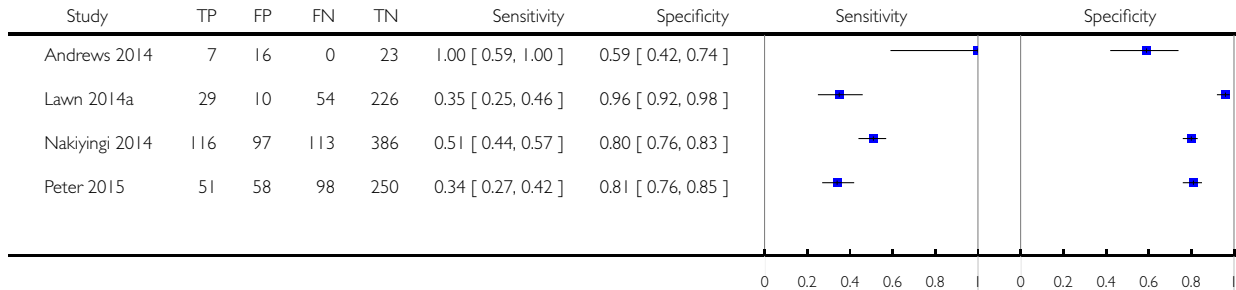
Test: 8 Diagnosis of TB against microbiological reference at Grade I, CD4 < 100



Test 9. Diagnosis of TB against microbiological reference at Grade I, CD4 > 50.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

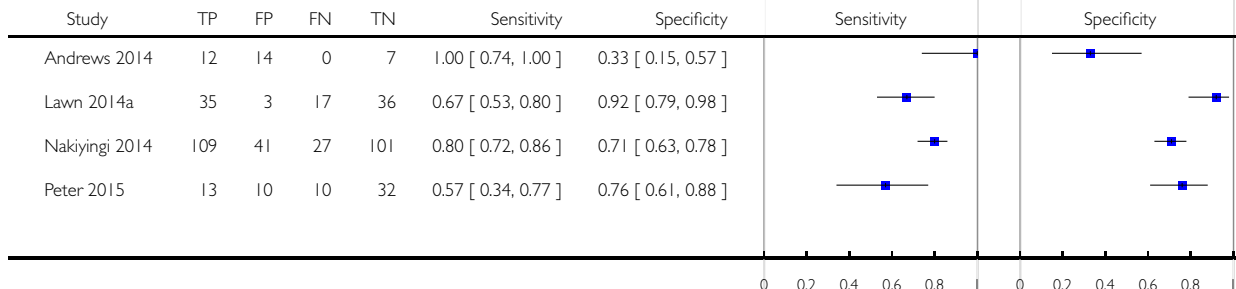
Test: 9 Diagnosis of TB against microbiological reference at Grade I, CD4 > 50



Test 10. Diagnosis of TB against microbiological reference at Grade I, CD4 < 50.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

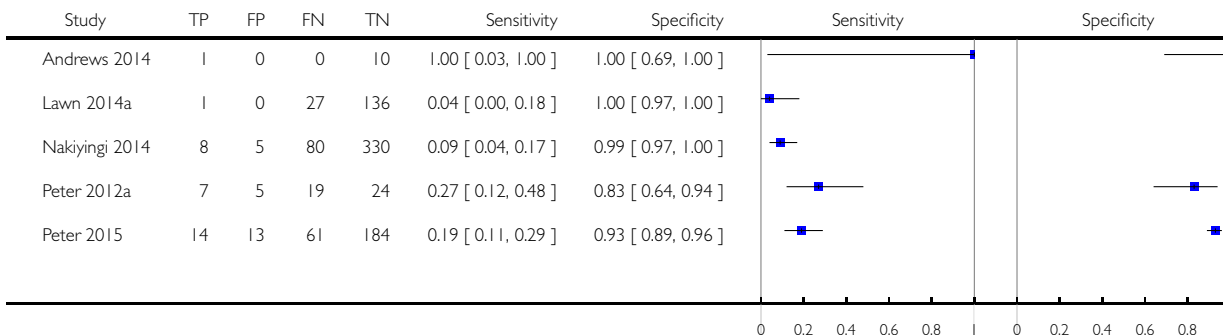
Test: 10 Diagnosis of TB against microbiological reference at Grade I, CD4 < 50



Test 11. Diagnosis of TB against microbiological reference at Grade 2, CD4 > 200.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

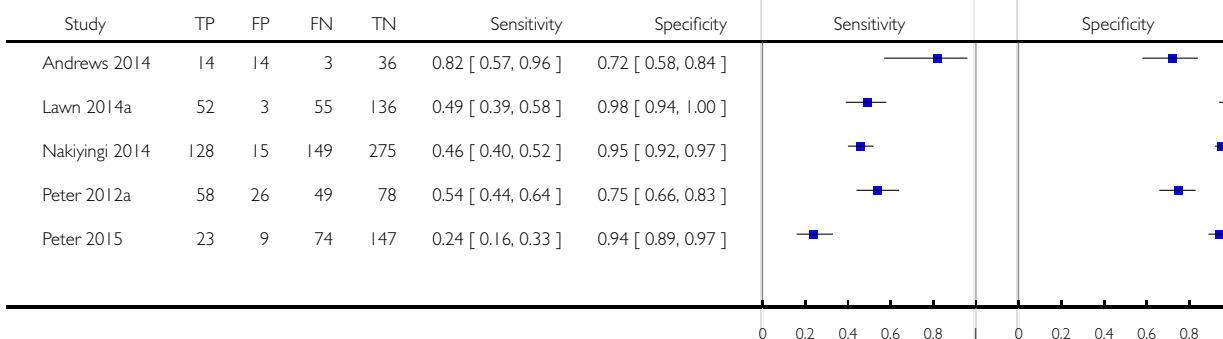
Test: 11 Diagnosis of TB against microbiological reference at Grade 2, CD4 > 200



Test 12. Diagnosis of TB against microbiological reference at Grade 2, CD4 < 200.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

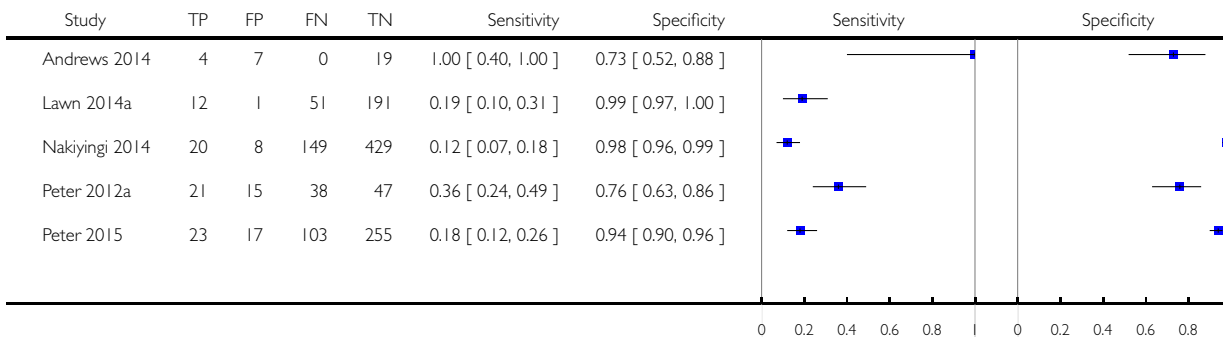
Test: 12 Diagnosis of TB against microbiological reference at Grade 2, CD4 < 200



Test 13. Diagnosis of TB against microbiological reference at Grade I, CD4 > 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

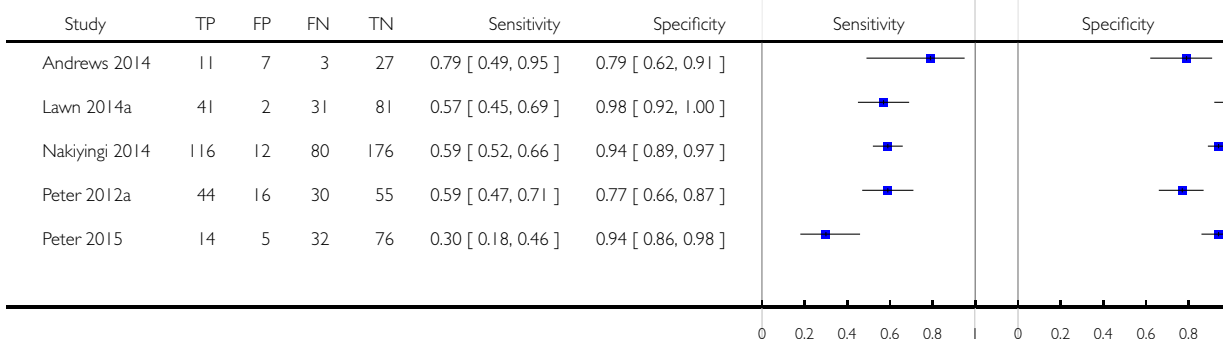
Test: 13 Diagnosis of TB against microbiological reference at Grade I, CD4 > 100



Test 14. Diagnosis of TB against microbiological reference at Grade I, CD4 < 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

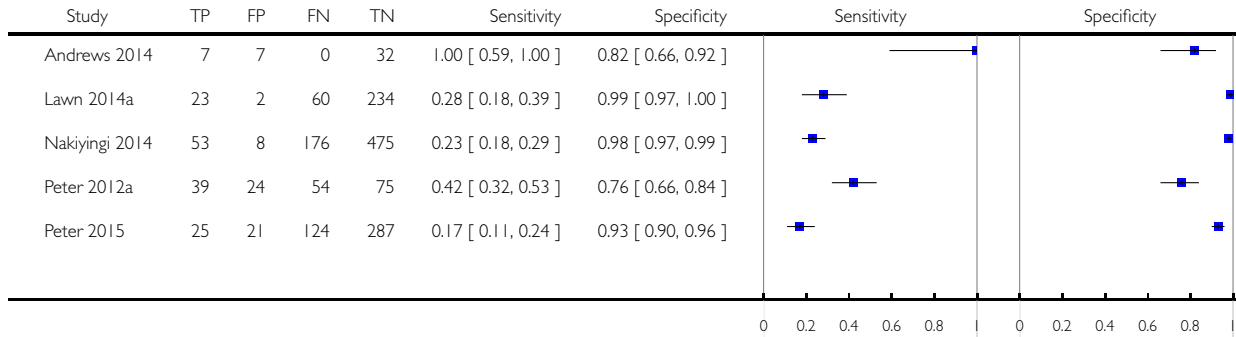
Test: 14 Diagnosis of TB against microbiological reference at Grade I, CD4 < 100



Test 15. Diagnosis of TB against microbiological reference at Grade I, CD4 > 50.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

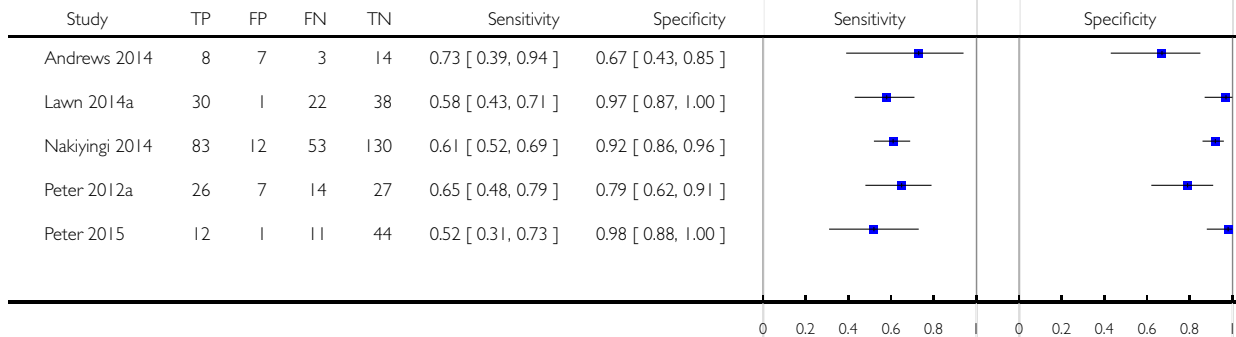
Test: 15 Diagnosis of TB against microbiological reference at Grade I, CD4 > 50



Test 16. Diagnosis of TB against microbiological reference at Grade I, CD4 < 50.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

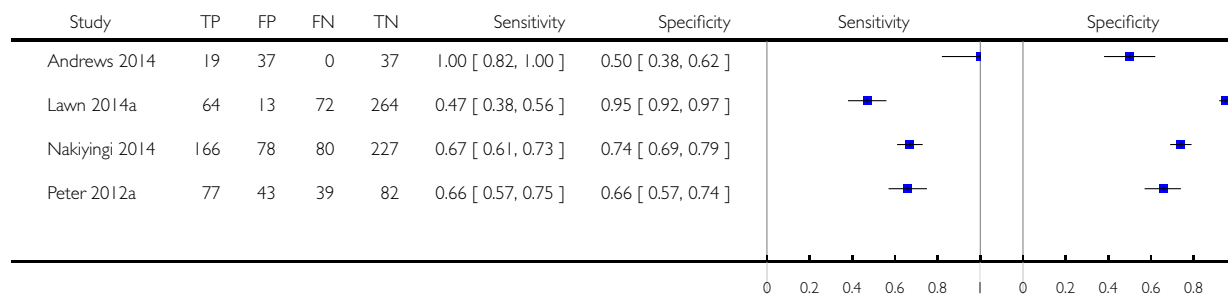
Test: 16 Diagnosis of TB against microbiological reference at Grade I, CD4 < 50



Test 17. Diagnosis of TB against microbiological reference at Grade 1: inpatients.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

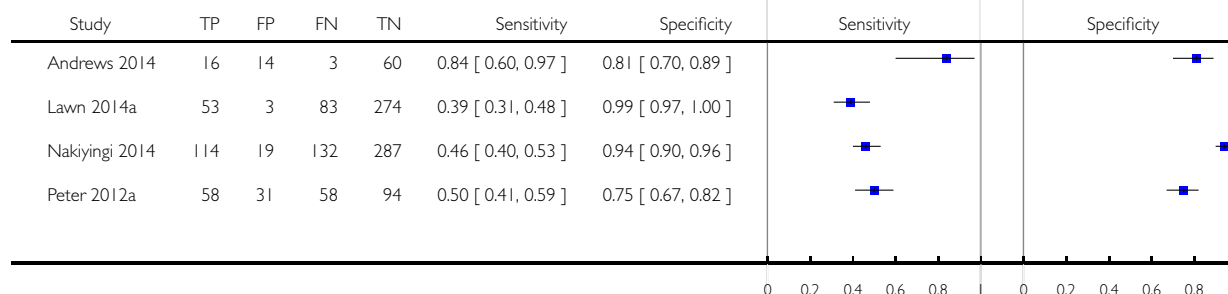
Test: 17 Diagnosis of TB against microbiological reference at Grade 1: inpatients



Test 18. Diagnosis of TB against microbiological reference at Grade 2: inpatients.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

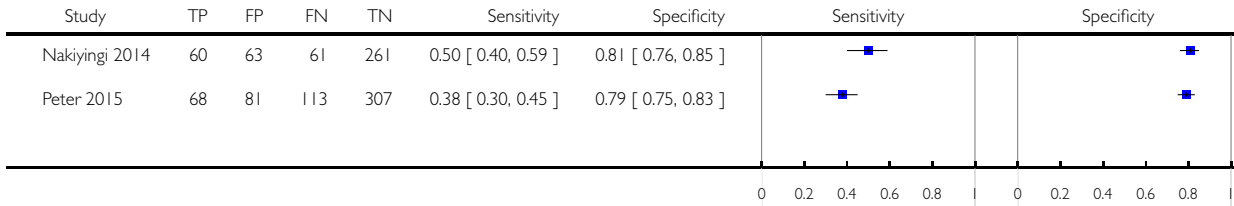
Test: 18 Diagnosis of TB against microbiological reference at Grade 2: inpatients



Test 19. Diagnosis of TB against microbiological reference at Grade 1: outpatients.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

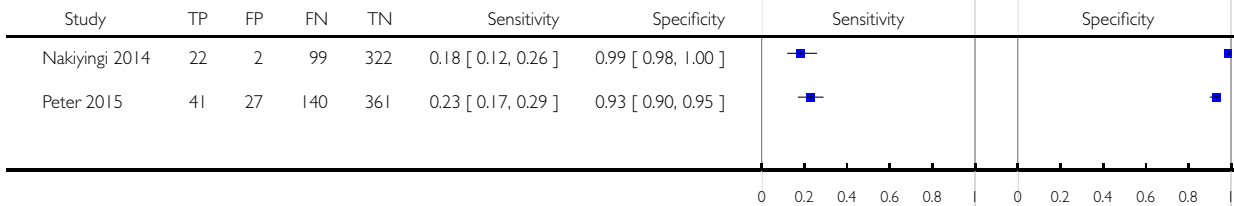
Test: 19 Diagnosis of TB against microbiological reference at Grade 1: outpatients



Test 20. Diagnosis of TB against microbiological reference at Grade 2: outpatients.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

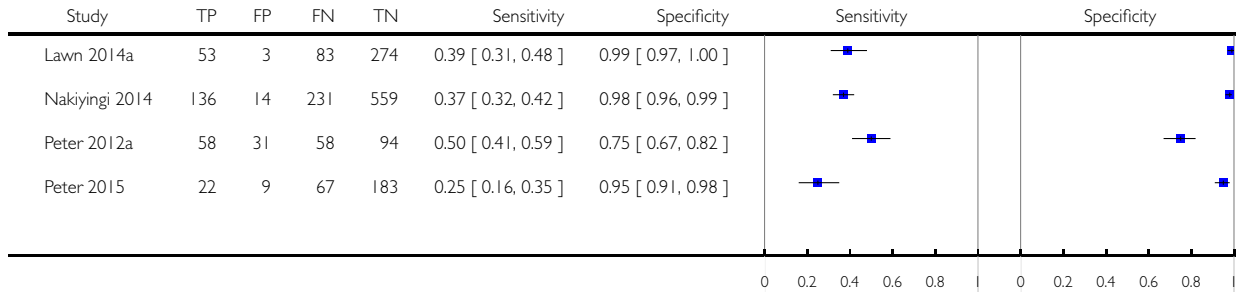
Test: 20 Diagnosis of TB against microbiological reference at Grade 2: outpatients



Test 21. Diagnosis of TB with LAM at Grade 2 (among studies comparing with smear).

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

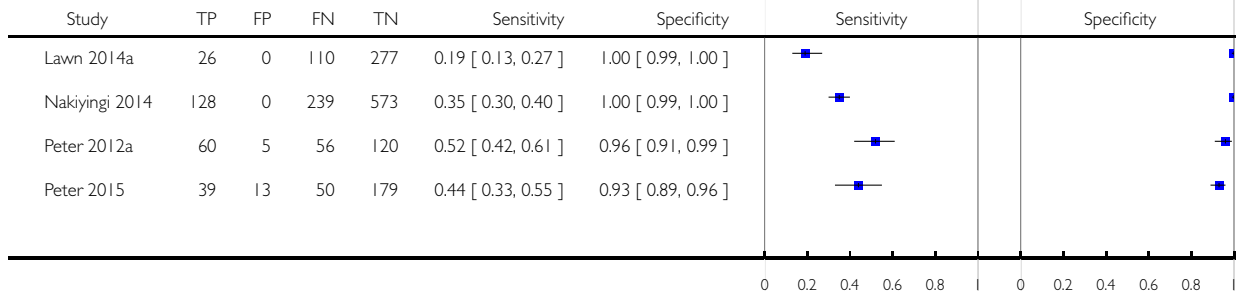
Test: 21 Diagnosis of TB with LAM at Grade 2 (among studies comparing with smear)



Test 22. Diagnosis of TB with smear (among studies comparing with LAM).

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

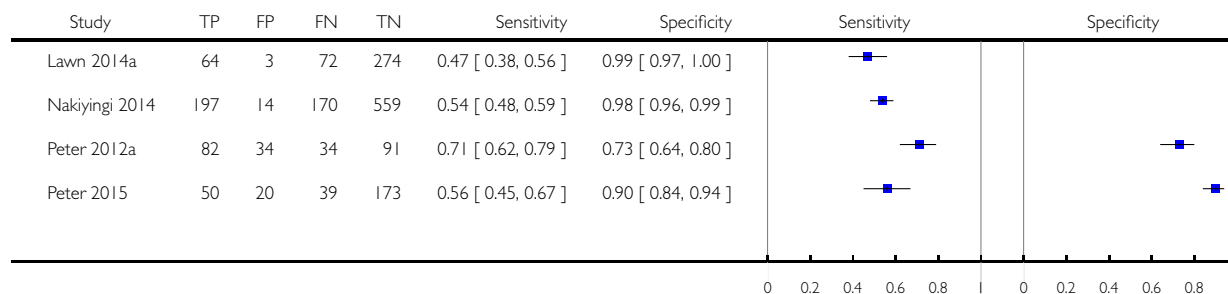
Test: 22 Diagnosis of TB with smear (among studies comparing with LAM)



Test 23. Diagnosis of TB with combination of LAM at Grade 2 with smear.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

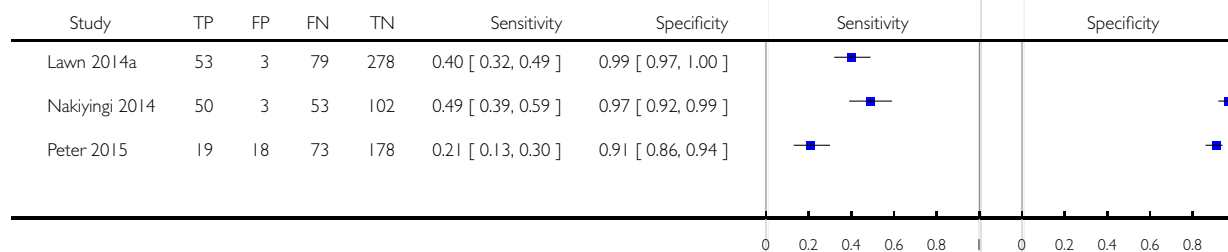
Test: 23 Diagnosis of TB with combination of LAM at Grade 2 with smear



Test 24. Diagnosis of TB with LAM at Grade 2 (among studies comparing with Xpert®).

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

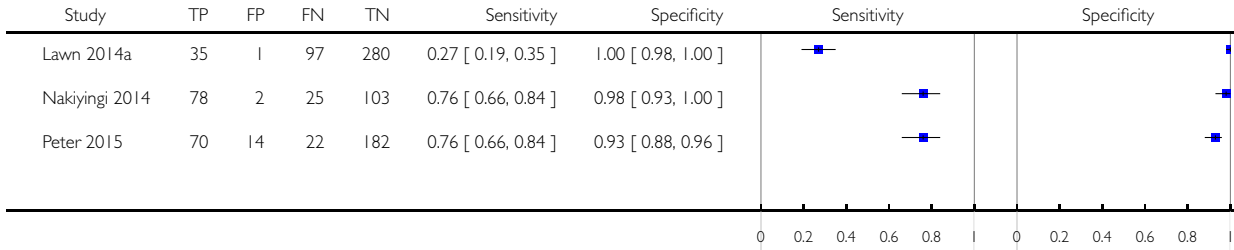
Test: 24 Diagnosis of TB with LAM at Grade 2 (among studies comparing with Xpert®)



Test 25. Diagnosis of TB with Xpert® (among studies comparing with LAM).

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

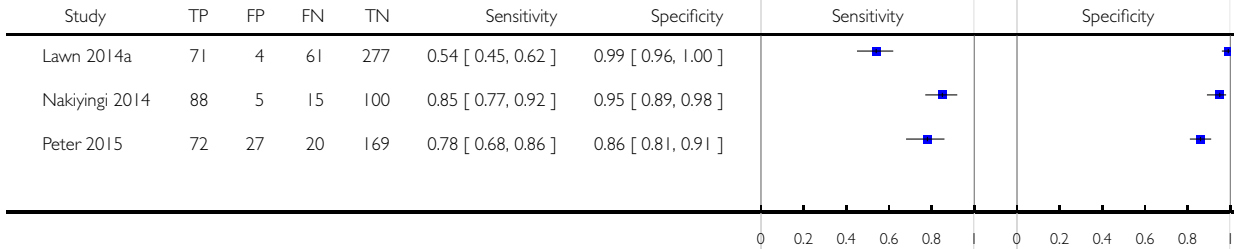
Test: 25 Diagnosis of TB with Xpert (among studies comparing with LAM)



Test 26. Diagnosis of TB with combination of LAM at Grade 2 with Xpert®.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

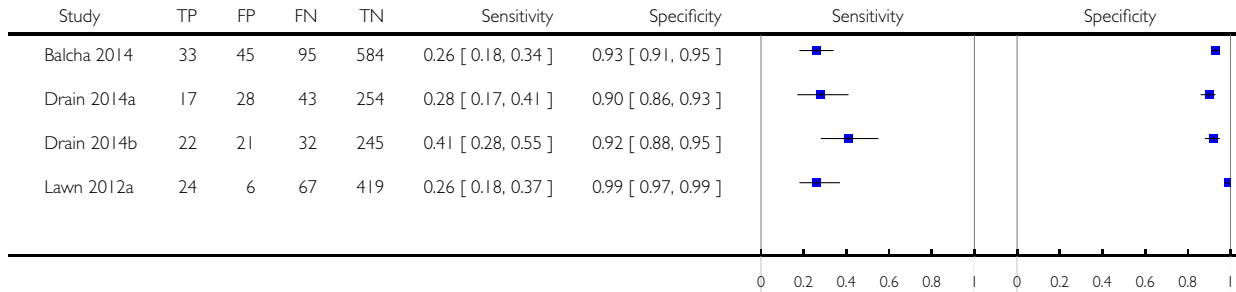
Test: 26 Diagnosis of TB with combination of LAM at Grade 2 with Xpert



Test 27. Screening for TB against microbiological reference at Grade 1: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

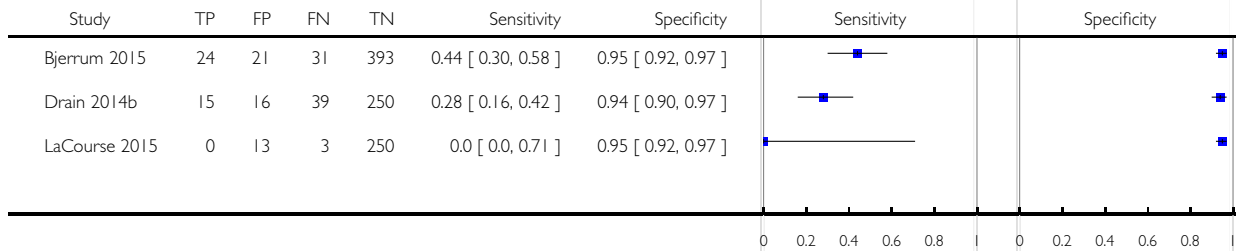
Test: 27 Screening for TB against microbiological reference at Grade 1: all participants



Test 28. Screening for TB against microbiological reference at Grade 2: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

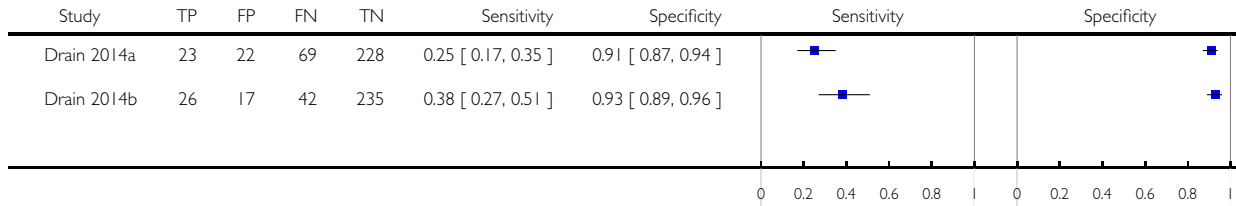
Test: 28 Screening for TB against microbiological reference at Grade 2: all participants



Test 29. Screening for TB against composite reference at Grade 1: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

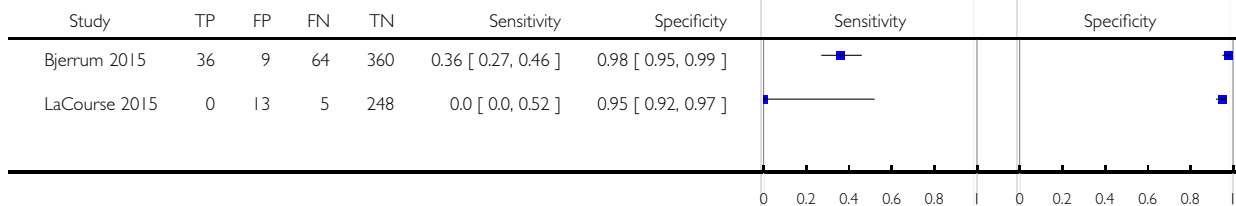
Test: 29 Screening for TB against composite reference at Grade 1: all participants



Test 30. Screening for TB against composite reference at Grade 2: all participants.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

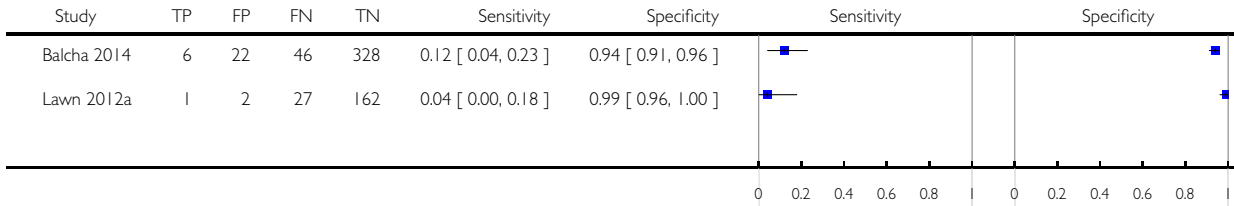
Test: 30 Screening for TB against composite reference at Grade 2: all participants



Test 31. Screening for TB against microbiological reference at Grade I, CD4 > 200.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

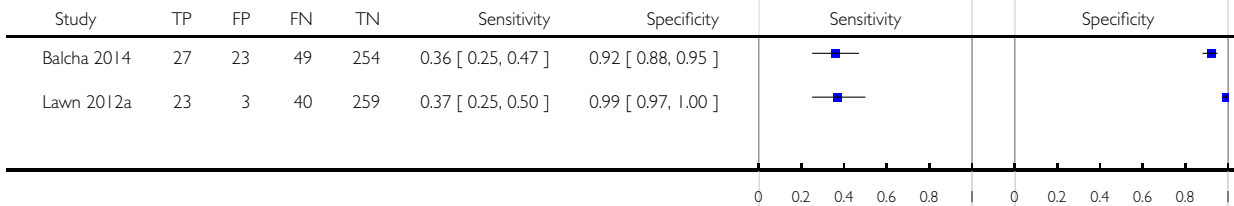
Test: 31 Screening for TB against microbiological reference at Grade I, CD4 > 200



Test 32. Screening for TB against microbiological reference at Grade I, CD4 < 200.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

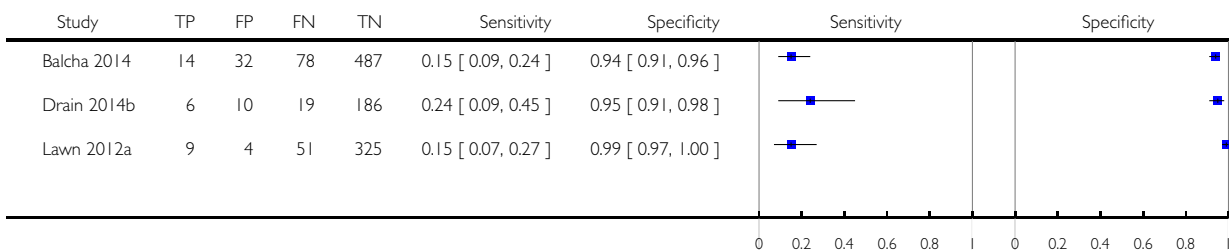
Test: 32 Screening for TB against microbiological reference at Grade I, CD4 < 200



Test 33. Screening for TB against microbiological reference at Grade I, CD4 > 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

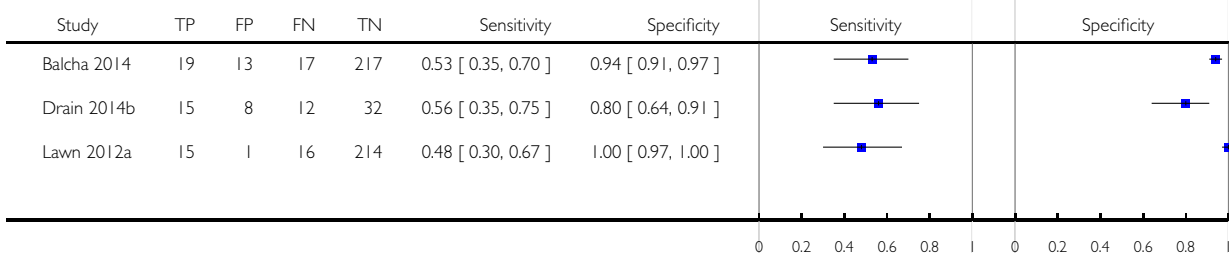
Test: 33 Screening for TB against microbiological reference at Grade I, CD4 > 100



Test 34. Screening for TB against microbiological reference at Grade I, CD4 < 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

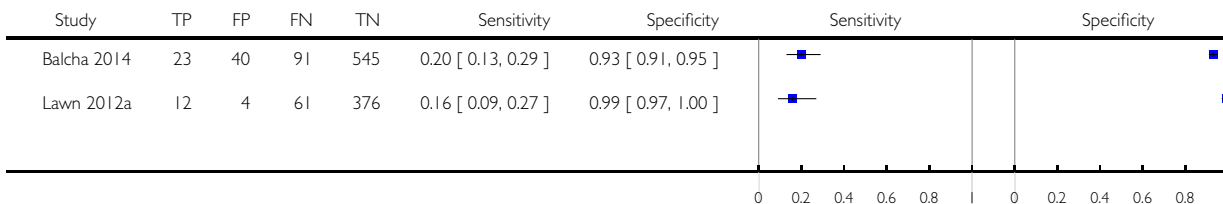
Test: 34 Screening for TB against microbiological reference at Grade I, CD4 < 100



Test 35. Screening for TB against microbiological reference at Grade I, CD4 > 50.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

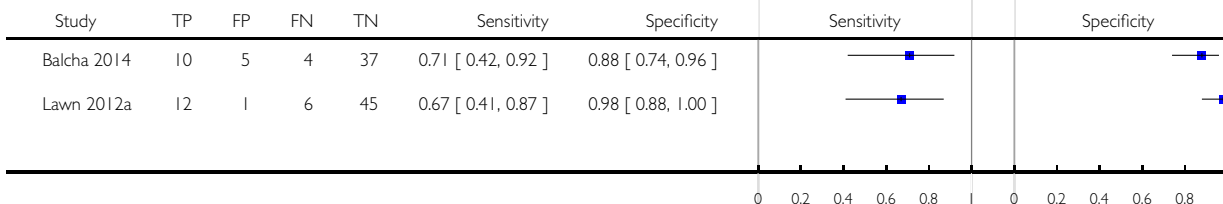
Test: 35 Screening for TB against microbiological reference at Grade I, CD4 > 50



Test 36. Screening for TB against microbiological reference at Grade I, CD4 < 50.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

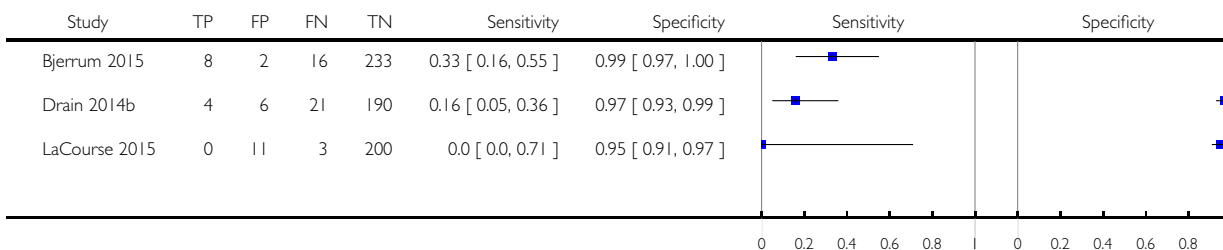
Test: 36 Screening for TB against microbiological reference at Grade I, CD4 < 50



Test 37. Screening for TB against microbiological reference at Grade 2, CD4 > 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

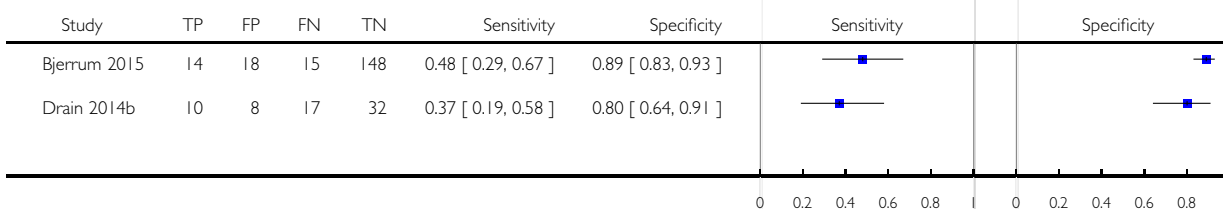
Test: 37 Screening for TB against microbiological reference at Grade 2, CD4 > 100



Test 38. Screening for TB against microbiological reference at Grade 2, CD4 < 100.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

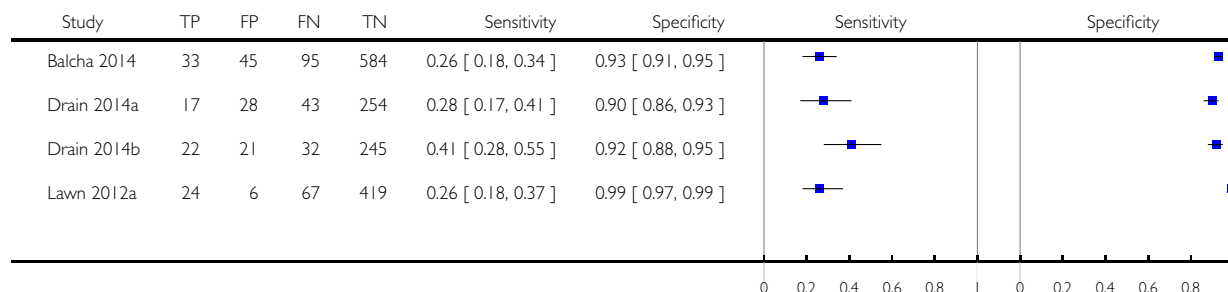
Test: 38 Screening for TB against microbiological reference at Grade 2, CD4 < 100



Test 39. Screening for TB against microbiological reference at Grade I: outpatients.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

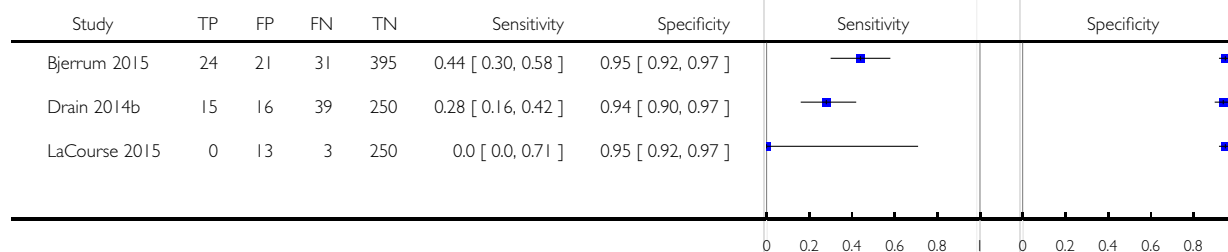
Test: 39 Screening for TB against microbiological reference at Grade I: outpatients



Test 40. Screening for TB with LAM at Grade 2 (among studies comparing with smear microscopy).

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

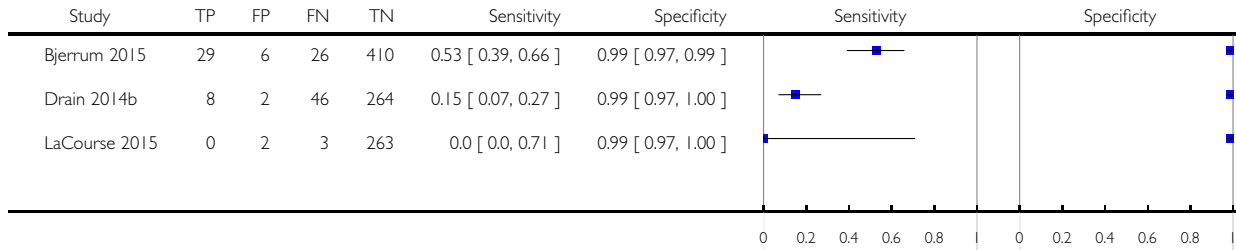
Test: 40 Screening for TB with LAM at Grade 2 (among studies comparing with smear microscopy)



Test 41. Screening for TB with smear microscopy (among studies comparing with LAM).

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

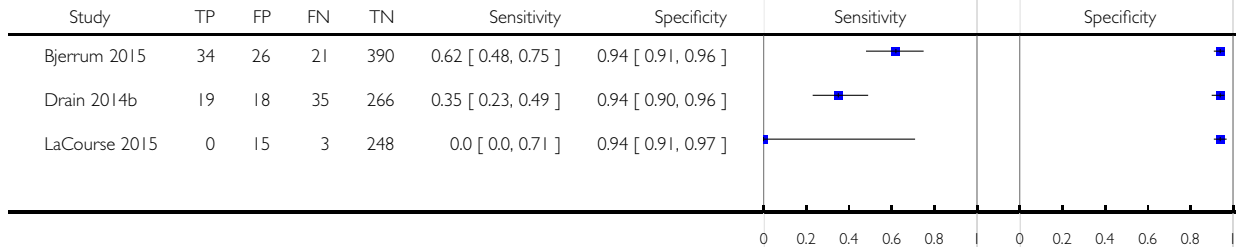
Test: 41 Screening for TB with smear microscopy (among studies comparing with LAM)



Test 42. Screening for TB with the combination of LAM (Grade 2) and smear microscopy.

Review: Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults

Test: 42 Screening for TB with the combination of LAM (Grade 2) and smear microscopy



ADDITIONAL TABLES

Table 1. LF-LAM pooled sensitivity and specificity for TB diagnosis, by grade

Type of analysis	Grade 1				Grade 2			
	Studies (total partici-	Partic- ipants with	Pooled sen- sitivity	Pooled specificity	Studies (to- tal partici-	Partic- ipants with	Pooled sen- sitivity	Pooled specificity

Table 1. LF-LAM pooled sensitivity and specificity for TB diagnosis, by grade (Continued)

	pants)	TB (%)	(95% CrI)	(95% CrI)	pants)	TB (%)	(95% CrI)	(95% CrI)
Overall accuracy, microbiological reference standard	6 studies (2402)	876 (36%)	59% (43 to 77)	78% (64 to 88)	5 studies (2313)	819 (35%)	45% (29 to 63)	92% (80 to 97)
Overall accuracy, composite reference standard ¹	3 studies (1585)	799 (50%)	Not applicable	Not applicable	3 studies (1586)	799 (50%)	Not applicable	Not applicable
Inpatient settings	4 studies (1298)	517 (40%)	69% (53 to 85)	76% (54 to 89)	4 studies (1299)	517 (40%)	53% (38 to 70)	90% (73 to 96)
Outpatient settings ²	2 studies (1014)	302 (20%)	Not applicable	Not applicable	2 studies (1014)	302 (30%)	Not applicable	Not applicable
CD4 > 200	4 studies (870)	192 (22%)	27% (13 to 49)	85% (70 to 93)	5 studies (925)	218 (24%)	15% (8 to 27)	96% (89 to 99)
CD4 ≤ 200	4 studies (1131)	499 (44%)	68% (46 to 88)	77% (59 to 89)	5 studies (1344)	605 (45%)	50% (35 to 67)	90% (78 to 95)
CD4 > 100	4 studies (1289)	362 (28%)	38% (25 to 59)	83% (66 to 92)	5 studies (1410)	421 (30%)	26% (16 to 46)	92% (78 to 97)
CD4 ≤ 100	4 studies (712)	329 (46%)	75% (56 to 89)	74% (54 to 87)	5 studies (859)	402 (47%)	56% (41 to 70)	90% (81 to 95)
CD4 > 50	4 studies (1534)	468 (31%)	49% (32 to 72)	82% (64 to 91)	5 studies (1726)	561 (33%)	34% (21 to 60)	93% (81 to 97)
CD4 ≤ 50	4 studies (467)	223 (48%)	77% (61 to 89)	72% (50 to 86)	5 studies (543)	262 (48%)	62% (49 to 73)	89% (77 to 95)
LF-LAM combined with sputum microscopy	1 study ³ (413)	136 (33%)	Not applicable	Not applicable	4 studies (1876)	708 (38%)	59% (47 to 70)	92% (73 to 97)
LF-LAM combined with sputum Xpert®	0 studies	Not applicable	Not applicable	Not applicable	3 studies (909)	327 (36%)	75% (61 to 87)	93% (81 to 97)

Abbreviations: CrI: credible interval; LF-LAM: lateral flow urine lipoarabinomannan assay; TB: tuberculosis.

¹For the composite reference test, sensitivity ranged from 23% to 60% and specificity ranged from 78% to 96% at grade 1; sensitivity ranged from 11% to 45% and specificity ranged from 96% to 98% at grade 2.

²For the outpatient setting, sensitivity ranged from 38% to 50% and specificity ranged from 79% to 81% at grade 1; sensitivity ranged from 18% to 23% and specificity ranged from 93% to 96% at grade 2.

³Sensitivity of smear was 19%, compared to 47% for LAM, and 54% for the combination of smear and LAM. Specificity of smear was 100%, compared to 95% for LAM, and 95% for the combination of smear and LAM.

Table 2. Comparison of mortality in LF-LAM-positive and LF-LAM-negative participants

Study	Study population	Mortality assessment	Mortality in LF-LAM positive	Mortality in LF-LAM negative
Balcha 2014	757 outpatients (screening)	148 TB cases	20% (7/35)	3% (3/113)
Bjerrum 2015	469 outpatients (screening)	469 enrollees	49% (22/45)	14% (59/424)
		55 TB cases	54% (13/24)	16% (5/31)
		39 TB starting treatment	32% (5/16)	4% (1/23)
Drain 2015	90 outpatients started on TB therapy	Baseline	31% (9/29); AHR 1.41	31% (16/61)
		2-month LF-LAM (grade 2)	50% (4/8) AHR 5.58	19% (12/65)
Lawn 2012c	535 ART enrollees (screening)	59 TB cases	22% (5/23)	0% (0/36)
Manabe 2014	351 hospitalized (diagnosis)	351 enrollees	28% (38/134)	17% (37/217)
		145 TB cases	28% (25/90)	13% (7/55)
		185 with no evidence of TB	34% (12/35)	19% (28/150)
Peter 2015	583 outpatients (diagnosis)	583 enrollees	25% (9/32)	11% (40/361)
		123 TB cases	35% (6/17)	14% (15/106)

Abbreviations: AHR: adjusted hazard ratio; ART: antiretroviral therapy; LF-LAM: lateral flow urine lipoarabinomannan assay; TB: tuberculosis.

APPENDICES

Appendix I. Minimum specifications for a point-of-care TB diagnostic test

Test specification	Minimum required value
Medical decision	Treatment initiation
Sensitivity: adults (for pulmonary TB only; regardless of HIV status)	Pulmonary TB <ul style="list-style-type: none"> ● 95% for smear positive, culture positive ● 60% to 80% for smear negative, culture positive (detection of extrapulmonary TB being a preferred but not minimal requirement)
Sensitivity: children (including extrapulmonary TB; regardless of HIV status)	<ul style="list-style-type: none"> ● 80% compared to culture of any specimen and ● 60% of probable TB (noting problem of lack of a gold standard)
Specificity: adults	<ul style="list-style-type: none"> ● 95% compared to culture
Specificity: children	<ul style="list-style-type: none"> ● 90% for culture-negative probable TB (noting problem of lack of a gold standard) ● 95% compared to culture
Time to results	3 hours maximum (patient must receive results the same day) (desirable would be < 15 minutes)
Throughput	20 tests/day minimum, by one laboratory technician
Specimen type	Adults: urine, oral, breath, venous blood, sputum (desired: non sputum-based sample type and use of finger prick instead of venous blood) Children: urine, oral, capillary blood (finger or heel prick)
Sample preparation	<ul style="list-style-type: none"> ● 3 steps maximum ● Safe: biosafety level 1 ● Ability to use approximate volumes (that is, no need for precise pipetting) ● Preparation that is not highly time sensitive
Number of samples	1 sample per test
Readout	<ul style="list-style-type: none"> ● Easy-to-read, unambiguous, simple “yes”, “no”, or “invalid” answer ● Readable for at least one hour
Waste disposal	<ul style="list-style-type: none"> ● Simple burning or sharps disposal; no glass component ● Environmentally acceptable disposal

(Continued)

Controls	<ul style="list-style-type: none"> • Positive control included in test kit • Quality control simpler and easier than with sputum smear microscopy
Reagents	<ul style="list-style-type: none"> • All reagents in self-contained kit • Kit contains sample collection device and water (if needed)
Storage/stability required	<ul style="list-style-type: none"> • Shelf life of 24 months, including reagents • Stable at 30°C, and at higher temperatures for shorter time periods • Stable in high humidity environments
Instrumentation	<ul style="list-style-type: none"> • If instrument needed, no maintenance required • Instrument works in tropical conditions • Acceptable replacement cost • Fits in backpack • Shock resistant
Power requirement	Can work on battery
Training	<ul style="list-style-type: none"> • 1 day maximum training time • Can be performed by any health worker
Cost	< USD 10 per test after scale-up

Copyright © [Batz 2011](#); reproduced with permission.
Abbreviations: TB, tuberculosis; USD, United States Dollar.

Appendix 2. Detailed search strategies

MEDLINE (PubMed) search history

Search	Query
#9	Search (#3) AND (#7) AND #8
#8	Search test OR assay OR antigen OR Ag OR lateral flow assay*OR urine antigen OR point of care Field: Title/Abstract
#7	Search (#4) OR #5 OR #6
#6	Search LAM ; Field: Title/Abstract
#5	Search “ lipoarabinomannan ” [Supplementary Concept]
#4	Search lipoarabinomannan ; Field: Title/Abstract
#3	Search (#1) OR #2

(Continued)

#2	Search tuberculosis Or TB Field: Title/Abstract
#1	Search (“ Tuberculosis ”[Mesh]) OR “ Mycobacterium tuberculosis ”[Mesh]

.....

Web of Science Core Collection - Indexes: SCI-EXPANDED, CPCI-S, Biosis previews

You searched for: **TOPIC:** (tuberculosis OR TB OR mycobacterium) *AND* **TOPIC:** (lipoarabinomannan OR LAM) *AND* **TOPIC:** (10. test OR assay OR antigen OR Ag OR lateral flow assay*OR urine antigen OR point of care) ...More **TOPIC:** (tuberculosis OR TB OR mycobacterium) *AND* **TOPIC:** (lipoarabinomannan OR LAM) *AND* **TOPIC:** (test OR assay OR antigen OR Ag OR lateral flow assay*OR urine antigen OR point of care)

.....

SCOPUS

(TITLE-ABS-KEY (tuberculosis OR tb) AND TITLE-ABS-KEY (lipoarabinomannan OR lam) AND PUBYEAR > 2013) AND (test OR diagnos* OR urine OR assay)

<p>Database: EMBASE search strategy</p> <p>-----</p> <p>1 tuberculosis.mp. or tuberculosis/ or Mycobacterium tuberculosis/ 2 lipoarabinomannan.mp. or lipoarabinomannan/ 3 LAM.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 4 2 or 3 5 1 and 4 6 (test or assay or antigen or Ag or lateral flow assay* or urine antigen or point of care).mp 7 4 and 5</p>
<p>CIDG Specialized Register, LILACS, Medion, Proquest dissertations, Current Controlled trials, WHO trials register Tuberculosis AND (lipoarabinomannan OR LAM)</p>

Appendix 3. Data collection form

Lateral flow urine lipoarabinomannan assay for diagnosing active tuberculosis in people living with HIV		
Data form		
1	First author	
2	Corresponding author and email	
3	Title of study	
4	Publication status of study	1. Published 2. Unpublished, what is the anticipated study completion date?

(Continued)

5	For the diagnosis of pulmonary TB, what reference standard was used to identify TB and not TB?	1. Sputum: solid culture 2. Sputum: liquid culture 3. Sputum: both solid and liquid culture 4. Nucleic acid amplification test, specify 6. Other, specify
6	Was sputum induction performed for individuals unable to produce expectorated sputum?	1. Yes, specify N/% (____) requiring sputum induction 2. No
7	Were patients without sputum specimens (for example, no expectorated, no induced sputum) included in this study?	1. Yes, specify N/% of patients included without sputum____ 2. No, specify N/% of patients excluded due to lack of sputum____
8	Were non-pulmonary specimens evaluated to allow diagnosis of extrapulmonary TB?	1. All participants received testing of non-pulmonary samples, please specify sites/fluids: 2. Some participants received testing of non-pulmonary samples, please specify which patients were tested, and sites/fluids: 3. Extrapulmonary TB was not evaluated 6. Other, please specify:
9	For the diagnosis of extrapulmonary TB, what tests were used to identify TB and not TB (circle all that apply)?	1. Solid culture 2. Liquid culture 3. Both solid and liquid culture 4. Nucleic acid amplification test, specify 6. Other, specify_____
10	How many sputum specimens were obtained in order to detect pulmonary TB?	Mean (standard deviation (SD)) number of specimens per patient = Not applicable
11	If applicable, how many specimens from body sites or fluids other than sputum were obtained in order to detect extrapulmonary TB?	Mean (SD) number of specimens per patient = Not applicable
12	In which country or countries was the study conducted?	Please list all countries:
13	What was the clinical setting of the study?	1. Outpatient 2. Inpatient 3. Both out-patient and in-patient 6. Other, describe:

(Continued)

13a	How would you describe the health facility where the study took place?	1. Primary care clinic, stand alone 2. Primary care clinic, connected to a referral hospital 3. Referral hospital 6. Other, describe:
14	Please select the statement that best describes the selection of participants into your study	1. HIV-positive participants with signs or symptoms suggestive of active TB were tested using LF-LAM. Please provide study definition of 'signs and symptoms': 2. A predetermined target population of HIV-positive individuals, with or without signs or symptoms of active TB, were tested using LF-LAM: please specify target population 3. Both 1 and 2 4. Neither 1 nor 2. This is what we did:
15	What was the manner of participant selection into the study?	1. Consecutive 2. Random 3. Convenience 6. Other, specify
16	Was LF-LAM performed on fresh or stored urine	1. Fresh 2. Stored, specify type of storage (e.g. frozen) 3. Both fresh and stored
17	Which package insert reference card was used to grade LF-LAM results in the study?	1. "Original": 5 grades, prior to Jan 2014 2. "New": 4 grades, after Jan 2014 date 9. I do not know
18	What specific grade LF-LAM value was used to define positivity for the primary analysis?	1. Grade 1 2. Grade 2 6. Other, specify 9. I do not know
18	Was data collected on alternative thresholds (grades) to define positivity	1. Yes, please specify 2. No
19	Was LF-LAM result interpreted without knowledge of the result of the reference standard result?	1. Yes 2. No
20	Was the reference standard result interpreted without knowledge of the result of LF-LAM?	1. Yes 2. No

(Continued)

21	Did more than one individual read the LF-LAM result?	1. Yes: please specify kappa or other measure of correlation 2. No
22	Were clinical outcomes evaluated, such as time to diagnosis or mortality?	1. Yes, If yes, please list the outcomes 2. No
23	Were there any LF-LAM results that were invalid (no bar in control window)?	1. Yes, if yes, please list number of invalid tests: _____ Please list if invalid tests were repeated: a) Yes b) No 2. No
24	Were any LF-LAM results uninterpretable for other reasons?	1. Yes, if yes please list number of uninterpretable tests and reasons: 2. No
25	What was the median or mean CD4 of participants?	Median (IQR) or mean (SD)
26	What percentage of participants was female?	
27	What was the median or mean age of participants?	Median (IQR) or mean (SD)

Abbreviations: IQR, interquartile range; SD, standard deviation.

Table 1a. Microbiological reference standard

TB is defined as positive culture or NAAT from sputum or any body fluid or site.

Not TB is defined as negative cultures or NAATs from sputum or any body fluid or site.

□ Please check box if data not available

LAM result (positive at grade 1)	TB	Not TB	Total	Unknown
Positive				
Negative				
Total				
Uninterpretable or invalid				

□ Please check box if data not available

LAM result (positive at grade 2)	TB	Not TB	Total	Unknown
Positive				
Negative				
Total				
Uninterpretable or invalid				

Table 1b. Composite reference standard

TB defined as satisfying at least one of the following components:

(a) a positive mycobacterial culture or NAAT from any body fluid or site; (b) positive sputum smear microscopy; or (c) a clinical diagnosis of TB, in which the patient was started on TB treatment by a health care provider, and, after at least one month of follow-up, was classified as 'TB' or 'Not TB'.

'Not TB' defined as

(a) patients had no positive culture, NAAT, or smear for TB **AND**

(b) patients were not started on anti-TB treatment by health care providers

(c) patients were alive at one month of follow-up and were clinically classified by a health care provider as Not TB

People who do not meet the criteria for 'TB' or 'Not TB' based on these definitions will be excluded as **'unclassifiable'**.

Please check box if data not available or not applicable

LAM result (Positive at grade 1)	TB	Not TB	Total	Unknown or unclassifiable
Positive				
Negative				
Total				
Uninterpretable or invalid				

Please check box if data not available or not applicable

LAM result (Positive at grade 2)	TB	Not TB	Total	Unknown or unclassifiable
Positive				
Negative				
Total				
Uninterpretable or invalid				

Please fill in the cells in the following table for study participants stratified by CD4 count

Please check box if CD4 stratified data are not available

For Tables 2a through 5a, use the microbiological reference standard

Table 2a. CD4 > 200

LAM result Grade ____	TB	Not TB	Total	Unknown
Positive				
Negative				
Total				
Uninterpretable or invalid				

Table 3a. CD4 101 to 200

LAM result Grade ____	TB	Not TB	Total	Unknown
Positive				
Negative				
Total				
Uninterpretable or invalid				

Table 4a. CD4 51 to 100

LAM result Grade ____	TB	Not TB	Total	Unknown
Positive				
Negative				
Total				
Uninterpretable or invalid				

Table 5a. CD4 ≤ 50

LAM result Grade ____	TB	Not TB	Total	Unknown
Positive				
Negative				
Total				
Uninterpretable or invalid				

Appendix 4. QUADAS-2

Domain 1: patient selection

Risk of bias: could the selection of patients have introduced bias?

Signalling question 1: Was a consecutive or random sample of patients enrolled?

We scored 'yes' if the study enrolled a consecutive or random sample of eligible participants; 'no' if the study selected participants by convenience; and 'unclear' if the study did not report the manner of participant selection or we could not tell.

Signalling question 2: Was a case-control design avoided?

We scored 'yes' to all included studies given that we are excluding case-control study designs.

Signalling question 3: Did the study avoid inappropriate exclusions?

We scored 'yes' to studies which included: a) all HIV-positive participants regardless of CD4 count and b) participants who were unable to produce sputum (expectorated or induced). We scored 'no' if studies excluded participants on the basis of CD4 count or the inability to produce sputum (no attempts at sputum induction). We also scored 'no' if studies excluded patients presumed to have extrapulmonary TB. We scored 'unclear' if we could not tell.

Applicability: Are there concerns that the included patients and setting do not match the review question?

We were interested in how LF-LAM performs in patients whose urine specimens were evaluated as they would be in routine practice. We expected to score most studies as 'low concern' since we planned to determine test accuracy separately for TB screening and TB diagnosis.

For LF-LAM used as a TB diagnostic test, we scored 'high concern' if the study participants did not resemble people with presumed TB; 'low concern' if the study population did resemble a population with presumed TB, and 'unclear concern', if we could not tell.

For LF-LAM used as a TB screening test, we have defined 'screening' for active TB in accordance with World Health Organization (WHO) guidance, as "the systematic identification of people with suspected active TB, in a predetermined target group, using tests, examinations or other procedures that can be applied rapidly ([WHO Tuberculosis Screening 2013](#))". We scored 'low concern' for studies in which the LF-LAM was performed uniformly within the predetermined study target populations of HIV-infected individuals, 'high concern' if LF-LAM was not performed uniformly within the predetermined study target populations of HIV-infected individuals, and 'unclear concern' if we could not tell.

Domain 2: index test

Risk of bias: could the conduct or interpretation of the index test have introduced bias?

Signalling question 1: were the index test results interpreted without knowledge of the results of the reference standard?

We answered 'yes' if the study interpreted the result of LF-LAM blinded to the result of the reference standard; we answered 'no' if the study did not interpret the result of LF-LAM blinded to the result of the reference standard. We answered 'yes' for studies in which LF-LAM was performed on fresh specimens, since reference standard results would be unavailable at the time of test interpretation. We answered 'unclear' if stored specimens were tested or we could not tell if the index test results were interpreted without knowledge of the reference standard results.

Signalling question 2: if a LF-LAM threshold was used to define positivity, was it prespecified?

We answered 'yes' if the threshold was prespecified, 'no' if the threshold was not prespecified, and 'unclear' if we could not determine if the threshold was prespecified or not.

Applicability: are there concerns that the index test, its conduct, or its interpretation differ from the review question?

If index test methods vary from those specified in the review question, concerns about applicability may exist. We judged 'high concern' if the test procedure was inconsistent with the manufacturer recommendations, 'low concern' if the test procedure was consistent with the manufacturer recommendations, and 'unclear concern' if we could not tell.

Domain 3: reference standard

Risk of bias: 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?

We considered this question separately for each reference standard.

a) Microbiological reference standard

HIV-infected TB patients may have pulmonary TB, extrapulmonary TB, or both pulmonary and extrapulmonary TB. A microbiological reference standard, primarily culture, is considered the gold standard for TB. Due to the difficulties in diagnosing HIV-associated TB, it is recommended that multiple cultures from sputum and other specimens be evaluated.

We answered 'yes' when appropriate specimens were obtained for the diagnosis of HIV-associated TB. For presumed pulmonary TB, sputum specimens should be obtained for culture, NAAT, or both culture and NAAT. If the patient cannot produce sputum, induced sputum should be performed. For presumed extrapulmonary TB, specimens should be consistent with Standard 4 of the International Standards for TB Care which states: "For all patients, including children, suspected of having extrapulmonary tuberculosis, appropriate specimens from the suspected sites of involvement should be obtained for microbiological and histological examination" (TB CARE I 2014). We answered yes if multiple specimens were collected from different sites for extrapulmonary TB. An Xpert® MTB/RIF test is recommended as the preferred initial microbiological test for suspected TB meningitis because of the need for a rapid diagnosis". We also answered 'yes' if studies followed a standardized approach of collecting appropriate specimens from "suspected sites of involvement", for example, blood or lymph nodes on all patients.

We answered 'no' when the reference standard was restricted to sputum specimens or the reference standard was restricted to extrapulmonary specimens (for example, urine, blood, etc.). We also answered 'no' if a consistent approach was not followed for all patients (for example, some but not all patients with presumed TB lymphadenitis receive lymph node tissue sampling). We answered 'unclear' if we could not tell.

b) Composite reference standard

Considerable uncertainty surrounds a clinical diagnosis of TB. Therefore, a reference standard that uses clinical TB (in culture-negative patients) is considered a lower quality reference standard and may incorrectly classify patients without TB disease as having TB disease. We scored 'unclear' for all studies that evaluated a combination of microbiological tests or clinical features to confirm TB.

Signalling question 2: were the reference standard results interpreted without knowledge of the results of the index test?

We answered 'yes' if the study interpreted the result of the reference standard blinded to the result of LF-LAM, or if the reference standard result was reported on an automated instrument; 'no' if the study did not interpret the result of the reference standard blinded to the result of LF-LAM, and 'unclear' if we could not tell.

Applicability: are there concerns that the target condition as defined by the reference standard does not match the question?

In general, we thought there was low concern for almost included studies based on the current definitions of the reference standard. We judged 'high concern' if included studies did not speciate mycobacteria isolated in culture, 'low concern' if speciation was performed, and 'unclear' if we could not tell. We also judged high concern if there was no protocol to ensure a minimum standard of testing with a reference standard.

Domain 4: Flow and timing

Risk of bias: could the patient flow have introduced bias?

Signalling question 1: was there an appropriate interval between the index test and reference standard?

We expected urine specimens for LF-LAM and the reference standards to be obtained at the same time and answered 'yes' for all studies that meet this criterion, or if index and reference standard tests were performed on specimens collected no greater than seven days apart. We chose seven days as a time period during which either treatment of TB or natural progression of TB without treatment could impact test results. We answered 'no' if specimens were collected for index and reference standard tests greater than seven days apart, and 'unclear' if we could not tell.

Signalling question 2: did all patients receive the same reference standard?

We considered this question separately for each reference standard.

a) Microbiological reference standard

We answered 'yes' if all participants in the study received the reference standard to confirm TB; 'no' if not all patients received the reference standard to confirm TB, and 'unclear' if we could not tell.

b) Composite reference standard

We scored 'unclear' for all studies because, for a given study, it is unlikely that we will know if all participants received the same component tests and if these component tests were interpreted and combined in a fixed way for all participants.

Signalling question 3: were all patients included in the analysis?

We determined the answer to this question by comparing the number of participants enrolled in the study with the number of participants included in the two-by-two tables. We answered 'yes' if all participants enrolled in the study were tested with results presented and accounted for. We answered 'no' if participants meeting enrolment criteria were not tested or results were not presented, and 'unclear' if we could not tell.

Appendix 5. Statistical appendix

We list here two types of WinBUGS programs used to fit the bivariate meta-analysis models in this review: i) a program for estimating the accuracy of the index test (T2), ii) a program for estimating the accuracy of an alternative test (T1) and a composite test (T3) based on the combination of the index and the alternative tests. In the subsections below, we first describe the likelihood and prior distribution for each type of model followed by the WinBUGS program.

As is usual with Bayesian models, initial values must be provided for all unknown parameters. For all programs, we selected three independent sets of initial values for most parameters using the in-built `gen.inits()` function within WinBUGS. The Gelman-Rubin statistic within the WinBUGS program was used to assess convergence. We did not observe any convergence problems for the analyses presented. We treated the first 3000 iterations as burn-in iterations and dropped them. We obtained summary statistics based on a total of 15,000 iterations resulting from the three separate chains.

A. Estimation of index test accuracy

Notation: in the i -th study the cells in the cross-tabulation between the index and reference tests are denoted by TP_i , FP_i , TN_i , FN_i . The sensitivity in i -th study is denoted by se_i and the specificity by sp_i .

We denote the Binomial probability distribution with sample size N and probability p as $\text{Binomial}(p, N)$, the Bivariate Normal probability distribution with mean vector μ and variance-covariance matrix S as $\text{BVN}(\mu, S)$, the univariate Normal distribution with mean m and variance s by $\text{N}(m, s)$ and the Uniform probability distribution between a and b by $\text{Uniform}(a, b)$. Note that logit refers to log odds.

Likelihood:

Within studies:

$TP_i \sim \text{Binomial}(se_i, TP_i + FN_i)$, and

$TN_i \sim \text{Binomial}(sp_i, TN_i + FP_i)$

Between studies:

The bivariate vector $(\text{logit}(se_i), \text{logit}(sp_i)) \sim \text{BVN}(\mu = (\mu_1, \mu_2), S)$ where

S is a 2×2 matrix with entries

$S[1,1] = \text{variance of } \text{logit}(se_i) = \sigma_1^2$,

$S[2,2] = \text{variance of } \text{logit}(sp_i) = \sigma_2^2$ and

$S[1,2] = S[2,1] = \text{covariance between } \text{logit}(se_i) \text{ and } \text{logit}(sp_i) = \rho \times \sigma_1 \times \sigma_2$

and ρ is the correlation between $\text{logit}(se_i)$ and $\text{logit}(sp_i)$ across studies.

The pooled sensitivity is given by $1/(1+\exp(-\mu_1))$, and the pooled specificity is given by $1/(1+\exp(-\mu_2))$.

Prior distributions:

μ_1 and $\mu_2 \sim \text{N}(m=0, s=4)$,

$\rho \sim \text{Uniform}(-1, 1)$

$(1/\sigma_1^2)$ and $(1/\sigma_2^2) \sim \text{Gamma}(\text{shape}=2, \text{rate}=0.5)$

A.1 WinBUGS program for estimating a bivariate hierarchical meta-analysis model for sensitivity and specificity of the index test.

Observed data must be provided for L (the number of studies), and TP, FN, FP and TN in each study.

```
-----
model {
for(i in 1:L) { ## L is the number of studies in the Meta-analysis
# Likelihood
pos[i]<-TP[i]+FN[i]
neg[i]<-TN[i]+FP[i]
TP[i] ~ dbin(se[i],pos[i])
TN[i] ~ dbin(sp[i],neg[i])
logit(se[i])<-l[i,1]
logit(sp[i])<-l[i,2]
l[i,1:2] ~ dnmnorm(mu[1:2], T[1:2, 1:2])
}
# Prior Distributions
mu[1] ~ dnorm(0,0.25)
mu[2] ~ dnorm(0,0.25)
```

```

T[1:2,1:2]<-inverse(S[1:2,1:2])
# Between-study variance-covariance matrix
S[1,1] <- sigma[1]*sigma[1]
S[2,2] <- sigma[2]*sigma[2]
S[1,2] <- rho*sigma[1]*sigma[2]
S[2,1] <- rho*sigma[1]*sigma[2]
# prec is the between-study precision in the logit(sensitivity) and logit(specificity)
# rho is the correlation between logit(sensitivity) and logit(specificity) across studies
prec[1] ~ dgamma(2,0.5)
prec[2] ~ dgamma(2,0.5)
rho ~ dunif(-1,1)
sigma[1]<-pow(prec[1],-0.5)
sigma[2]<-pow(prec[2],-0.5)
# Pooled sensitivity and specificity
Pooled`S<-1/(1+exp(-mu[1]))
Pooled`C<-1/(1+exp(-mu[2]))
}

```

B. Estimation of the accuracy of an alternative test and a composite test

Notation: In the i -th study the four cells in the cross-tabulation between the index and alternative tests are denoted by $T12 \cdot D_i$ among disease positive and by $T12 \cdot ND_i$ among disease negative. The sensitivity and specificity of the alternative test in i -th study are denoted by se_{1i} and sp_{1i} , respectively. The sensitivity and specificity of the index test in i -th study are denoted by se_{2i} and sp_{2i} , respectively. The sensitivity and specificity of the composite based on the index and the alternative tests in i -th study are denoted by se_{3i} and sp_{3i} , respectively.

We denote the Multinomial probability distribution with sample size N and probability p as $\text{Multinomial}(p, N)$. For other probability distributions we use the same notation as for the previous section.

Likelihood:

Within studies:

$T12 \cdot D_i \sim \text{Multinomial}(pD_i, D_i)$, and

$T12 \cdot ND_i \sim \text{Multinomial}(pND_i, ND_i)$

The multinomial probabilities can be expressed in terms of the sensitivities and specificities of T1 and T2 and the covariance between T1 and T2 in the disease positive as follows:

$pD_i[1] = se_{1i} * se_{2i} + covs[i] \# P(\tau_1=1, \tau_2=1 \mid D=1)$

$pD_i[2] = se_{1i} * (1 - se_{2i}) - covs[i] \# P(\tau_1=1, \tau_2=0 \mid D=1)$

$pD_i[3] = (1 - se_{1i}) * se_{2i} - covs[i] \# P(\tau_1=0, \tau_2=1 \mid D=1)$

$pD_i[4] = (1 - se_{1i}) * (1 - se_{2i}) + covs[i] \# P(\tau_1=0, \tau_2=0 \mid D=1)$

$pND_i[1] = sp_{1i} * sp_{2i} \# P(\tau_1=1, \tau_2=1 \mid D=0)$

$pND_i[2] = sp_{1i} * (1 - sp_{2i}) \# P(\tau_1=1, \tau_2=0 \mid D=0)$

$pND_i[3] = (1 - sp_{1i}) * sp_{2i} \# P(\tau_1=0, \tau_2=1 \mid D=0)$

$pND_i[4] = (1 - sp_{1i}) * (1 - sp_{2i}) \# P(\tau_1=0, \tau_2=0 \mid D=0)$

The sensitivity and specificity of the index test can be expressed in terms of the sensitivities and specificities of the composite and alternative tests. Doing so allows us to estimate the pooled sensitivity and specificity of the composite and alternative tests. The form of the between-studies likelihoods for the accuracy parameters of the composite test and the alternative test, remains the same as in the previous section. Prior distributions are also expressed similarly. The covariance parameters follow a uniform prior distribution whose limits are determined by se_{1i} and se_{2i} .

B.1 WinBUGS program for estimating bivariate hierarchical meta-analysis models of the sensitivity and specificity of a comparator test (T1) and a composite (T3) based on T1 and index test (T2) ($T3+ = T1+ \text{ or } T2+$).

This program also provides the value of the index test (T2) as a new test over the alternative test (T1), while adjusting for the covariance between T1 and T2 among disease positive. We did not model covariance among disease negatives as specificities were generally high.

Observed data must be provided for L (the number of studies), the entries of the two-by-two table of index vs. the alternative test separately in disease positive patients (T12·D) and in disease negative patients (T12·ND), the total number of disease positive (D) and disease negative (ND) in each study. Initial values should be provided by the user for covs_i, se_{1i} and se_{2i}.

```

-----
model {
for(i in 1:L) { # L is the number of studies
# Likelihood - within-studies
# 2 by 2 table between T1 and T2 in Disease+
T12·D[i,1:4] ~ dmulti(pD[i,1:4], D[i])
# 2 by 2 table between T1 and T2 in Disease-
T12·ND[i,1:4] ~ dmulti(pND[i,1:4], ND[i])
pD[i,1] <- se[1,i] * se[2,i] + covs[i] # P(t1=1, t2=1 | D=1)
pD[i,2] <- se[1,i] * (1-se[2,i]) - covs[i] # P(t1=1, t2=0 | D=1)
pD[i,3] <- (1-se[1,i]) * se[2,i] - covs[i] # P(t1=0, t2=1 | D=1)
pD[i,4] <- (1-se[1,i]) * (1-se[2,i]) + covs[i] # P(t1=0, t2=0 | D=1)
pND[i,1] <- sp[1,i] * sp[2,i] # P(t1=1, t2=1 | D=0)
pND [i,2] <- sp[1,i] * (1-sp[2,i]) # P(t1=1, t2=0 | D=0)
pND [i,3] <- (1-sp[1,i]) * sp[2,i] # P(t1=0, t2=1 | D=0)
pND [i,4] <- (1-sp[1,i]) * (1-sp[2,i]) # P(t1=0, t2=0 | D=0)
# lower limit of covariance in ith study
ll[i]<- -(1-se[1,i])*(1-se[2,i])
# upper limit of covariance in ith study
ul[i]<-min(se[1,i], se[2,i])- se[1,i]*se[2,i])
# prior over covariance in ith study
covs[i]~dunif(ll[i],ul[i])
# By definition of P(T3+|D+) = P(T1+, T2+|D+) + P(T1+, T2-|D+) + P(T1-, T2+|D+)
# and P(T3-|D-) = P(T1-, T2-|D-)
# Using this, we can express se2 and (1-sp2) in terms of the accuracies of the
# composite test and the alternative test and place priors on se2, sp2, se3 and
# sp3 as follows.
se[2,i] <- (se[3,i]-se[1,i] + covs[i]) / (1-se[1,i])
sp[2,i] <- sp[3,i] / sp[1,i]
# calculating logit sensitivities and specificities
logit(se[1, i]) <- l1[i,1]
logit(sp[1, i]) <- l1[i,2]
logit(se[3, i]) <- l3[i,1]
logit(sp[3, i]) <- l3[i,2]
# Likelihood - Between-studies
l1[i,1:2] ~ dnorm(mu1[,], T1[,])
l3[i,1:2] ~ dnorm(mu3[,], T3[,])
}
# Prior distributions
mu1[1] ~ dnorm(0,0.25)
mu1[2] ~ dnorm(0,0.25)
mu3[1] ~ dnorm(0,0.25)
mu3[2] ~ dnorm(0,0.25)
T1[1:2,1:2]<-inverse(S1[,])
T3[1:2,1:2]<-inverse(S3[,])
S1[1,1] <- sigma1[1]*sigma1[1]
S1[2,2] <- sigma1[2]*sigma1[2]
S1[1,2] <- rho1*sigma1[1]*sigma1[2]
S1[2,1] <- rho1*sigma1[1]*sigma1[2]
S3[1,1] <- sigma3[1]*sigma3[1]

```

```

S3[2,2] <- sigma3[2]*sigma3[2]
S3[1,2] <- rho3*sigma3[1]*sigma3[2]
S3[2,1] <- rho3*sigma3[1]*sigma3[2]
sigma1[1]<-pow(prec1[1],-0.5)
sigma1[2]<-pow(prec1[2],-0.5)
prec1[1] ~ dgamma(2,0.5)
prec1[2] ~ dgamma(2,0.5)
rho1 ~ dunif(-1,1)
sigma3[1]<-pow(prec3[1],-0.5)
sigma3[2]<-pow(prec3[2],-0.5)
prec3[1] ~ dgamma(2,0.5)
prec3[2] ~ dgamma(2,0.5)
rho3 ~ dunif(-1,1)
# Pooled accuracy of T1 and T3
Pooled`S1<-1/(1+exp(-mu1[1]))
Pooled`C1<-1/(1+exp(mu1[2]))
Pooled`S3<-1/(1+exp(-mu3[1]))
Pooled`C3<-1/(1+exp(mu3[2]))
# Added value of T3 compared to T1
Add`S<-Pooled`S3-Pooled`S1 # gain in sensitivity
Add`C<-Pooled`C3-Pooled`C1 # gain in specificity
# Probability that sensitivity or specificity of T3 is greater than of T1
prob`S <- step(Add`S)
prob`C <- step(Add`C)
}

```

WHAT'S NEW

Date	Event	Description
11 May 2016	Amended	Acknowledgements amended to take into account NIH funding

CONTRIBUTIONS OF AUTHORS

MS, KRS, and ND wrote the protocol, with input from CH, ZYW, SDL, and CMD. MS and CH reviewed articles for inclusion and extracted data. MS, CH, ZYW, ND, and KRS analysed the data. MS, CH, ZYW, ND, and KRS interpreted the analyses. MS, CH, and KRS drafted the manuscript. ND drafted the statistical analysis section, and ZYW and ND drafted the statistical appendix ([Appendix 5](#)). ND, SDL, and CMD provided critical revisions to the manuscript. All review authors read and approved the final manuscript draft.

DECLARATIONS OF INTEREST

Development of the systematic review was in part made possible with financial support from the United States Agency for International Development (USAID) administered by the World Health Organization (WHO) Global TB Programme, Switzerland. KRS, MS, CH, and ZYW received funding to carry out the review from USAID. KRS served as Co-ordinator of the Evidence Synthesis and Policy Subgroup of Stop TB Partnership's New Diagnostics Working Group through 2015. SDL has received Wellcome Trust grant funds to conduct primary research evaluating diagnostic tests using this product. SDL published primary research that was included in the review and provided intellectual input to the review, but did not participate directly in the application of the inclusion criteria, 'Risk of bias' assessment, or data extraction. CMD is employed by FIND, a Swiss non-profit organization. FIND provided funding for an initial assessment of data available for the review. KRS, MS, and ZYW received funding from FIND for an initial assessment. We presented the findings from the initial assessment at the Union World Conference on Lung Health, Barcelona, October 2014. MS is supported through an National Institute of Health (NIH) K23 grant (K23AI089259). ND has no declarations of interest to declare. The review authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the review apart from those disclosed.

SOURCES OF SUPPORT

Internal sources

- Liverpool School of Tropical Medicine, UK.
- Cochrane Infectious Diseases Group, UK.

External sources

- Department for International Development, UK.

Grant: 5242

- United States Agency for International Development (USAID) administered by the World Health Organization (WHO) Global TB Programme, Switzerland.

Development of the systematic review was in part made possible with financial support from the USAID administered by the WHO Global TB Programme

- National Institute of Health, USA.

This work was supported by NIH K23 grant AI089259

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Our review differed from the Cochrane protocol, [Shah 2014](#), in several ways. In [Shah 2014](#) we wrote as the objective that we would assess the accuracy of LF-LAM at different cut-off values for test positivity (grade 1 and grade 2). Over the year we worked on the review, the test manufacturer revised the reference bands such that the band intensity for grade 1 on the new reference card (on a scale from 1 to 4) corresponded to the band intensity for grade 2 from the previous reference card (on a scale of 1 to 5; this scale was used to define grade in this review since it represents the grading system used in included studies). Therefore, we wanted the review to highlight LF-LAM grade 2 (on the prior scale) which corresponds to the current manufacturer recommendations. As such, we modified the review as follows: we removed the words "at different cut-off values for test positivity" from the objective. We presented results for both grade 1 and 2 for diagnosis in [Table 1](#) (there were few results for TB screening which we presented in the text). However, in the 'Summary of findings' tables, we only presented results for grade 2. In the protocol we wrote that we would exclude data reported only in abstracts. However, we decided to include abstracts in the review when they included sufficient data. We corresponded with all authors of included studies, including authors of abstracts, for missing data and clarifications. In the protocol we described two reference standards, a microbiological and a broad reference standard. We re-titled the latter "composite reference standard." In addition, as suggested by our advisory group, we amended the definition of the composite reference standard to include microscopy smear as one of the components, though, in actuality, the included studies rarely used smear. In the protocol, we wrote that we might consider an alternative approach in which we obtained pooled estimates of sensitivity and specificity via a two-part meta-analysis model. We did not do this. In the protocol we used the words "incremental value" to describe LF-LAM as a test combined with sputum smear microscopy

or Xpert® MTB/RIF. In the review, we removed the word 'incremental' because we thought it was confusing. Otherwise the methods in the protocol, [Shah 2014](#), and this review were similar.

INDEX TERMS

Medical Subject Headings (MeSH)

Biomarkers [urine]; CD4 Lymphocyte Count; HIV Seropositivity [*complications]; Lipopolysaccharides [*urine]; Point-of-Care Systems; Randomized Controlled Trials as Topic; Sensitivity and Specificity; Tuberculosis [*diagnosis]; Tuberculosis, Pulmonary [diagnosis]

MeSH check words

Adult; Humans