![Cochrane Library logo]

Cochrane Database of Systematic Reviews

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)

Bjerrum S, Schiller I, Dendukuri N, Kohli M, Nathavitharana RR, Zwerling AA, Denkinger CM, Steingart KR, Shah M

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV

Stephanie Bjerrum[1,2,3], Ian Schiller[4], Nandini Dendukuri[4], Mikashmi Kohli[5], Ruvandhi R Nathavitharana[6], Alice A Zwerling[7], Claudia M Denkinger[8,9], Karen R Steingart[10], Maunank Shah[11]

[1]Department of Clinical Research, Research Unit of Infectious Diseases, University of Southern Denmark, Odense, Denmark. [2]MyCRESD, Mycobacterial Research Centre of Southern Denmark, Department of Infectious Diseases, Odense University Hospital, Odense, Denmark. [3]OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark. [4]Division of Clinical Epidemiology, McGill University Health Centre - Research Institute, Montreal, Canada. [5]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada. [6]Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA. [7]School of Epidemiology & Public Health, University of Ottawa, Ottawa, Canada. [8]FIND, Geneva, Switzerland. [9]Center of Infectious Diseases, University Hospital Heidelberg, Heidelberg, Germany. [10]Honorary Research Fellow, Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK. [11]Department of Medicine, Division of Infectious Diseases, John Hopkins University School of Medicine, Baltimore, Maryland, USA

Contact address: Stephanie Bjerrum, Department of Clinical Research, Research Unit of Infectious Diseases, University of Southern Denmark, Odense, Denmark. steph@medicinsk.dk.

A B S T R A C T

Background

The lateral flow urine lipoarabinomannan (LF-LAM) assay Alere Determine™ TB LAM Ag is recommended by the World Health Organization (WHO) to help detect active tuberculosis in HIV-positive people with severe HIV disease. This review update asks the question, "does new evidence justify the use of LF-LAM in a broader group of people?", and is part of the WHO process for updating guidance on the use of LF-LAM.

Objectives

To assess the accuracy of LF-LAM for the diagnosis of active tuberculosis among HIV-positive adults with signs and symptoms of tuberculosis (symptomatic participants) and among HIV-positive adults irrespective of signs and symptoms of tuberculosis (unselected participants not assessed for tuberculosis signs and symptoms).

The proposed role for LF-LAM is as an add on to clinical judgement and with other tests to assist in diagnosing tuberculosis.

Search methods

We searched the Cochrane Infectious Diseases Group Specialized Register; MEDLINE, Embase, Science Citation Index, Web of Science, Latin American Caribbean Health Sciences Literature, Scopus, the WHO International Clinical Trials Registry Platform, the International Standard Randomized Controlled Trial Number Registry, and ProQuest, without language restriction to 11 May 2018.

Selection criteria

Randomized trials, cross-sectional, and observational cohort studies that evaluated LF-LAM for active tuberculosis (pulmonary and extrapulmonary) in HIV-positive adults. We included studies that used the manufacturer's recommended threshold for test positivity, either the updated reference card with four bands (grade 1 of 4) or the corresponding prior reference card grade with five bands (grade 2 of 5). The reference standard was culture or nucleic acid amplification test from any body site (microbiological). We considered a higher quality reference standard to be one in which two or more specimen types were evaluated for tuberculosis diagnosis and a lower quality reference standard to be one in which only one specimen type was evaluated.

Data collection and analysis

Two review authors independently extracted data using a standardized form and REDCap electronic data capture tools. We appraised the quality of studies using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool and performed meta-analyses to estimate pooled sensitivity and specificity using a bivariate random-effects model and a Bayesian approach. We analyzed studies enrolling strictly symptomatic participants separately from those enrolling unselected participants. We investigated pre-defined sources of heterogeneity including the influence of CD4 count and clinical setting on the accuracy estimates. We assessed the certainty of the evidence using the GRADE approach.

Main results

We included 15 unique studies (nine new studies and six studies from the original review that met the inclusion criteria): eight studies among symptomatic adults and seven studies among unselected adults. All studies were conducted in low- or middle-income countries. Risk of bias was high in the patient selection and reference standard domains, mainly because studies excluded participants unable to produce sputum and used a lower quality reference standard.

Participants with tuberculosis symptoms

LF-LAM pooled sensitivity (95% credible interval (CrI)) was 42% (31% to 55%) (moderate-certainty evidence) and pooled specificity was 91% (85% to 95%) (very low-certainty evidence), (8 studies, 3449 participants, 37% with tuberculosis).

For a population of 1000 people where 300 have microbiologically-confirmed tuberculosis, the utilization of LF-LAM would result in: 189 to be LF-LAM positive: of these, 63 (33%) would not have tuberculosis (false-positives); and 811 to be LF-LAM negative: of these, 174 (21%) would have tuberculosis (false-negatives).

By clinical setting, pooled sensitivity was 52% (40% to 64%) among inpatients versus 29% (17% to 47%) among outpatients; and pooled specificity was 87% (78% to 93%) among inpatients versus 96% (91% to 99%) among outpatients. Stratified by CD4 cell count, pooled sensitivity increased, and specificity decreased with lower CD4 cell count.

Unselected participants not assessed for signs and symptoms of tuberculosis

LF-LAM pooled sensitivity was 35% (22% to 50%), (moderate-certainty evidence) and pooled specificity was 95% (89% to 96%), (low-certainty evidence), (7 studies, 3365 participants, 13% with tuberculosis).

For a population of 1000 people where 100 have microbiologically-confirmed tuberculosis, the utilization of LF-LAM would result in: 80 to be LF-LAM positive: of these, 45 (56%) would not have tuberculosis (false-positives); and 920 to be LF-LAM negative: of these, 65 (7%) would have tuberculosis (false-negatives).

By clinical setting, pooled sensitivity was 62% (41% to 83%) among inpatients versus 31% (18% to 47%) among outpatients; pooled specificity was 84% (48% to 96%) among inpatients versus 95% (87% to 99%) among outpatients. Stratified by CD4 cell count, pooled sensitivity increased, and specificity decreased with lower CD4 cell count.

Authors' conclusions

We found that LF-LAM has a sensitivity of 42% to diagnose tuberculosis in HIV-positive individuals with tuberculosis symptoms and 35% in HIV-positive individuals not assessed for tuberculosis symptoms, consistent with findings reported previously. Regardless of how people are enrolled, sensitivity is higher in inpatients and those with lower CD4 cell, but a concomitant lower specificity. As a simple point-of-care test that does not depend upon sputum evaluation, LF-LAM may assist with the diagnosis of tuberculosis, particularly when a sputum specimen cannot be produced.

17 October 2019

Up to date

All studies incorporated from most recent search

All studies identified during the most recent search (11 May, 2018) have been incorporated in the review, and no ongoing studies identified.

PLAIN LANGUAGE SUMMARY

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV

Why is improving the diagnosis of tuberculosis important?

Tuberculosis causes more deaths in people living with HIV than any other disease. The lateral flow urine lipoarabinomannan assay (LF-LAM, Alere Determine™ TB LAM Ag assay) is a World Health Organization-recommended rapid test to assist in detection of active tuberculosis in HIV-positive people with severe HIV disease. Rapid and early tuberculosis diagnosis may allow for prompt treatment and alleviate severe illness and death. An incorrect tuberculosis diagnosis may result in anxiety and unnecessary treatment.

What is the aim of this review?

To find out how accurate LF-LAM is for diagnosing tuberculosis in HIV-positive people with tuberculosis symptoms (symptomatic participants) and those not assessed for tuberculosis symptoms (unselected participants). This is an update of the 2016 Cochrane Review.

What was studied in this review?

LF-LAM is a commercially available point-of-care test that detects lipoarabinomannan (LAM), a component of the bacterial cell walls, present in some people with active tuberculosis. The test is simple and shows results in 25 minutes. LF-LAM results were measured against culture or molecular tests (benchmark).

What are the main results of this review?

Fifteen studies: eight studies evaluated LF-LAM for tuberculosis among symptomatic participants and seven studies among unselected participants. All studies were conducted in low- or middle-income countries.

Tuberculosis diagnosis among symptomatic participants: LF-LAM registered positive in 42% (sensitivity) of people who actually had tuberculosis and did not register positive in 91% of people who were actually negative (specificity).

Tuberculosis diagnosis among unselected participants: LF-LAM sensitivity was 35% and specificity 95%.

How confident are we in the review's results?

Several studies excluded participants who could not produce sputum and most studies relied on a lower quality benchmark. Few studies and participants were included in some analyses and only one study was conducted outside of sub-Saharan Africa. Results should be interpreted with caution.

What do the results mean?

Among symptomatic participants, in theory, for a population of 1000 people where 300 have microbiologically-confirmed tuberculosis, the utilization of LF-LAM would result in: 189 to be LF-LAM positive: of these, 63 (33%) would not have tuberculosis (false-positives); and 811 to be LF-LAM negative: of these, 174 (21%) would have tuberculosis (false-negatives).

Among unselected participants, in theory, for a population of 1000 people where 100 have microbiologically-confirmed tuberculosis, the utilization of LF-LAM would result in: 80 to be LF-LAM positive: of these, 45 (56%) would not have tuberculosis (false-positives); and 920 to be LF-LAM negative: of these, 65 (7%) would have tuberculosis (false-negatives).

Who do the review's results apply to?

HIV-positive people with tuberculosis symptoms and those not assessed for tuberculosis symptoms.

What are the implications of this review?

LF-LAM has sensitivity around 40% to detect tuberculosis. As the test does not require sputum collection, LF-LAM may be the only way to diagnose tuberculosis when sputum cannot be produced.

How up-to-date is this review?

To 11 May 2018.

SUMMARY OF FINDINGS

Summary of findings 1.   LF-LAM for symptomatic participants

Review question: what is the diagnostic accuracy of LF-LAM for the diagnosis of active tuberculosis in HIV-positive adults with signs and symptoms of tuberculosis?

Studies: cross-sectional studies and randomized controlled trials

Participants: HIV-positive adults with tuberculosis signs and symptoms

Setting: all settings (inpatient and outpatient)

Index test: LF-LAM

Threshold for index tests: manufacturer's recommended threshold for positivity i.e. grade 1 of 4 (revised reference card) or the corresponding grade 2 of 5 (prior reference card)

Reference standard: microbiological (mycobacterial culture and/or nucleic acid amplification test)

Role: an add on to clinical judgement and with other tests to assist in tuberculosis diagnosis

Pooled sensitivity (95%CrI): 42% (31% to 55%); **pooled specificity (95% CrI): 91% (85% to 95%)**

Test result	Number of results per 1000 participants tested (95% CrI)			Number of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1% **Typically seen in asymptomatic persons in outpatient settings**	**Prevalence 10%** **Typically seen in symptomatic persons in all settings**	**Prevalence 30%** **Typically seen in seriously ill persons in inpatient settings**		
True positives	**4** (3 to 6)	**42** (31 to 55)	**126** (93 to 165)	1277 (8)	⊕⊕⊕⊖ **MODERATE**a,b,c,d
False negatives	**6** (4 to 7)	**58** (45 to 69)	**174** (135 to 207)		
True negatives	**901** (842 to 941)	**819** (765 to 855)	**637** (595 to 665)	2172 (8)	⊕⊕⊖⊖ **VERY LOW**c,e,f
False positives	**89** (49 to 148)	**81** (45 to 135)	**63** (35 to 105)		

Abbreviations: CrI: credible interval.

Explanations

aAs assessed by QUADAS-2, in the patient selection domain, we judged six studies (75%) at high risk of bias because they did not avoid inappropriate exclusions. We downgraded one level for risk of bias.

bFor individual studies, sensitivity estimates ranged from 23% to 68%. We thought that differences in enrolment criteria (different populations targeted) or CD4 count could explain in part the heterogeneity. We did not downgrade for inconsistency.

[c]The median tuberculosis prevalence in the studies was 42% and thus the results tend to be more applicable to settings with a higher tuberculosis prevalence. For tuberculosis prevalence of 1% and 10%, whether or not to downgrade remains unclear. It is possible the test will perform differently at lower prevalences. We did not downgrade for indirectness.

[d]We thought the 95% CrIs around true positives and false negatives would likely not lead to different decisions depending on which credible limits are assumed. We did not downgrade for imprecision.

[e]As assessed by QUADAS-2, in the reference standard domain, we judged seven studies (88%) at high risk of bias because we thought the reference standard used was unlikely to correctly classify the target condition. We downgraded two levels for risk of bias.

[f]We thought the 95% CrIs around true negatives and false positives would likely lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.

GRADE certainty of evidence (GRADEpro 2015; Balshem 2011)
High: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate: 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: our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very low: 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 normalized frequencies within a hypothetical cohort of 1000 patients at three different prevalences of tuberculosis (pre-test probabilities): 1%, 10% and 30%. Credible limits (CrIs) 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.

Summary of findings 2.   LF-LAM for unselected participants

Review question: what is the diagnostic accuracy of LF-LAM for the diagnosis of active tuberculosis in HIV-positive adults irrespective of signs and symptoms of tuberculosis?

Studies: cross-sectional studies and randomized controlled trials

Participants: HIV-positive unselected adults not assessed for signs and symptoms of tuberculosis i.e. irrespective of symptoms

Setting: all settings (inpatient and outpatient)

Index test: LF-LAM

Threshold for index tests: manufacturer's recommended threshold for positivity i.e. grade 1 of 4 (revised reference card) or the corresponding grade 2 of 5 (prior reference card)

Reference standard: microbiological (mycobacterial culture and/or nucleic acid amplification test)

Role: an add on to clinical judgement and with other tests to assist in tuberculosis diagnosis

Pooled sensitivity (95% CrI): 35% (22% to 50%); pooled specificity (95% CrI): (89% to 98%)

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)

Test result	Number of results per 1000 participants tested (95% CrI)			Number of partici-pants (studies)	Certainty of the evidence (GRADE)
	Prevalence 1%	**Prevalence 10%**	**Prevalence 30%**		
	Typically seen in asympto-matic persons in outpatient settings	**Typically seen in sympto-matic persons in all settings**	**Typically seen in seriously ill persons in inpatient settings**		
True positives	**3** (2 to 5)	**35** (22 to 50)	**105** (66 to 150)	432 (7)	⊕⊕⊕⊖ **MODERATE**[a,b,c]
False negatives	**7** (5 to 8)	**65** (50 to 78)	**195** (150 to 234)		
True negatives	**941** (881 to 970)	**855** (801 to 882)	**665** (623 to 686)	2933 (7)	⊕⊕⊖⊖ **LOW**[d,e,f]
False positives	**49** (20 to 109)	**45** (18 to 99)	**35** (14 to 77)		

Abbreviations: CrI: Credible interval.

Explanations

[a]As assessed by QUADAS-2, in the patient selection domain, we judged four studies (57%) at high risk of bias because they did not avoid inappropriate exclusions. We downgraded one level for risk of bias.

[b]For individual studies, sensitivity ranged from 0% to 67%. We thought that the percentage of patients with tuberculosis symptoms or differences in CD4 count could explain in part the heterogeneity. One study, LaCourse 2016, with sensitivity of 0% differed from the other studies by including a) a population of exclusively pregnant women attending an antenatal care setting, b) a low proportion of symptomatic participants (19%), c) a low tuberculosis prevalence (1%), and d) a high median CD4 cell count (437 cells per μL). One study, Thit 2017, with sensitivity 67% differed from the other studies by being conducted in Myanmar, and was the only study included in this review that evaluated LF-LAM in a setting outside sub-Saharan Africa. We did not downgrade for inconsistency.

[c]We thought the 95% CrI around true positives and false negatives would likely not lead to different decisions depending on which credible limits are assumed. We did not downgrade for imprecision.

[d]As assessed by QUADAS-2, in the reference standard domain, we judged five studies (71%) at high risk of bias because we thought the reference standard used was unlikely to correctly classify the target condition. We downgraded one level for risk of bias.

[e]For individual studies, specificity ranged from 67% to 99%. Eighty-six per cent (6/7) of the included studies had specificity of 94% or higher. One study, Thit 2017, with specificity 67% differed from the other studies by being conducted in Myanmar, and is the only study included in this review that evaluated LF-LAM in a setting outside sub-Saharan Africa. We did not downgrade for inconsistency.

[f]We thought the wide 95% CrIs around true negatives and false positives would lead to different decisions depending on which credible limits are assumed. We downgraded one level for imprecision.

GRADE certainty of evidence (GRADEpro 2015; Balshem 2011)
High: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate: 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: our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very low: 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 normalized frequencies within a hypothetical cohort of 1000 patients at three different prevalences of tuberculosis (pre-test probabilities): 1%, 10% and 30%. Credible limits (CrIs) 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.

BACKGROUND

Tuberculosis is an infectious airborne disease caused by the bacillus *Mycobacterium tuberculosis* (*M tuberculosis*). In 2017, an estimated 10.0 million people developed tuberculosis disease of which 920,000 (9%) occurred among people living with HIV (WHO Global Tuberculosis Report 2018). Although the number of tuberculosis deaths has fallen since 2000, tuberculosis was responsible for 1.6 million deaths in 2017 (WHO Global Tuberculosis Report 2018), and tuberculosis has surpassed HIV as the world's leading cause of death from an infectious disease. Among people living with HIV, tuberculosis is the major cause of hospitalisation and in-hospital death despite increased access to antiretroviral treatment (ART) (Ford 2016). A systematic review of the prevalence of tuberculosis identified at autopsy in resource-limited settings, suggests that tuberculosis is responsible for up to 40% of all HIV-related deaths and that tuberculosis often is disseminated and undiagnosed at the time of death (Gupta 2015). Globally in 2017, only 51% of the incident tuberculosis cases were reported among people living with HIV (WHO Global Tuberculosis Report 2018). However, most deaths from tuberculosis are preventable if the disease is detected early and effectively treated. Overall, it is estimated that 45 million lives were saved between 2000 and 2017 through effective diagnosis and treatment (WHO Global Tuberculosis Report 2018).

Geographically, HIV and tuberculosis are often concentrated in areas of poverty with limited resources for diagnosis, treatment, and prevention of tuberculosis. Most of the 30 high tuberculosis/HIV burden countries are situated in sub-Saharan Africa where HIV-infection represents a major driver of the tuberculosis epidemic (WHO Global Tuberculosis Report 2018). HIV-positive individuals have a 20- to 37-fold increased risk of developing tuberculosis compared to HIV-negative individuals (Getahun 2010). The risk of tuberculosis increases with decreasing CD4 cell count, and remains elevated throughout the course of HIV (Gupta 2012; Lawn 2009).

Tuberculosis predominantly affects the lungs (pulmonary tuberculosis), but can affect most parts of the body, such as the lymph nodes, pleura, brain, or spine (extrapulmonary tuberculosis). Extrapulmonary tuberculosis represented 14% of the incident tuberculosis cases notified in 2017 (WHO Global Tuberculosis Report 2018). In comparison with HIV-negative people, HIV-positive people have higher rates of extrapulmonary tuberculosis or mycobacteraemia (tuberculosis bloodstream infection) (Pai 2016; Shivakoti 2017). Signs and symptoms of tuberculosis in people living with HIV vary depending on the progression of immunodeficiency and involved site(s). Complaints are often non-specific such as fever, weight loss, and fatigue. A cough for longer than two weeks is a common distinguishing feature of tuberculosis that prompts diagnostic testing for tuberculosis in people who are HIV negative, but is present in less than a third of people with tuberculosis who are HIV positive (Cain 2010). Similarly, radiographic features of tuberculosis in people living with HIV may be misleading or atypical. Whereas upper lobe cavitary lesions are often seen in HIV-negative people with tuberculosis, such lesions are less common in HIV-positive people with tuberculosis (Cain 2010). Identifying those who warrant further testing for tuberculosis can therefore be challenging. Tuberculosis diagnosis further relies heavily on a sputum-based test strategy that may have limited accuracy as well as applicability among a population that may be sputum scarce and frequently has

extrapulmonary and disseminated tuberculosis disease. Moreover, people living with HIV often have paucibacillary tuberculosis disease that makes diagnosis of tuberculosis by smear microscopy, nucleic acid amplification tests (NAATs), and culture less sensitive.

Non-sputum-based point-of-care tuberculosis diagnostic tests are highly desired to improve tuberculosis case detection in people living with HIV and ensure timely treatment (WHO TTP 2014). Desired characteristics of such a test would include minimal or non-invasive sample collection, short time to result (under one hour), and ability to implement the test without need for special instruments, electricity, or specimen preparation (WHO TTP 2014).

Detection of mycobacterial antigen in urine has attracted great attention over time. Urine-based antigen testing would allow for a tuberculosis diagnosis that is non-site specific. In addition, urine is easy to collect and store, and lacks the infection control risks associated with sputum collection. Multiple platforms have been developed to detect lipoarabinomannan (LAM), a component of mycobacterial cell walls, initially as enzyme-linked immunosorbent (ELISA) assays that were evaluated in several clinical settings (Minion 2011). Later, the lateral flow urine lipoarabinomannan (LF-LAM, Alere Determine TB-LAM assay) was developed as a simple point-of-care test for diagnosis of active tuberculosis in people living with HIV. The test is commercially available, does not require access to special laboratory equipment, and produces a result after 25 minutes (Alere 2017), meeting many of the desired target product-profile requirements (WHO TTP 2014).

The original Cochrane Review of LF-LAM for tuberculosis included 12 studies (Shah 2016). In that review, six studies evaluated accuracy among symptomatic individuals and found a low pooled sensitivity of 45% (Credible Interval (CrI) 29% to 63%) and specificity of 92% (CrI 80% to 97%) against a microbiological reference standard (Shah 2016). In participants with CD4 ≤ 100 cells/µL, pooled sensitivity increased to 56% (CrI 41% to 70%) and specificity decreased to 90% (81% to 95%). In the original review, the accuracy of LF-LAM was also evaluated in participants not assessed of symptoms (i.e. termed 'TB screening' and now described as studies among 'unselected participants') with sensitivity ranging from 0% to 44% (Shah 2016).

In 2015, informed by the original Cochrane Review, the WHO made a conditional recommendation for using LF-LAM to assist with the diagnosis of tuberculosis in HIV-positive people with advanced disease (described below) and a strong recommendation against using the test "as a screening test for tuberculosis" based on the data among unselected participants (WHO Lipoarabinomannan Policy Guidance 2015).

Since 2015, additional evidence for the use and clinical impact of LF-LAM has emerged. This Cochrane Review update includes published studies evaluating the accuracy of the commercially available LF-LAM, Alere Determine TB LAM Ag assay for diagnosis of active tuberculosis (pulmonary and extrapulmonary tuberculosis) in people living with HIV and informed the updated WHO guidelines on the use of the test. Of note, in 2018, preliminary performance characteristics of a second commercially developed lateral flow assay to detect LAM for the diagnosis of tuberculosis was announced based on data from frozen biobank specimens (Fujifilm SILVAMP TB LAM, Japan; FujiLAM) (Broger 2019). The test is projected to become commercially available in 2020; no studies of FujiLAM accuracy are included in this updated review.

Target condition being diagnosed

The target condition is active tuberculosis, which includes pulmonary and extrapulmonary tuberculosis.

Index test(s)

The urine-based lateral flow lipoarabinomannan immunocapture assay (LF-LAM) is a commercially available point-of-care test for active tuberculosis (Alere Determine™ TB LAM Ag, Abbott, Palatine, IL, USA, previous Alere Inc., Waltham, MA, USA). Lipoarabinomannan (LAM) is a lipopolysaccharide present in mycobacterial cell walls (Brennan 2003), which is released from metabolically active or degenerating bacterial cells during tuberculosis disease (Briken 2004). LAM is detectable in urine of people with active tuberculosis disease and evaluated for both LAM ELISA and the LF-LAM testing platforms (Lawn 2012; Minion 2011; Peter 2010; Shah 2016). The original Cochrane Review of LF-LAM (Shah 2016) and a meta-analysis of an earlier generation LAM ELISA test (Minion 2011) both demonstrated that the accuracy of urinary LAM detection was improved among people living with HIV with advanced immunosuppression. Several hypotheses may explain the higher sensitivity of urine LAM detection in people living with HIV including higher bacillary burden and antigen load (Shah 2010), greater likelihood of genitourinary tract tuberculosis involvement, and greater glomerular permeability to allow increased antigen levels in urine (Lawn 2016; Minion 2011).

LF-LAM testing is performed manually by applying 60 μL of unprocessed urine to the sample pad of the DetermineTB LAM Ag test and leave the strip test to incubate at room temperature for 25 minutes (Alere 2017; Appendix 1). The strip is then inspected by eye. The intensity of any visible band is graded by comparing it with the intensities of the bands on a manufacturer-supplied reference scale card. Of note, the reference scale was revised in January 2014. Prior to January 2014, the reference scale card included five bands (grade 1 representing a very low intensity band to grade 5 representing a high/dark intensity band). After January 2014, the manufacturer revised the reference scale 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 (Appendix 2). Under the current manufacturer recommendations (using the revised four bands reference card), only bands that are grade 1 or higher are considered positive (Alere 2017; Appendix 2).

Clinical pathway

Based on current WHO guidelines (WHO Lipoarabinomannan Policy Guidance 2015), the proposed role for the LF-LAM test is as an 'add on' to clinical judgement and with other tests to assist in tuberculosis diagnosis. The test does not have a role as a replacement or triage test. The reason for specifying 'clinical judgement' is that, in our view, clinical judgement for HIV-associated tuberculosis carries more importance than in many other cases (i.e. tuberculosis treatment may be provided regardless of the test results). The reason for specifying 'add on' is that for settings where the test is not available, the same clinical judgement is used; hence for settings where the test is available, it is used 'in addition'. Importantly, some HIV-positive people may have difficulties in producing a good quality sputum specimen or any sputum at all.

In the WHO policy guidance on the use of LAM, it is recommended that LF-LAM "may be used to assist in the diagnosis of tuberculosis in HIV-positive adult inpatients with signs and symptoms of tuberculosis (pulmonary/and/or extrapulmonary) and a CD4 cell count less than or equal to 100 cells/μL, or in people living with HIV who are 'seriously ill' regardless of CD4 count or if the CD4 count is unknown" (WHO Lipoarabinomannan Policy Guidance 2015). The recommendations also apply to HIV-positive outpatients and children with signs and symptoms of tuberculosis (pulmonary and/or extrapulmonary) based on the generalization of data from adult inpatients while acknowledging the limitation of available data (WHO Lipoarabinomannan Policy Guidance 2015). The WHO recommends that LF-LAM should not be used for general tuberculosis screening "owing to suboptimal sensitivity" (WHO Lipoarabinomannan Policy Guidance 2015).

In 2016, WHO developed an algorithm for managing people living with HIV presumed to have tuberculosis that was published as part of the guidelines on use of ART for treating and preventing HIV infection (WHO ART Guidelines 2016). The algorithm recommends the Xpert® MTB/RIF assay (Cepheid, Sunnyvale, USA) as the initial diagnostic test in adults and children with presumed HIV-associated tuberculosis (WHO Xpert® Policy Update 2013). It is recommended that Xpert® MTB/RIF should be used rather than conventional microscopy and culture for testing sputum and specific extra-pulmonary specimens like cerebrospinal fluid and lymph nodes (WHO Xpert® Policy Update 2013). The algorithm includes recommendations on use of LF-LAM as a test that may assist in diagnosing active tuberculosis among seriously ill adults and children living with HIV, regardless of CD4 count (WHO ART Guidelines 2016). Use of LF-LAM is also included as one of the components in the package for diagnosis of tuberculosis in the WHO guidelines for managing people presenting with advanced HIV (WHO Managing Advanced HIV Disease 2017).

When extrapulmonary tuberculosis is suspected, it is recommended to obtain "appropriate specimens from the suspected sites of involvement for microscopy, culture, and histopathological examination" (TB CARE I 2014). A newly published Cochrane Review found that Xpert MTB/RIF may be helpful in confirming the diagnosis of extrapulmonary tuberculosis (Kohli 2018). However, evaluation for extrapulmonary tuberculosis often requires invasive diagnostic procedures that may have low yield even in people with advanced disease.

To identify those who need referral for tuberculosis diagnostic testing, 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 tuberculosis and other diseases" referred to as the WHO symptom screening rule (Getahun 2011; WHO ICF 2011). People living with HIV with tuberculosis may however not exhibit common symptoms of tuberculosis disease. In some settings with high prevalence of tuberculosis, clinicians have therefore considered evaluating all patients presenting for care for tuberculosis without assessment or consideration of specific signs or symptoms of tuberculosis.

In the context of LF-LAM being recommended in combination with existing tuberculosis tests to increase early tuberculosis diagnosis and treatment, the downstream consequences of LF-LAM testing include the following.

- True-positive (TP): people would benefit from rapid non-site specific diagnosis and early initiation of tuberculosis treatment.
- True-negative (TN): people would be spared unnecessary treatment and would benefit from reassurance and pursuit of an alternative diagnosis.
- False-positive (FP): people would likely experience anxiety and morbidity caused by additional testing, unnecessary treatment, and possible adverse effects; possible stigma associated with a tuberculosis diagnosis; and a FP result may halt further diagnostic evaluation. However, as FP results increase (specificity decrease) as CD4 count decrease, the observed FP results may be due to an inability of the sickest patients to produce a sputum specimen and are TP results being misclassified as FP. Given the high mortality in persons living with HIV, acting on all positive LF-LAM results may balance the possibly adverse effects associated with unnecessary diagnosis and treatment.
- False-negative (FN): people would experience increased risk of morbidity and mortality and delayed treatment initiation; there would be continuous risk of tuberculosis transmission. These concerns may, however, be limited when using LF-LAM in combination with existing tuberculosis tests.

Importantly, it should be noted that LF-LAM does not provide information about drug resistance and some individuals with unidentified drug-resistant tuberculosis may be inappropriately treated with a regimen appropriate for drug-susceptible tuberculosis.

Alternative test(s)

In this section, we briefly describe selected alternative tests for detection of tuberculosis. For a comprehensive review of these tests, we refer the reader to several excellent resources (Lewinsohn 2017; Unitaid 2017).

Smear microscopy (light microscopy (Ziehl-Neelsen), fluorescence microscopy, or light-emitting diode (LED) fluorescence microscopy) is the examination of smears for acid-fast bacilli (tuberculosis bacteria) under a microscope. Advantages of smear microscopy include its simplicity, low cost, speed, and high specificity in high tuberculosis burden areas. In addition, smear microscopy identifies the most infectious tuberculosis patients. Smear microscopy can be performed in basic laboratories. Drawbacks of smear microscopy include the need for specialized training and its relatively low sensitivity, 50% to 60% on average for a direct smear. Among people living with HIV sensitivity is reported as low as 22% to 43% (Getahun 2007). Around 5000 to 10,000 organisms per mL must be present in the specimen for tuberculosis bacteria to be visible by microscopy (Lewinsohn 2017). Although, the sensitivity of microscopy can be improved by approximately 10% with fluorescence (Steingart 2006), a large number of tuberculosis cases still go undiagnosed if tuberculosis diagnosis solely relies on sputum smear microscopy. Smear-negative tuberculosis is disproportionately more common in HIV-positive than HIV-negative individuals, accounting for 24% to 61% of all pulmonary cases in people living with HIV, the yield decreasing with lower CD4 cell counts (Getahun 2007; Perkins 2007).

Mycobacterial culture is a method used to grow bacteria on nutrient-rich media. In comparison with microscopy, a positive culture requires only around 100 organisms per mL and therefore can detect lower numbers of tuberculosis bacteria (Lewinsohn 2017). Additionally, culture is essential for species identification and drug susceptibility testing. However, culture may take up to six to eight weeks and requires a highly equipped laboratory.

Nucleic acid amplification tests (NAATs) are molecular systems that can detect small quantities of genetic material (DNA or RNA) from microorganisms, such as *M tuberculosis*. The key advantage of NAATs is that they are rapid diagnostic tests, potentially providing results in a few hours. A variety of molecular amplification methods are available, of which polymerase chain reaction (PCR) is the most common. NAATs are available as commercial kits and in-house tests (based on a protocol developed in a laboratory) and are used routinely in high-income countries for tuberculosis detection. In-house PCR is widely used in low-income countries because these tests are less expensive than commercial kits. However, in-house PCR is known to produce highly inconsistent results (Flores 2005).

Xpert MTB/RIF and Xpert Ultra, the newest version (Cepheid, Sunnyvale, USA), are fully automated NAATs, that simultaneously and rapidly detect *M tuberculosis* complex and rifampicin resistance (WHO Xpert® Policy Update 2013; WHO Xpert Ultra 2017). A Cochrane Review found that Xpert MTB/RIF was sensitive and specific for both pulmonary tuberculosis detection and rifampicin resistance detection (Horne 2019). Compared with Xpert MTB/RIF, Xpert Ultra had higher sensitivity and lower specificity for tuberculosis and similar sensitivity and specificity for rifampicin resistance (one study) (Horne 2019). Although sputum testing with Xpert MTB/RIF has high sensitivity for smear-positive pulmonary tuberculosis (98%), sensitivity is lower for smear-negative pulmonary tuberculosis (67%) (Horne 2019). In another Cochrane Review including 66 studies evaluating Xpert MTB/RIF for detection of extrapulmonary tuberculosis, pooled sensitivity was found to vary across different types of specimens from 31% in pleural tissue to 97% in bone or joint fluid, whereas specificity varied less from 82% in bone or joint tissue to 99% in pleural fluid and urine (Kohli 2018). These finding have bearing for HIV-associated tuberculosis where smear-negative tuberculosis and extrapulmonary tuberculosis are disproportionately higher. In 2017, based on a non-inferiority analysis of Xpert Ultra compared with Xpert MTB/RIF, the WHO stated that recommendations on the use of Xpert MTB/RIF also apply to the use of Xpert Ultra as the initial diagnostic test for all adults and children with signs and symptoms of tuberculosis (WHO Xpert Ultra 2017).

The loop-mediated isothermal amplification test, TB-LAMP (Eiken Chemical Co., Tokyo, Japan), has been recommended by the WHO since 2016, for diagnosing pulmonary tuberculosis in adults (WHO TB-LAMP 2016). It is a manual and simple assay that can be performed directly on sputum samples with a result provided in less than one hour (Yuan 2014). It is suitable for use at peripheral health centres and is promoted as a test more sensitive than microscopy, but inferior to Xpert MTB/RIF (WHO TB-LAMP 2016).

FujiLAM is a novel, lateral flow urine-based, point-of-care test for tuberculosis diagnosis in people living with HIV in low-resource settings. This assay was developed to detect LAM with results available in less than one hour and was announced on 26 September 2018, the day of the first United Nations General Assembly high-level meeting on tuberculosis (FIND 2018). Fuji LAM has been evaluated on frozen biobank samples originating from various studies of hospitalized HIV-positive patients with an estimated sensitivity of 70% (Broger 2019). Studies are ongoing

to evaluate FujiLAM performance on bio-samples stored from outpatient HIV-cohorts and prospective studies to evaluate the diagnostic accuracy of FujiLAM are now called for as the next step forward. These trials are being supported by the Global Health Innovative Technology fund and the German Federal Ministry of Education and Research. The Foundation for Innovative New Diagnostics (FIND) participated in development of the test, that was supported by the governments of the Netherlands and Australia, UK aid from the UK government, and the Bill & Melinda Gates Foundation.

Rationale

To address tuberculosis as the leading cause of morbidity and mortality among people living with HIV, new tests and strategies for detection of tuberculosis are urgently needed. Among the key priorities identified by the WHO, healthcare providers, patients, and advocacy groups is development of point-of-care, non-site specific tests for tuberculosis (Batz 2011; Pai 2012; Weyer 2011; WHO TTP 2014). To date, LF-LAM is the only commercially available point-of-care test for tuberculosis. LF-LAM, if sufficiently accurate, would satisfy many of the established minimum specifications for a point-of-care test for tuberculosis (Appendix 3; Batz 2011). LF-LAM testing could provide obvious benefits for HIV-positive people by earlier detection of pulmonary tuberculosis that may be missed by sputum smear microscopy and sputum Xpert MTB/RIF and extrapulmonary tuberculosis that may be missed by sputum-based testing. Studies that evaluate the impact of the use of LF-LAM on mortality and other patient outcomes are becoming available.

WHO guidelines on the use of urine LF-LAM were published in 2015 (WHO Lipoarabinomannan Policy Guidance 2015). Since 2015, additional evidence on the diagnostic accuracy of LF-LAM has emerged and is summarized in this Cochrane Review update. A draft of this review was used to inform the 2019 updated WHO guidelines on the use of LF-LAM.

OBJECTIVES

Primary objectives

We had two primary objectives:

1. To assess the accuracy of the lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis among HIV-positive adults with signs and symptoms of tuberculosis.
2. To assess the accuracy of LF-LAM for the diagnosis of active tuberculosis among HIV-positive adults irrespective of signs and symptoms of tuberculosis (i.e. unselected participants without consideration or assessment of tuberculosis signs and symptoms).

To estimate accuracy in HIV-positive individuals with signs and symptoms of tuberculosis (Objective 1), we combined studies in which presentation with signs and symptoms suggestive of tuberculosis was an inclusion criterion and refer to these as 'Studies with symptomatic participants'. To estimate accuracy in HIV-positive adults irrespective of signs and symptoms of tuberculosis (Objective 2), we combined studies that considered all HIV-positive individuals eligible to participate, including both individuals with and individuals without symptoms of tuberculosis and refer to these as 'Studies with unselected participants'. These

studies enrolled participants without assessment or consideration of specific signs or symptoms of tuberculosis.

Secondary objectives

Our secondary objective was to investigate potential sources of heterogeneity in test accuracy, including clinical setting, CD4 cell count, and prevalence of tuberculosis in the studies.

METHODS

Criteria for considering studies for this review

Types of studies

We included primary studies that evaluated the diagnostic accuracy of urine LF-LAM assay for the detection of active tuberculosis in people living with HIV and compared the index test results with a defined microbiological reference standard. We included studies from which we could extract true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) values.

Diagnostic studies for tuberculosis are largely cross-sectional in design, but may include some clinical follow-up as part of patient classification. We included cross-sectional studies and observational cohort studies. In addition, we included randomized controlled trials that evaluated the use of the index test on patient health outcomes, but that also reported sensitivity and specificity. Although the study design was a randomized trial for the purpose of determining the impact of the test on participant outcomes, the study design was a cross-sectional study for the purpose of determining the diagnostic accuracy of the index test in this review. We excluded case-control studies and other study designs. We excluded data reported only in abstracts, reviews, comments, and editorial notes. We did not include unpublished studies.

Participants

People living with HIV are at increased risk of tuberculosis and may present with symptoms of tuberculosis but may also be asymptomatic or have symptoms not routinely associated with tuberculosis. We included participants who were adults (15 years and older is considered 'adult' for purpose of tuberculosis surveillance) and HIV positive. We included studies in which there was a suspicion of tuberculosis among study participants based on the presence of signs and symptoms compatible with tuberculosis (studies with symptomatic participants), as well as studies that included participants who presented for medical care irrespective of signs and symptoms of tuberculosis (studies with unselected participants). Signs and symptoms of tuberculosis include cough, fever, weight loss, and night sweats. Participants who were known to have active tuberculosis and were taking anti-tuberculosis drugs were not included.

Index tests

We included studies that evaluated the lateral flow lipoarabinomannan (LF-LAM) assay Alere Determine™ TB LAM Ag test (Abbott, Palatine, IL, USA, previous Alere Inc., Waltham, MA, USA) on urine samples. As of May 2019, the Alere Determine™ TB LAM Ag test was the only commercially available LF-LAM assay that had been evaluated in published studies.

We included studies that evaluated the test at the manufacturer's recommended threshold for positivity, i.e. grade 1 and above on the updated reference scale card with four band intensities graded on a scale of 1 to 4. For studies that used the prior reference scale card with band intensities graded on a scale of 1 to 5, we included those that evaluated the test at grade 2 and above corresponding to the current recommended positivity threshold. We excluded studies that did not use a positivity threshold corresponding to the manufacturer's recommendations. Results summarizing diagnostic accuracy at older thresholds (grade 1 on a scale of 1 to 5) can be found in the original review (Shah 2016).

Target conditions

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

Reference standards

We required studies to diagnose tuberculosis using the following microbiological reference standard.

- 'Tuberculosis' is defined as a positive *M tuberculosis* culture or NAAT.
- 'Not tuberculosis' is defined as a negative *M tuberculosis* culture and NAAT (if performed).

NAAT tests 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); Xpert® MTB/RIF assay (Cepheid, Sunnyvale, USA); and Xpert® MTB/RIF Ultra.

We considered a higher quality reference standard to be one in which two or more specimen types were evaluated for tuberculosis diagnosis in all participants as part of a standardized study algorithm. We considered a lower quality reference standard to be one in which only one specimen type was evaluated for tuberculosis diagnosis, or if there was no algorithm defined to ensure a standardized approach for specimen collection and testing.

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 tuberculosis diagnosis may be missed for the following reasons: people living with HIV 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 tuberculosis cannot produce sputum at all (Lawn 2013), or have extrapulmonary tuberculosis without pulmonary tuberculosis. 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 tuberculosis, may reduce the number

of cases of tuberculosis incorrectly classified as 'not tuberculosis' by culture or NAAT if performed.

In the original Cochrane Review, we additionally considered a 'composite microbiological and clinical reference standard' recognizing that microbiological reference standards alone may fail to detect tuberculosis in patients with tuberculosis. However, our original review found relatively little data using a composite reference standard; found heterogeneity in defining and applying composite reference standards; and found a relatively modest impact on pooled estimates of sensitivity and specificity comparing microbiological and composite reference standards. Results assessing diagnostic accuracy against a composite reference standard can be found in the original review (Shah 2016).

Given the limitations of the reference standard, we might have considered pursuing latent class analysis (Chu 2009; Kohli 2018). However, we lacked patient-level data on the type of specimen and LAM is not site-specific, meaning a positive LAM alone does not tell us whether the patient has pulmonary or extrapulmonary tuberculosis.

Search methods for identification of studies

We attempted to identify all relevant studies regardless of language or publication status (published, unpublished, in press, and ongoing). As mentioned, we only included published studies in this review.

Electronic searches

We performed literature searches up to 11 May 2018 in the following databases using the search terms reported in Appendix 4: the Cochrane Infectious Diseases Group Specialized Register; MEDLINE (PubMed, from 1966); Embase (OVID, from 1947); 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; LILACS (BIREME, from 1982); and SCOPUS (from 1995). We also searched Clinicaltrials.gov 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. We included search results from the original review and re-evaluated previously included studies to determine if the studies met the refined inclusion criteria.

Searching other resources

We further examined reference lists of relevant reviews and studies and searched the WHO websites.

Data collection and analysis

Selection of studies

We used Covidence systematic review software to manage the selection of studies (Covidence 2017). Two review authors (MS and SB) independently examined all titles and abstracts identified from the electronic search to determine potentially eligible studies. We 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). We included studies from the original

review if still eligible according to the predefined eligibility criteria. We maintained a list of excluded studies and the reasons for exclusion, and recorded these details in the Characteristics of excluded studies table and we prepared a PRISMA diagram.

Data extraction and management

We developed a standardized data extraction form and piloted the form on two of the included studies. Based on the pilot, we finalized the form (Appendix 5). Then two review authors (MS and SB) independently extracted data from each included study on the following characteristics.

- Author, publication year, study design, country/countries, clinical setting (outpatient or inpatient).
- Participants: age, gender, HIV-status, CD4 count, tuberculosis history, clinical status (asymptomatic, symptomatic).
- Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) items.
- Cut-off used for determining a positive index test result and the reference card used.
- Samples collected (sputum and/or extrapulmonary samples).
- Reference standard(s).
- The number of tuberculosis cases in the study.
- Number of TP, FN, FP, and TN values.
- Missing or unavailable test results.

We assigned country income status (high income, upper- and lower-middle income, and low income) as classified by the World Bank (World Bank 2018). In addition, we classified a country as being high burden or not high burden for tuberculosis/HIV according to the post-2015 era classification by the WHO (WHO Global Tuberculosis Report 2018). For studies that included both participants with HIV and without HIV infection, we extracted data only for participants with HIV. We contacted study authors for clarifications on the LF-LAM positivity threshold used if data were missing.

We used REDCap electronic data capture tools (Harris 2009) hosted at OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark (SDU Open) to collect and manage study data. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless downloads to common statistical packages; and 4) procedures for importing data from external sources. With regard to the use of REDCap, the content in this review is solely the responsibility of the authors.

Assessment of methodological quality

We used the QUADAS-2 tool tailored to this review to assess the quality of the included studies (Whiting 2011; Appendix 6). QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing (flow and timing domain includes differential verification of tuberculosis 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 the guidance on how to appraise the questions in each domain. Then, one review author (SB) piloted the tool with two of the included studies and finalized the

QUADAS-2 tool. Two review authors (MS and SB) independently completed the QUADAS-2 assessment. We resolved disagreements through discussion or consulted a third review author (KRS). We present the results of this quality assessment in the text, tables, and graphs.

Statistical analysis and data synthesis

We performed descriptive analyses of the characteristics of the included studies using Stata 15 (StataCorp 2017), 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 (CIs). We presented individual study results graphically by plotting the estimates of sensitivity and specificity (and their 95% CIs) in forest plots using Review Manager 5 (RevMan 5) (Review Manager 2014).

We presented results at the current manufacturer reference scale card for test interpretation, with band intensities graded 1 to 4, and considered all test results at grade 1 and above as positive. The prior reference scale card with five band intensities was used in the original Cochrane Review with grade 2 considered as positivity threshold that corresponds to the current grade 1 band intensity (Appendix 2). The original review also included several analyses at grade 1 which is no longer recommended for determining test positivity. To allow consistent comparisons, we converted results from older studies that used the 'grade 2' threshold and treated these as 'grade 1' in the updated review. As such, analyses labelled at 'grade 2' in the original Cochrane Review are in this review considered according to the new manufacturer reference card as 'grade 1'. Studies in the original review that used the 'grade 1' threshold on the prior reference card were not included as this threshold is no longer recommended for determining test positivity.

We grouped the studies evaluating LF-LAM for: (I) diagnosis of tuberculosis in HIV-positive people with signs and symptoms of tuberculosis i.e. 'Studies with symptomatic participants' and (II) diagnosis of tuberculosis in HIV-positive people, irrespective of signs and symptoms of tuberculosis i.e. 'Studies with unselected participants'.

When data were sufficient, we carried out meta-analyses to estimate LF-LAM pooled sensitivity and specificity with a bivariate random-effects model (Chu 2006; 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 estimated all models using a Bayesian approach implemented using OpenBUGS (Lunn 2009). 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. Because most meta-analyses involved few studies (eight or less), which could lead the model to be just identified, we chose to use low-information prior distributions for most parameters and a more informative prior on the between-study standard deviations

which are particularly sensible in meta-analyses with few studies (Spiegelhalter 2004).

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 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. The 2.5% and 97.5% prior distribution quantiles for the pooled sensitivity or pooled specificity are 2.0% and 98.0%, slightly wider than for a standard normal distribution. 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. The resulting median 2.5% and 97.5% prior distribution quantiles for the predicted sensitivity or predicted specificity are 0.1% and 99.9%. 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 distributions) and the OpenBUGS programs we used to estimate them in Appendix 7.

To study the sensitivity of 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 OpenBUGS 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% confidence interval (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.

We also estimated the predicted sensitivity and specificity in a future study together with their 95% CrIs. The predicted estimate is our best guess for the estimate in a future study and is the same as the pooled estimate. The CrIs, however, may be different. These values were derived from the predicted region typically reported in a bivariate meta-analysis plot. If there is no heterogeneity between the included studies, the CrI around the predicted estimate will

be the same as the CrI around the pooled estimate. On the other hand, if there is considerable heterogeneity between studies, the CrI around the predicted estimate will be much wider than the CrI around the pooled estimate. We generated summary plots displaying the individual study estimates for sensitivity and specificity, the pooled estimate for sensitivity and specificity with the 95% credible region and the 95% prediction region using R (R Statistical Computing 2018).

In our original review we evaluated the incremental change in sensitivity and specificity when combining LF-LAM with smear microscopy or Xpert MTB/RIF (Shah 2016). We did not undertake analysis of incremental benefit in the current review as it was beyond the scope of this review, and data within published manuscripts were limited.

Approach to uninterpretable LF-LAM results

We excluded uninterpretable test results from the meta-analyses of sensitivity and specificity, but we reported the number and proportion of uninterpretable test results from each study when such data were available.

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: clinical setting (inpatient versus outpatient); CD4 count (CD4 ≤ 200; CD4 ≤ 100; CD4 101-200; CD4 > 200 and; CD4 > 100 cells per μL). To further investigate heterogeneity, we performed a subgroup analysis by the prevalence of tuberculosis in the studies and classified prevalence as greater than the median value versus less than or equal to the median value. We investigated heterogeneity separately for studies with symptomatic participants and studies with unselected participants. We generated the plots depicting the pooled results within CD4 count categories using R (R Statistical Computing 2018).

Sensitivity analyses

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 with a higher quality reference standard, for example studies that included two or more specimen types. 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).
- Studies that used only fresh urine specimens for LAM testing.
- Studies initially categorized as 'studies among unselected participants' that included more than 80% of symptomatic participants were re-categorized as 'studies with symptomatic participants'. We conducted this analysis to explore the possibility that these studies represented a comparable population to the studies of symptomatic participants even

though participants were not explicitly enrolled in the study on the basis of specific tuberculosis symptoms.

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).

Assessment of the certainty of the evidence

We assessed the certainty of evidence as recommended using the GRADE approach (Balshem 2011; Schünemann 2008; Schünemann 2016). As recommended, we rated the certainty 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 certainty 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 (SB, MS, ND, and KRS) discussed judgments and applied GRADE in the following way.

- Risk of bias: we used QUADAS-2 to assess risk of bias.
- Indirectness: 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 index test and asked are differences sufficient to lower certainty in results?
- 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 tuberculosis 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 and extensive outreach to tuberculosis researchers to identify studies.

R E S U L T S

Results of the search

We identified 15 unique studies that met the inclusion criteria of this review. We included data from six published manuscripts from the original WHO Guidelines (WHO Lipoarabinomannan Policy Guidance 2015), and Cochrane Review (Shah 2016) that met the refined inclusion criteria (Bjerrum 2015; Drain 2015a; LaCourse 2016; Nakiyingi 2014; Peter 2012a; Peter 2015), and nine new studies identified in the updated search (Drain 2016; Floridia 2017; Hanifa 2016; Huerga 2017; Juma 2017; Lawn 2017; Pandie 2016; Peter 2016; Thit 2017). Of six previously included studies, three were excluded because they did not use the currently recommended threshold for test positivity (Balcha 2014; Drain 2014a; Lawn 2012a); one previously included abstract (Lawn 2014) was added as an updated published manuscript (Lawn 2017); one abstract remained unpublished (Andrews 2014); and one abstract was published but did not provide diagnostic accuracy data (Drain 2014b). Eight studies evaluated the accuracy of LF-LAM for tuberculosis diagnosis in participants with signs and symptoms suggestive of tuberculosis. Seven studies evaluated the accuracy of LF-LAM for diagnosis of unselected participants without assessment of symptoms (i.e. patients may or may not have had tuberculosis signs and symptoms at enrolment).

Figure 1 shows the flow of studies in the review. We listed the excluded studies and the reasons for their exclusion in the 'Characteristics of excluded studies' section. All studies were written in English.

Figure 1. Study flow diagram
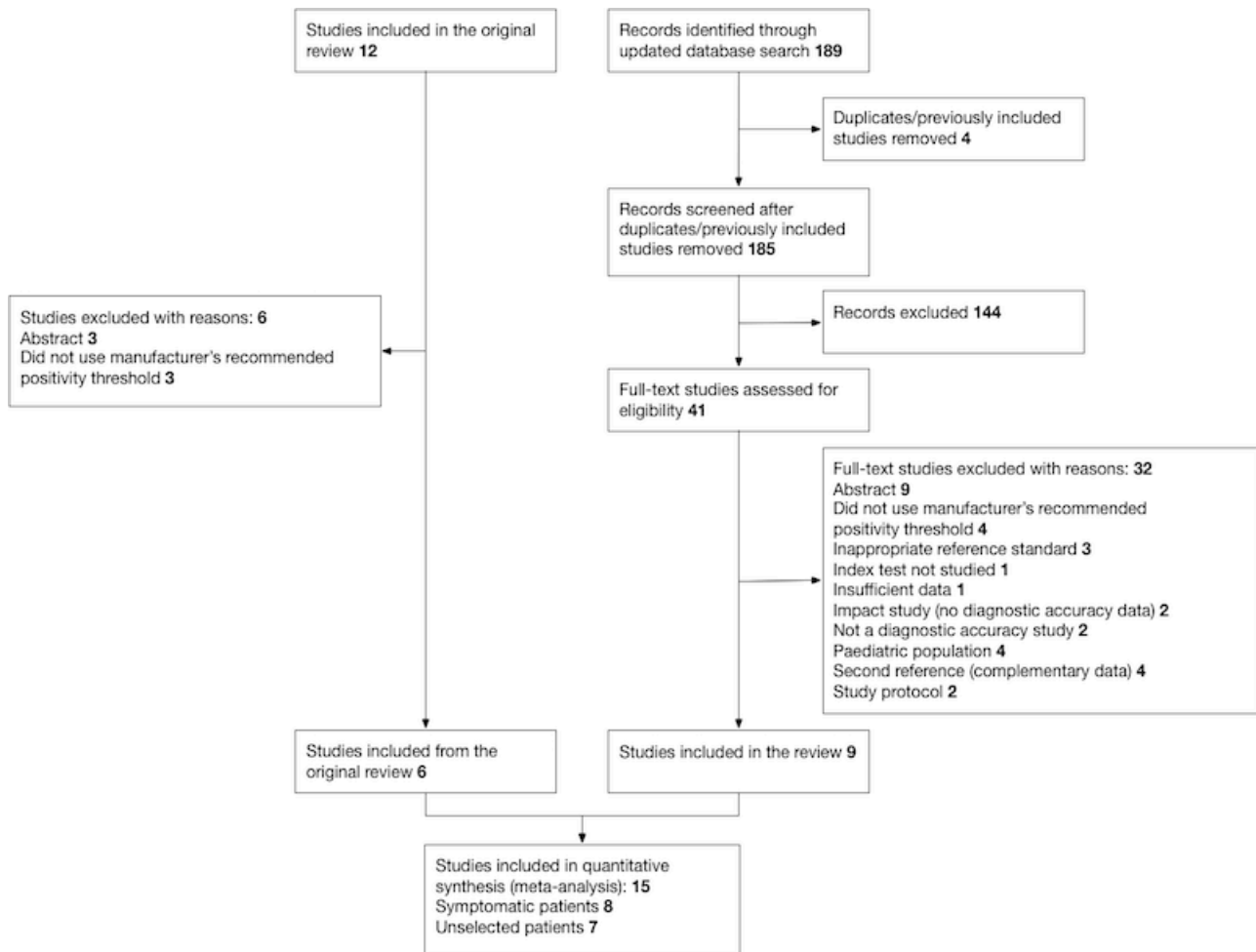
Methodological quality of included studies

Risk of bias and applicability concerns for each of the 15 included studies is shown in Figure 2; Figure 3.

Figure 2. Studies with symptomatic participants - Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

Figure 3. Studies with unselected participants - Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

Studies with symptomatic participants

Eight studies were included that evaluated LF-LAM for tuberculosis diagnosis among symptomatic participants suspected of tuberculosis. Risk of bias and applicability concerns for the studies with symptomatic participants are shown in Figure 2.

In the patient selection domain, we considered six studies (75%) to be at high risk of bias because: (1) the study excluded all smear-positive participants (Drain 2016); (2) the studies excluded participants who could not expectorate or produce sputum despite sputum induction (Drain 2016; Nakiyingi 2014; Peter 2015); (3) the study excluded participants from analysis if, in the absence of a positive reference standard result, there was one sample with no Xpert MTB/RIF result or contaminated result (culture) (Huerga 2017); (4) the study only included patients suspected of extrapulmonary tuberculosis and excluded patients suspected

of pulmonary tuberculosis (Juma 2017); (5) the study only included participants with pericardial effusion and suspected tuberculosis and excluded participants suspected of other forms of tuberculosis (Pandie 2016). All studies were cross-sectional, cohort or randomized controlled studies. Regarding applicability, seven studies (88%) had low concern in the patient selection domain because the studies included the appropriate participants and settings. We judged one study (12%) to have high concern for applicability as the participants did not resemble people with presumed HIV/tuberculosis co-infection i.e. participants were smear-negative HIV-positive and HIV-negative patients with a Karnofsky Performance score < 50 (Drain 2016).

In the index test domain, we judged one study (12%) at high risk of bias as the study used grade 2 (on the updated reference scale card) as the test positivity threshold, as opposed to the

current manufacturer recommendation to use grade 1 to define test positivity (Juma 2017). We also considered this study to have high concern of applicability because the test procedure was inconsistent with the manufacturers' recommendations (Juma 2017). The remaining studies all used the recommended threshold for positivity and interpreted the test without knowledge of the results of the reference standard, and we considered them to have low concern for applicability.

In the reference standard domain, we considered seven studies (88%) to be at high risk of bias because: (1) the studies did not include testing of any extrapulmonary specimens (Drain 2016, Peter 2015); (2) the study did not include testing of any respiratory samples (Juma 2017); (3) the study only tested respiratory samples for some of the participants (Pandie 2016); (4) the study only tested extrapulmonary specimens in addition to respiratory samples for some of the participants (Huerga 2017); (5) health providers selected the sites for testing based on their own clinical suspicion (Peter 2012a; Peter 2016). We deemed three studies at high concern for applicability as they lacked a study or protocol directed testing (Pandie 2016; Peter 2012a; Peter 2016). In these studies, 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 tuberculosis.

In the flow and timing domain, we considered four studies (50%) to be at high risk of bias because not all participants received the same reference standard (Huerga 2017; Peter 2012a; Peter 2016), or because not all participants were included in the two-by-two tables (Huerga 2017; Pandie 2016). We judged the remaining studies to be at low risk of bias because all participants received the index test, the same reference standard and no participants were excluded from the two-by-two table.

Studies with unselected participants

Seven studies contributed data for the purpose of evaluating LF-LAM for tuberculosis diagnosis among unselected participants who may or may not have tuberculosis signs or symptoms. Risk of bias and applicability concerns for the studies with symptomatic participants are shown in Figure 3.

In the patient selection domain, we considered four studies (57%) to be at high risk of bias because these studies excluded participants who could not expectorate or produce sputum samples (Bjerrum 2015; Drain 2015a; Floridia 2017; LaCourse 2016). All studies were cross-sectional or cohort studies. Regarding applicability, we judged that all studies (100%) included the appropriate participants and settings.

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 at the recommended positivity threshold without knowledge of the results of the reference standard. We considered the test conduct and interpretation in all studies to be applicable.

In the reference standard domain, we considered five studies (71%) to be at high risk of bias because these studies did not include microbiological testing on extrapulmonary specimens (Bjerrum 2015; Drain 2015a; Floridia 2017; LaCourse 2016; Thit 2017). Thit 2017 also did not report if the reference standard results were interpreted without knowledge of the index test result. One study

(Thit 2017) did not report if they speciated mycobacteria isolates and was judged to have unclear concern for applicability. We judged the remaining six studies to be of low concern in terms of applicability.

In the flow and timing domain, we considered two studies (29%) to be at high risk of bias because the studies collected specimens for index and reference standard tests up to six months apart (Hanifa 2016; Thit 2017), and Hanifa 2016 excluded clinical tuberculosis cases from analysis. We considered the remaining five studies (71%) to be at low risk of bias because all participants received the index test and the same reference standard, and no participants enrolled were excluded from the two-by-two table.

Findings

The 15 included studies involved 6814 participants, 1761 (26%) with tuberculosis. Eight studies evaluated the accuracy of LF-LAM for tuberculosis diagnosis in participants with signs and symptoms suggestive of tuberculosis involving 3449 participants, 1277 (37%) with tuberculosis. Seven studies evaluated the accuracy of LF-LAM for diagnosis of unselected participants that may or may not have had tuberculosis signs and symptoms at enrolment involving 3365 participants, 432 (13%) with tuberculosis.

All studies were performed in countries with a high tuberculosis/HIV burden (WHO Global Tuberculosis Report 2018), and classified as low-income or middle-income countries (World Bank 2018). We noted substantial differences in the studies for the following characteristics: type of population ('studies with symptomatic participants' and 'studies with unselected participants'); setting (inpatients versus outpatients); median CD4 cell count; tuberculosis prevalence; inclusion and exclusion of participants based on whether or not they could produce sputa; and whether patients were evaluated for pulmonary tuberculosis, extrapulmonary tuberculosis, or both. The key study characteristics are summarized in Table 1 (Summary characteristics of included studies) and in Characteristics of included studies.

Table 2 presents pooled sensitivity and specificity results for LF-LAM grouped by the type of population, 'studies among symptomatic participants' and 'studies among unselected participants'.

I. Diagnostic accuracy of LF-LAM for tuberculosis diagnosis in HIV-positive adults with signs and symptoms of tuberculosis

Of the 15 included studies, eight evaluated the accuracy of LF-LAM for tuberculosis diagnosis in participants with signs and symptoms suggestive of tuberculosis (Drain 2016; Huerga 2017; Juma 2017; Nakiyingi 2014; Pandie 2016; Peter 2012a; Peter 2015; Peter 2016). The suggestive signs and symptoms of tuberculosis varied from study to study, as described in Table 1 and in Characteristics of included studies, but were often based on any of cough, fever, weight loss, or night sweats as per the WHO symptom screening rule WHO ICF 2011. The tuberculosis prevalence in the study population ranged from 29% to 63%. Two studies were conducted exclusively among patients with presumed extrapulmonary tuberculosis (Juma 2017; Pandie 2016). Four studies were conducted exclusively in an inpatient setting (Juma 2017; Pandie 2016; Peter 2012a; Peter 2016); two studies exclusively in an outpatient setting (Drain 2016; Peter 2015); and two studies in both inpatient and outpatient settings (Huerga 2017; Nakiyingi 2014). The median CD4 cell count ranged from 81 to 210

cells per µL across the eight studies, lower in studies evaluating inpatients (median CD4 between 81 to 139 cells per µL) compared to studies evaluating outpatients (median CD4 was 168 to 210 cells per µL). See Table 1 and Characteristics of included studies.

Primary analysis, LF-LAM for tuberculosis in HIV-positive adults with signs and symptoms of tuberculosis

For determining the overall accuracy of LF-LAM in HIV-positive adults with signs and symptoms of tuberculosis, eight studies provided data for 3449 participants, including 1277 (37%) with tuberculosis. Sensitivity estimates ranged from 23% to 68%, and specificity estimates from 75% to 100% (Figure 4).

Figure 4.   Forest plots of LF-LAM sensitivity and specificity for tuberculosis against a microbiological reference standard for studies among symptomatic participants. The individual studies are ordered by decreasing sensitivity. 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).

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Juma 2017	15	3	7	42	0.68 [0.45, 0.86]	0.93 [0.82, 0.99]
Huerga 2017	102	19	54	100	0.65 [0.57, 0.73]	0.84 [0.76, 0.90]
Peter 2012a	58	31	58	94	0.50 [0.41, 0.59]	0.75 [0.67, 0.82]
Peter 2016	156	94	186	736	0.46 [0.40, 0.51]	0.89 [0.86, 0.91]
Nakiyingi 2014	136	21	231	609	0.37 [0.32, 0.42]	0.97 [0.95, 0.98]
Pandie 2016	12	0	24	2	0.33 [0.19, 0.51]	1.00 [0.16, 1.00]
Drain 2016	13	1	44	32	0.23 [0.13, 0.36]	0.97 [0.84, 1.00]
Peter 2015	41	27	140	361	0.23 [0.17, 0.29]	0.93 [0.90, 0.95]
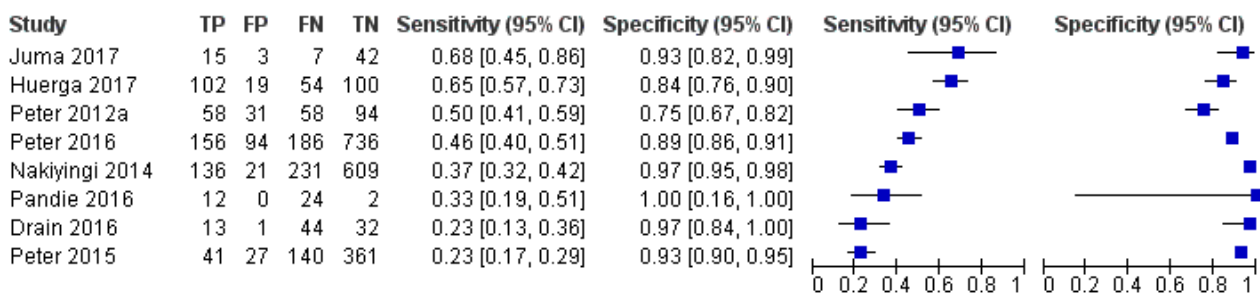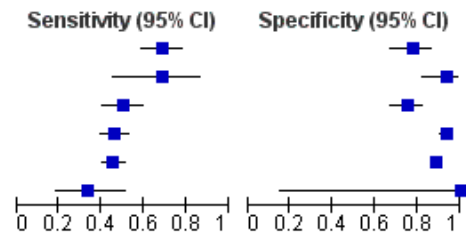
Juma 2017 evaluated diagnostic accuracy for extrapulmonary tuberculosis (all forms) exclusively and had the highest sensitivity of 68%. Pandie 2016 evaluated accuracy for pericardial tuberculosis and found sensitivity of 33%. Sensitivity was lowest in the studies by Peter 2015 and Drain 2016 that both reported a sensitivity of 23%. Differences between these studies and the other studies in this analysis were the setting (outpatient only), focus on pulmonary tuberculosis (no extrapulmonary samples were taken), and exclusion of participants unable to produce sputum. In particular, Drain 2016 included smear-negative participants with presumed tuberculosis and a small number of HIV-negative participants. Drain 2016 further excluded participants with a low Karnofsky score in order to target relatively well outpatients, where smear-negative tuberculosis is often seen. Specificity was lowest for Peter 2012a a study that included only inpatients and differed from other studies in that both sputum and non-sputum-based sampling was performed at the discretion of the attending clinical team and not study directed. Pandie 2016 reported the highest specificity of 100%. This study, as mentioned, differed from others by evaluating accuracy only for pericardial tuberculosis. The authors further excluded a number of participants in the analysis for unknown reasons and reported two true negatives and zero false positives that may have inflated specificity. The pooled sensitivity and specificity (95% credible interval (CrI)) were 42% (31% to 55%) and 91% (85% to 95%) (Table 2). Appendix 8 presents the pooled and predicted sensitivity and specificity estimates together with the credible and prediction regions for LF-LAM for tuberculosis detection. Prediction intervals were very wide in all meta-analysis models indicating between-study heterogeneity (data not shown).

Uninterpretable index test results

Studies reported few uninterpretable test results. Peter 2012a reported that 1% to 2% of 423 tests remained 'indeterminate' (defined as a broken band in the patient window) after repeat testing. Peter 2016 reported that 0.3% (3/1172) test results remained 'invalid' after a repeat test (defined as no control band identified in the patient window or a broken/ incomplete band in the patient window). Peter 2015 reported that fewer than 1% of LF-LAM strip tests failed on the first attempt and required a second strip to produce valid results. Four studies reported that a valid LF-LAM result was obtained on the first attempt for all tests (Drain 2016; Huerga 2017; Juma 2017; Nakiyingi 2014), and one study did not provide information on uninterpretable tests (Pandie 2016).

Investigations of heterogeneity

LF-LAM accuracy among symptomatic participants, stratified by setting

LF-LAM accuracy in inpatient settings

Six studies were conducted among inpatients involving 2253 participants, 868 (39%) with tuberculosis (Huerga 2017; Juma 2017; Nakiyingi 2014; Pandie 2016; Peter 2012a; Peter 2016). Sensitivity estimates ranged from 33% to 69% and specificity estimates ranged from 75% to 100% (Figure 5). The highest sensitivity (69%) was reported by Huerga 2017 with a relatively low specificity (78%). This study enrolled inpatients with tuberculosis symptoms who had at least one of the following features: who were severely ill or who had CD4 < 200 cell per µL (median CD4 109) or who had a low body mass index (BMI). The study did not include microbiological or histological evaluation of extrapulmonary specimens for tuberculosis in their reference standard, which may have led to LAM-positive participants with extrapulmonary tuberculosis being misclassified as 'false-positive' and lowered specificity. Pandie 2016 reported the lowest sensitivity (33%) and a specificity of 100%. As mentioned, this study evaluated accuracy for pericardial tuberculosis and excluded participants from specificity analysis that may have affected the specificity

estimate. The pooled sensitivity and specificity (95% CrI) among inpatients were 52% (40% to 64%) and 87% (78% to 93%) (Table 2).

Figure 5. Forest plots of LF-LAM sensitivity and specificity for tuberculosis against a microbiological reference standard for studies among symptomatic participants, by health care setting. TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. The individual studies are ordered by decreasing sensitivity. 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 accuracy in outpatient settings

Four studies were conducted among outpatients involving 1196 participants, 409 (34%) with tuberculosis (Drain 2016; Huerga 2017; Nakiyingi 2014; Peter 2015). Sensitivity estimates ranged from 18% to 58% and specificity estimates ranged from 93% to 99% (Figure 5). The highest sensitivity (58%) was reported by Huerga 2017 where outpatients included were severely ill, or had a CD4 < 200 cell per µL, or had a low BMI below 17 kg/m2. Among outpatients, pooled

sensitivity and specificity were 29% (17% to 47%) and 96% (91% to 99%) (Table 2).

LF-LAM accuracy among symptomatic participants, stratified by CD4 count

We included five studies in the analyses by CD4 count (Nakiyingi 2014; Peter 2012a; Peter 2015; Peter 2016). See Figure 6 and Figure 7.

Figure 6. Forest plots of LF-LAM sensitivity and specificity for tuberculosis against a microbiological reference standard for studies with symptomatic participants, stratified by CD4 (CD4 > 200 and CD4 ≤ 200; CD4 > 100 and CD4 ≤ 100). The individual studies are ordered by decreasing sensitivity. 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).
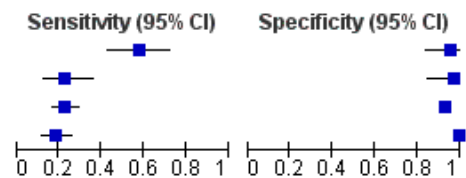
Figure 7. Plot by CD4 of diagnostic accuracy in adults with signs and symptoms of tuberculosis. (A) Sensitivity by CD4 strata; (B) Specificity by CD4 strata. The circle represents the pooled estimates (median), with bars representing 95% credible intervals.
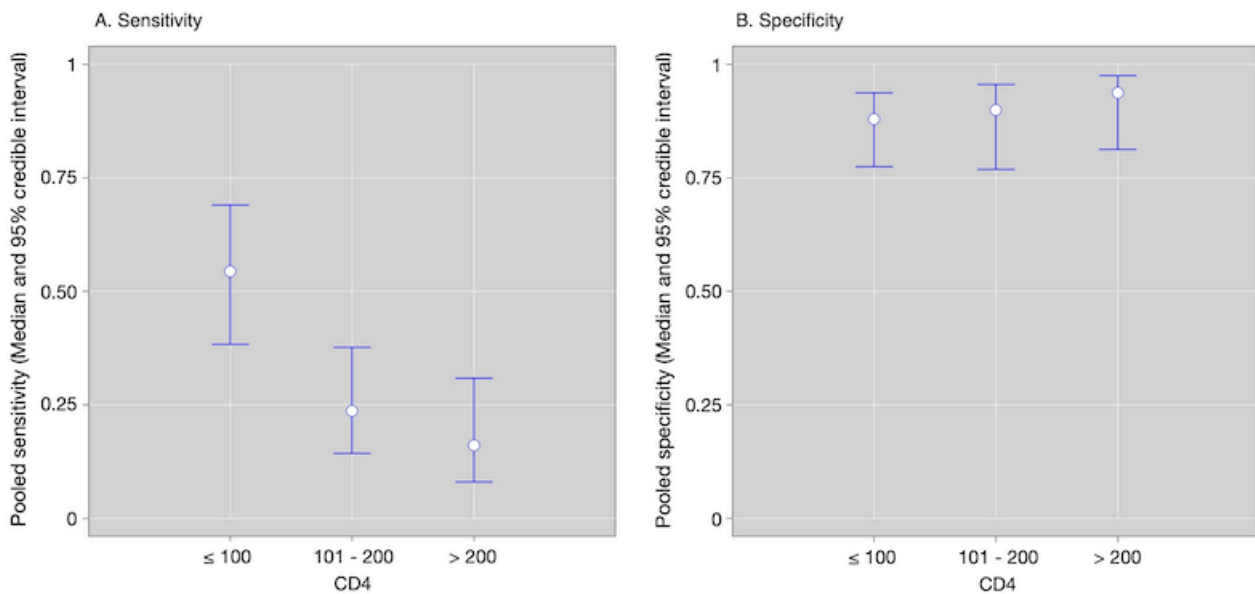
LF-LAM accuracy stratified by CD4 > 200 cells per μL and ≤ 200 cells per μL

Three studies evaluated LF-LAM in participants with CD4 > 200 cells per μL (Nakiyingi 2014; Peter 2012a; Peter 2016). Sensitivity estimates ranged from 9% to 27% and specificity estimates ranged from 83% to 99% (Figure 6). In the four studies that evaluated participants with CD4 ≤ 200 cells per μL, sensitivity estimates ranged from 24% to 58% and specificity estimates ranged from 72% to 95% (Nakiyingi 2014; Peter 2012a; Peter 2015; Peter 2016), (Figure 6). The pooled sensitivity (95% CrI) was higher among participants with CD4 ≤ 200 cells per μL at 45% (31% to 61%) (4 studies, 1825 participants, 40% with tuberculosis) versus 16% (8% to 31%) among those with CD4 > 200 cells per μL (3 studies, 738 participants, 22% with tuberculosis) (Table 2); a significant difference in pooled sensitivity (95% CrI) of -29% (-47% to -9%). The pooled specificity was 89% (77% to 94%) for participants with CD4 ≤ 200 cells per μL and 94% (81% to 97%) for those with CD4 > 200 cells per μL (Table 2), a difference in pooled specificity of 5% (-8% to 17%).

When we limited the analysis to studies involving inpatients with CD4 ≤ 200 cells per μL, the pooled sensitivity and specificity were 54% (34% to 73%) and 80% (58% to 91%) (2 studies, 1009 participants, 34% with tuberculosis) (Peter 2012a; Peter 2016). Only one study reported data for outpatients with CD4 ≤ 200 cells per μL (Peter 2015).

LF-LAM accuracy stratified by CD4 > 100 cells per μL and ≤ 100 cells per μL

Four studies evaluated LF-LAM in participants with CD4 > 100 cells per μL (Nakiyingi 2014; Pandie 2016; Peter 2015; Peter 2016). Sensitivity estimates ranged from 12% to 19% and specificity estimates ranged from 92% to 100% (Figure 6). In the five studies that evaluated participants with CD4 ≤ 100 cells per μL, sensitivity estimates ranged from 30% to 65% and specificity estimates ranged from 75% to 94% (Nakiyingi 2014; Pandie 2016; Peter 2012a; Peter

2015; Peter 2016). One study had no estimable specificity, as the authors reported zero true negatives (Pandie 2016). The pooled sensitivity (95% CrI) was higher among participants with CD4 ≤ 100 cells per μL at 54% (38% to 69%) (4 studies, 1239 participants, 41% with tuberculosis) versus 17% (10% to 27%), (4 studies, 1519 participants, 28% with tuberculosis) among those with CD4 > 100 cells per μL, a significant difference in pooled sensitivity of -37% (-53% to -19%). The pooled specificity was 88% (77% to 94%) for participants with CD4 ≤ 100 cells per μL and 95% (89% to 98%) for those with CD4 > 100 cells per μL (Table 2), a difference in pooled specificity of 7% (-1% to 18%).

When we limited the analysis to studies involving inpatients with CD4 ≤ 100 cells per μL (Peter 2012a; Peter 2016), the pooled sensitivity and specificity were 61% (40% to 78%) and 81% (61% to 91%). Only one study reported data for outpatients with CD4 ≤ 100 cells per μL (Peter 2015).

LF-LAM accuracy stratified by CD4 strata

We observed that LF-LAM pooled sensitivity increased as the degree of immunodeficiency increased, from 16% (8% to 31%) in patients with CD4 cell count >200 cells per μL; to 24% (14% to 38%) in patients with CD4 count between 101 to 199; to 54% (38% to 69%) in patients with CD4 ≤ 100 (Figure 7). Also, we observed that most participants contributing data for the CD4 ≤ 200 cells per μL stratum (1825 participants including 722 (40%) with tuberculosis) were participants with CD4 ≤ 100 cells per μL (1239 participants including 512 (41%) with tuberculosis) (Table 2).

LF-LAM accuracy among symptomatic participants, stratified by tuberculosis prevalence

The median prevalence of tuberculosis in studies with symptomatic participants was 43% (interquartile range (IQR) 32% to 60%). In an analysis by tuberculosis prevalence, we found that pooled sensitivity and specificity for symptomatic participants in settings with tuberculosis prevalence of greater than 43% were 44% (27% to

![Cochrane Library logo] Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

62%) and 86% (73% to 94%) and 39% (21% to 63%) and 95% (89% to 97%) in settings with a tuberculosis prevalence less than 43%.

Sensitivity analyses

In evaluation of the accuracy of LF-LAM for detection of tuberculosis, we undertook sensitivity analyses by limiting inclusion in the meta-analysis to the following.

- Studies with low risk of bias for patient selection (two studies)
- Studies with low risk of bias in the reference standard domain (one study)
- Studies that used fresh urine samples (four studies) rather than stored urine sample

Two studies with greater than 80% symptomatic participants (Bjerrum 2015; Lawn 2017) were re-assigned from 'studies of unselected participants' to 'studies of symptomatic participants'.

These sensitivity analyses only included few studies and participants and made little difference to any of the findings (Table 3).

II. Diagnostic accuracy of LF-LAM for tuberculosis diagnosis in HIV-positive unselected adults irrespective of signs and symptoms for tuberculosis

Of the 15 studies included, seven studies evaluated the accuracy of LF-LAM for diagnosis in participants, irrespective of sign and symptoms ('unselected participants') (Bjerrum 2015; Drain 2015a; Floridia 2017; Hanifa 2016; LaCourse 2016; Lawn 2017; Thit 2017). The tuberculosis prevalence in the study population varied from 1% in LaCourse 2016 to 33% in Lawn 2017. Six of the studies reported the proportion of symptomatic participants that were included (e.g. having a positive WHO symptoms screen) which varied from 19% (LaCourse 2016) to more than 90% of participants in two studies (Bjerrum 2015; Lawn 2017). No studies reported results specifically for individuals who were asymptomatic. Four studies were carried out in an outpatient setting (Drain 2015a; Floridia 2017; Hanifa 2016; LaCourse 2016), one study exclusively in an inpatient setting (Lawn 2017), and two studies in both inpatient and outpatient settings (Bjerrum 2015; Thit 2017). The median CD4 cell count across studies of unselected adults ranged from 111 to 437 cells per μL across studies. See Table 1 and Characteristics of included studies.

Primary analysis, LF-LAM for tuberculosis in HIV-positive adults irrespective of signs and symptoms of tuberculosis

For determining the overall accuracy of LF-LAM in HIV-positive adults irrespective of signs and symptoms of tuberculosis, seven studies provided data for 3365 participants, including 432 (13%) with tuberculosis. Sensitivity estimates ranged from 0% to 67%, and specificity estimates from 64% to 99% (Figure 8).

Figure 8. Forest plots of LF-LAM sensitivity and specificity for tuberculosis against a microbiological reference standard for studies among unselected participants not assessed for signs and symptoms of tuberculosis. The individual studies are ordered by decreasing sensitivity. 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).

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Thit 2017	36	165	18	298	0.67 [0.53, 0.79]	0.64 [0.60, 0.69]
Bjerrum 2015	24	21	31	393	0.44 [0.30, 0.58]	0.95 [0.92, 0.97]
Lawn 2017	53	3	83	274	0.39 [0.31, 0.48]	0.99 [0.97, 1.00]
Floridia 2017	26	8	64	874	0.29 [0.20, 0.39]	0.99 [0.98, 1.00]
Drain 2015a	15	16	39	250	0.28 [0.16, 0.42]	0.94 [0.90, 0.97]
Hanifa 2016	3	5	37	363	0.07 [0.02, 0.20]	0.99 [0.97, 1.00]
LaCourse 2016	0	13	3	250	0.00 [0.00, 0.71]	0.95 [0.92, 0.97]
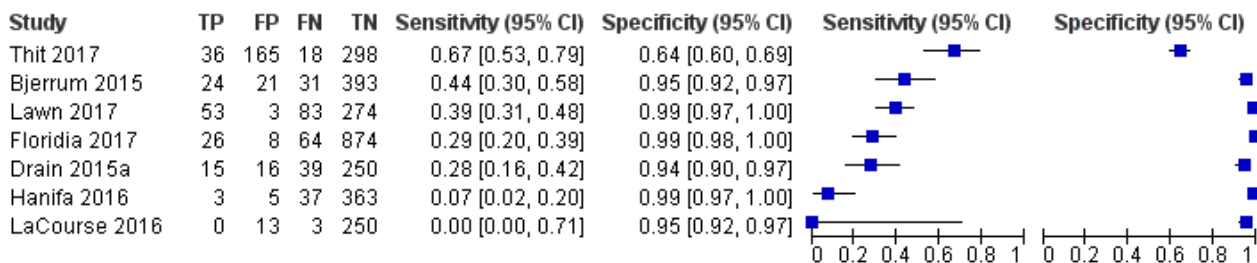
Sensitivity was lowest (0%) in LaCourse 2016, that differed from the other studies by including a) a population of exclusively pregnant women attending an antenatal care setting, b) a low proportion of symptomatic participants (19%), c) a low tuberculosis prevalence (1%), and d) a high median CD4 cell count (437 cells per μL). Specificity was lowest (64%) in the study by Thit 2017 that also reported the highest sensitivity (67%). This study reported that more than 90% of the FP results were grade 1 positive results (classified as positive according to current manufacturer recommendations). Participants included had a median CD4 at 270 cells per μL and 33% were symptomatic at enrolment. The study evaluated sputum samples only and allowed a follow-up for six months from LF-LAM testing at enrolment to final classification of participants as 'Tuberculosis' or 'Not tuberculosis' cases. Thit 2017 further differed from the other studies by being conducted in Myanmar, and is the only study included in this review that

evaluated LF-LAM in a setting outside sub-Saharan Africa. The pooled sensitivity and specificity (95% CrI) were 35% (22% to 50%) and 95% (89% to 98%) (Table 2). Appendix 9 presents the pooled and predicted sensitivity and specificity estimates together with the credible and prediction regions for LF-LAM for tuberculosis detection. Prediction intervals were very wide in all meta-analysis models indicating between-study heterogeneity (data not shown).

Uninterpretable index test results

The included studies reported no uninterpretable results.

Investigations of heterogeneity

LF-LAM accuracy among unselected participants, stratified by setting

LF-LAM accuracy in inpatient settings

We identified three studies that evaluated accuracy among inpatients involving 537 participants, 159 (30%) with tuberculosis

(Bjerrum 2015; Lawn 2017; Thit 2017). Sensitivity estimates ranged from 39% to 88% and specificity estimates ranged from 39% to 99% (Figure 9).

Figure 9.   Forest plots of LF-LAM sensitivity and specificity for tuberculosis against a microbiological reference standard for studies among unselected participants, by health care setting. TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. The individual studies are ordered by decreasing sensitivity. 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).

Thit 2017 reported a very low specificity (39%) and a high sensitivity (88%), based on a relatively small sample size of 54 inpatients; eight (15%) with tuberculosis. The authors reported that 41 of the inpatients (76%) were symptomatic at enrolment with a median CD4 of 96 (IQR 37 to 277) cells per μL, which is comparable to evaluation of LF-LAM in a population with advanced HIV disease. The study further differed from the other studies by allowing a follow-up for six months for classification of participants and by being conducted in Myanmar as mentioned above. Most of the participants (91%) included in the studies by Bjerrum 2015 and Lawn 2017 were symptomatic and thus comparable to studies with symptomatic participants. See Table 1 and Characteristics of included studies. The pooled sensitivity and specificity (95% CrI) among unselected inpatients were 62% (41% to 83%) and 84% (48% to 96%) (Table 2).

LF-LAM accuracy in outpatient settings

Six studies were conducted among unselected outpatients, involving 2828 participants; 273 (10%) with tuberculosis (Bjerrum 2015; Drain 2015a; Floridia 2017; Hanifa 2016; LaCourse 2016; Thit 2017). Sensitivity estimates ranged from 0% to 63% and specificity estimates ranged from 67% to 99% (Figure 9). Pooled sensitivity and specificity (95% CrI) were 31% (18% to 47%) and 95% (87% to 99%) (Table 2).

LF-LAM accuracy among unselected participants, stratified by CD4 count

There were limited data to evaluate LF-LAM by CD4 threshold for unselected participants. Five studies contributed data to the analyses by CD4 count (Bjerrum 2015; Drain 2015a; Hanifa 2016; LaCourse 2016; Lawn 2017). See Appendix 10; Appendix 11 and Table 2.

LF-LAM accuracy stratified by CD4 ≤ 200 cells per μL and > 200 cells per μL

Two studies evaluated LF-LAM in unselected participants with CD4 ≤ 200 cells per μL, all settings. Sensitivity and specificity were 45% and 93% for Bjerrum 2015, and 7% and 99% for Hanifa 2016, (Appendix 10). Pooled sensitivity and specificity were 26% (9% to 56%) and 96% (87% to 98%) (706 participants, 12% with tuberculosis) (Table 2).

For unselected outpatients with a CD4 ≤ 200 cells per μL, two studies contributed data (652 participants; 10% with tuberculosis). Sensitivity and specificity were 36% and 94% for Bjerrum 2015, and 7% and 99% for Hanifa 2016 (Appendix 10). Pooled sensitivity and specificity were 21% (8% to 48%) and 96% (89% to 99%) (Table 2).

When we limited the analysis to unselected inpatients with CD4 ≤ 200 cells per μL, only one study (Bjerrum 2015) contributed data and found a sensitivity of 64% (95% CI: 35% to 87%) and specificity of 82% (67% to 93%) (54 participants, 26% with tuberculosis).

For comparison, we assessed diagnostic accuracy among participants with CD4 > 200 cells per µL, all settings. Only one study reported data for participants with CD4 > 200 cells per µL and reported a sensitivity of 27% (95% CI: 6% to 61%) and specificity of 99% (95% CI; 96% to 100%) (Bjerrum 2015).

LF-LAM accuracy stratified by CD4 ≤ 100 cells per µL and > 100 cells per µL

Three studies evaluated patients with CD4 ≤ 100 cells per µL, all settings, and sensitivity estimates ranged from 37% to 55% and specificity from 80% to 98% (Appendix 11). The pooled sensitivity and specificity (95% CrI) among participants with CD4 ≤ 100 cells per µL were 47% (30% to 64%) and 90% (77% to 96%) (3 studies, 417 participants, 31% with tuberculosis) (Table 2).

When we limited the analysis to studies involving unselected inpatients with CD4 ≤ 100 cells per µL two studies contributed data, (200 participants; 42% with tuberculosis). Sensitivity and specificity were 60% and 80% for Bjerrum 2015, and 55% and 98% for Lawn 2017 (Appendix 11). The pooled sensitivity and specificity were 57% (33% to 79%) and 90% (69% to 97%) (Table 2).

Two studies contributed data for unselected outpatients with CD4 ≤ 100 cells per µL (Bjerrum 2015; Hanifa 2016) (Appendix 11). The pooled sensitivity and specificity were 40% (20% to 64%) and 87% (68% to 94%) (217 participants, 21% with tuberculosis) (Table 2).

Four studies evaluated LF-LAM in participants with CD4 > 100 cells per µL (Bjerrum 2015; Drain 2015a; LaCourse 2016; Lawn 2017). Sensitivity estimates ranged from 0% to 33% and specificity estimates ranged from 95% to 99% (Appendix 11). Pooled sensitivity and specificity (95% CrI) were 20% (10% to 35%) and 98% (95% to 99%), (952 participants, 12% with tuberculosis) (Table 2).

LF-LAM accuracy among unselected participants, stratified by tuberculosis prevalence

The median prevalence of tuberculosis in studies with unselected participants was 10% (IQR 9% to 17%). In the analysis by tuberculosis prevalence, we found that pooled sensitivity and specificity for unselected participants in settings with tuberculosis prevalence of 10% or more were 45% (31% to 61%) and 92% (79% to 97%) (4 studies) compared to 16% (5% to 36%) and 98% (94% to 99%) in settings with tuberculosis prevalence less than 10% (3 studies). In general, tuberculosis prevalence increased in studies with a higher proportion of symptomatic participants.

Sensitivity analyses

As for studies among symptomatic participants, we undertook sensitivity analyses by limiting inclusion in the meta-analysis to the following.

- Studies with low risk of bias for patient selection (three studies)
- Studies with low risk of bias in the reference standard domain (two studies)
- Studies that used fresh urine samples (five studies) rather than stored urine sample

We further excluded the two studies with more than 80% of participants being symptomatic at inclusion (Bjerrum 2015; Lawn 2017).

These sensitivity analyses made little difference to any of the findings (Table 3).

DISCUSSION

This systematic review updates the current literature and includes 15 unique studies on the accuracy of lateral flow urine lipoarabinomannan assay (LF-LAM, Alere Determine™ TB LAM Ag) for tuberculosis in adults living with HIV and integrates nine new studies identified since the original WHO guidelines (WHO Lipoarabinomannan Policy Guidance 2015) and Cochrane Review (Shah 2016). A priori, we chose to evaluate performance of LF-LAM separately for studies that enrolled strictly symptomatic participants and studies that enrolled unselected participants irrespective of signs and/or symptoms of tuberculosis, and we conducted investigations of heterogeneity based on CD4 counts and study setting (inpatient/outpatient). Eight studies used LF-LAM for tuberculosis diagnosis in symptomatic adult participants with signs and symptoms suggestive of tuberculosis. These studies largely focused on inpatient settings and had high tuberculosis prevalence. Seven studies evaluated LF-LAM for diagnosing tuberculosis in unselected participants who may or may not have had signs or symptoms suggestive of tuberculosis when enrolled in the study. The studies with unselected participant were conducted predominantly in outpatient settings and, compared to studies with exclusively symptomatic participants, had lower tuberculosis prevalence and involved patients with higher CD4 counts; the proportion of symptomatic participants in these studies ranged from 19% to 90%. All studies were conducted in low- and middle-income countries with a high tuberculosis/HIV burden, and only one study outside sub-Saharan Africa.

Summary of main results

We have summarized the main findings in Summary of findings 1; Summary of findings 2. The main findings of this updated review include the following.

For tuberculosis diagnosis in HIV-positive adults presenting with signs and symptoms of tuberculosis, the diagnostic accuracy of LF-LAM is (95% credible interval (CrI)):

- In all settings, sensitivity 42% (31% to 55%) and specificity 91% (85% to 95%);
- In inpatient settings, sensitivity 52% (40% to 64%) and specificity 87% (78% to 93%);
- In outpatient settings, sensitivity 29% (17% to 47%) and specificity 96% (91% to 99%).

For tuberculosis diagnosis in HIV-positive adults irrespective of signs and symptoms of tuberculosis, the diagnostic accuracy of LF-LAM is:

- In all settings, sensitivity 35% (22% to 50%) and specificity 95% (89% to 98%);
- In inpatient settings, sensitivity 62% (41% to 83%) and specificity 84% (48% to 96%);
- In outpatient settings, sensitivity 31% (18% to 47%) and specificity 95% (87% to 99%).

We note that the band intensity of grade 1 in this review corresponds to the current manufacturer threshold for positivity (equivalent to that of grade 2 on the old manufacturer reference

card) and all results were evaluated against a microbiological reference standard.

LF-LAM for tuberculosis diagnosis in studies with symptomatic participants

Summary of findings 1.

In planned investigations of heterogeneity, we found an inverse correlation between LF-LAM sensitivity and CD4 count, with increasing sensitivity as patient CD4 count decreased (increased from 16% (CrI 8% to 31%) in patients with CD4 cell count > 200 cells per μL to 24% (CrI 14% to 38%) in patients with CD4 cell count between 101 to 199 cells per μL, to 54% (CrI 38% to 69%) in patients with CD4 ≤ 100 cells per μL). Similarly, we a priori planned to investigate and expected to find higher sensitivity in patients who were hospitalized while specificity decreased.

Results of these studies indicate that in theory, for a population of 1000 people where 300 have microbiologically-confirmed tuberculosis, the utilization of LF-LAM would result in: 189 to be LF-LAM positive: of these, 63 (33%) would not have tuberculosis (false-positives); and 811 to be LF-LAM negative: of these, 174 (21%) would have tuberculosis (false-negatives).

LF-LAM for tuberculosis diagnosis in studies with unselected participants (signs and symptoms not assessed as a condition of study inclusion)

Summary of findings 2.

In the investigations of heterogeneity, we expected and found a higher sensitivity in patients with CD4 ≤ 100 cells per μL (47%) and among inpatients (62%) compared to patients with CD4 > 100 cells per μL (20%) among outpatients (31%) respectively, though data to inform subgroup analyses were limited. We noted that participants included in the studies with unselected participants often presented with sign and symptoms suggestive of tuberculosis (a positive WHO tuberculosis screen), and in the studies evaluating inpatients most participants (> 80%) were in fact presenting with signs and symptoms suggestive of tuberculosis. These studies may be considered more similar to studies with exclusively symptomatic participants. In additional analysis of heterogeneity, we examined diagnostic accuracy based on tuberculosis prevalence within the studied cohort, as an alternative surrogate to presence of symptoms or CD4 count as an assessment of pre-test probability. We found that pooled sensitivity was 45% when tuberculosis prevalence within the study population was ≥ 10%, compared to only 16% when tuberculosis prevalence in the study population was < 10%.

Results of these studies indicate that in theory, for a population of 1000 people where 100 have microbiologically-confirmed tuberculosis, the utilization of LF-LAM would result in: 80 to be LF-LAM positive: of these, 45 (56%) would not have tuberculosis (false-positives); and 920 to be LF-LAM negative: of these, 65 (7%) would have tuberculosis (false-negatives).

LF-LAM for tuberculosis diagnosis, overall

The findings of this updated review are consistent with those of the original review (Shah 2016; WHO Lipoarabinomannan Policy Guidance 2015). Inclusion of additional studies in this updated review provided the basis for a more precise estimate of the

LF-LAM overall sensitivity and specificity. It further allowed us to address key questions regarding test accuracy and sources of heterogeneity including clinical setting and CD4 cell count in studies with symptomatic individuals and in studies with unselected participants.

Overall, we found lower sensitivity for diagnosis of tuberculosis among people living with HIV than the internationally suggested target of minimum 65% overall for rapid non-sputum tuberculosis tests (WHO TTP 2014). We found that sensitivity increased when considering inpatients and individuals with lower CD4 counts, whether considering studies with exclusively symptomatic participants or those with unselected participants. The reasons for this heterogeneity remain unclear, and several hypotheses have been postulated including differences in bacterial burden, sites of tuberculosis disease, or degree of glomerular dysfunction that may be different across different settings or CD4 counts.

When restricting analysis to studies that included participants unable to produce a sputum sample, the estimates of sensitivity increased. Sputum-scarce patients may be the potential target population to benefit the most from urine-based testing as they cannot have other sputum-based diagnostic testing and are likely to have a high yield of urine lipoarabinomannan (LAM) test positivity (Sabur 2017). However, only a few studies included participants who could not provide sputum samples for diagnostic testing. To the extent that inability to produce sputum is correlated with severity of tuberculosis disease and/or LAM positivity, this approach to participant selection could have lowered sensitivity estimates within these studies. Sensitivity analysis further revealed a higher sensitivity among studies evaluating LF-LAM on fresh non-stored urine samples without it affecting specificity. However, no study has made a direct comparison of performance on fresh versus frozen/stored urine samples and the significance of this is unclear.

Overall, we found that the estimated specificity approached the recommended targets for non-site-specific, non-sputum based test (WHO TTP 2014), although lower specificity was found among inpatients and those with advanced immunosuppression compared to outpatients and those with higher CD4 counts. We expected that, if restricting the analysis to studies using a higher quality reference standard (e.g. inclusion of more than one specimen type), that estimates of specificity would increase, but had limited data to conduct such a sensitivity analysis. When evaluating an assay that is not site-specific, future studies may consider systematic sampling of multiple sites to improve the quality of the reference standard. Future studies should include systematic sampling of sputum, blood, and urine for mycobacterial culture and/or nucleic acid amplification test (NAAT) testing from all participants as common sites of tuberculosis that may be identified by urinary LAM antigen detection; additional site-specific samples (e.g. cerebrospinal fluid (CSF), tissue biopsies) should also be collected for culture and/or NAAT based on clinical presentation.

We decided a priori to evaluate performance of LF-LAM in HIV-positive individuals with signs and symptoms of tuberculosis (symptomatic) separately from HIV-positive individuals irrespective of signs and symptoms of tuberculosis (unselected participants). We considered evaluating LF-LAM performance among specifically asymptomatic (i.e. exclusively those without symptoms) participants to assess the use of LF-LAM for the purpose of tuberculosis screening or community-based active case-finding, but such data were lacking among included

studies. We did find that several studies among unselected participants reported that a high proportion of study participants had signs and symptoms of tuberculosis, suggesting relative similarities to studies that enrolled exclusively symptomatic participants. Consequently, the overall performance of LF-LAM among asymptomatic patients remains largely unknown.

The overall differences in pooled estimates of sensitivity and specificity between studies of symptomatic versus unselected participants may have been attributable to differences in study setting and relative degree of immunosuppression of included participants, rather than type of population (i.e. unselected versus symptomatic participants). When examining inpatients, the pooled estimates for sensitivity were 52% (40% to 64%) and 62% (41% to 83%), when comparing studies of symptomatic participants and those including unselected participants. Among outpatients, the pooled sensitivity was 29% (17% to 47%) compared to 31% (18% to 47%) among studies of symptomatic participants and unselected participants, respectively. This indicates that other characteristics than lower CD4 may explain the higher sensitivity among inpatients like higher tuberculosis prevalence, higher mycobacterial burden, renal or genitourinary tract tuberculosis with LAM secretion in urine.

Overall, our findings suggest that the diagnostic accuracy of LF-LAM may vary by study setting, CD4 count, and tuberculosis prevalence among the target population. The authors hypothesize that these attributes (inpatients, low CD4 counts, or high tuberculosis prevalence) may collectively be surrogate indicators of participants with advanced tuberculosis disease or higher bacillary burden and LAM antigenuria in whom LF-LAM may aid in the diagnosis of tuberculosis, including both pulmonary and extrapulmonary tuberculosis. 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 these finding in that inpatients, those with low CD4, or cohorts with higher tuberculosis prevalence are likely to have higher disease severity or higher bacillary burden. While the test does not identify all tuberculosis cases, our findings suggest that it may be of particular value in diagnosing tuberculosis among patients with increased disease severity. Other factors that may be considered in evaluating LF-LAM may include ability to perform the test on individuals unable to produce sputum who cannot be diagnosed with other tuberculosis diagnostic tests, and ability to implement the test at the point-of-care with non-invasive specimen collection (WHO TTP 2014).

We identified three published studies of LF-LAM in children as the result of a broader search for studies in adults and children using the same inclusion criteria (Kroidl 2015; LaCourse 2018a; Nicol 2014). All three studies took place in countries in Africa with a high tuberculosis/HIV burden. Prevalence of microbiologically-confirmed tuberculosis in the studies was 7% in LaCourse 2018a, 22% in Nicol 2014; and 40% in Kroidl 2015. Kroidl 2015 enrolled children six weeks to 14 years, median age (interquartile range (IQR) 6.8 years (3.9 to 9.5) for all participants, including HIV-positive and HIV-negative children. LaCourse 2018a enrolled children aged 12 years or less, median age (IQR) 24 months (13 to 58). Nicol 2014 enrolled children aged 15 years or less, median age (IQR) 42.5 months (19.1 to 66.3) for all participants, including HIV-positive and HIV-negative children. A table summarizing the characteristics of studies in children can be found in Appendix 12. Kroidl 2015 and Nicol 2014 involved HIV-positive children with tuberculosis symptoms. LaCourse 2018a involved HIV-positive children hospitalized for acute illness irrespective of tuberculosis signs and symptoms. Kroidl 2015 was conducted in an outpatient setting, LaCourse 2018a in an inpatient setting, and Nicol 2014 in both an inpatient and an outpatient setting (Appendix 12). Given the differences in population and setting, we did not perform meta-analyses and report sensitivity and specificity estimates for the individual studies, all measured against a microbiological reference standard. Sensitivity and specificity (95% CI) were 42% (15% to 72%) and 94% (73% to 100%), (30 participants) (Kroidl 2015); 56% (21% to 86%) and 95% (90% to 98%), (130 participants) (LaCourse 2018a); and 43% (23% to 66%) and 80% (69% to 88%), (106 participants) (Nicol 2014). We found limited evidence on the accuracy of LF-LAM in children living with HIV. There were too few studies and participants to draw conclusions.

We acknowledge that patient outcomes are clearly important to patients, decision makers, and the wider tuberculosis community. While, the primary focus of this Cochrane Review update was to evaluate the diagnostic accuracy of LF-LAM, in the process of performing the review, we also assessed available data on LF-LAM and patient-important outcomes including mortality. We identified two randomized controlled trials (Gupta-Wright, 2018a; Peter 2016) that included data on the impact of LF-LAM implementation on mortality and other patient outcomes. Peter 2016 enrolled participants with tuberculosis symptoms, and Gupta-Wright, 2018a enrolled unselected participants irrespective of symptoms. Both trials were conducted in multiple countries in sub-Saharan Africa. Both trials involved hospitalized, HIV-positive patients, used the results of LF-LAM to guide therapy, and assessed all-cause mortality at eight weeks. Peter 2016 found that mortality was significantly reduced in participants receiving LF-LAM compared to those receiving routine diagnostic tests alone (no LF-LAM); adjusted relative risk reduction of 0.83 (95% CI: 0.73 to 0.96, P = 0.012). In Gupta-Wright, 2018a, the use of rapid urine screening with LF-LAM and urine Xpert MTB/RIF (intervention group) was not associated with a statistically significant reduction in mortality compared with routine diagnostic tests alone (standard-of-care group, no LF-LAM); adjusted risk reduction (aRD) of -2.8% (95% CI -5.8 to 0.3, P = 0.074). However, of importance, Gupta-Wright, 2018a, demonstrated that mortality was significantly lower in the intervention group than in the standard-of-care group for three of 12 prespecified subgroups: CD4 counts less than 100 cells per μL (aRD -7.1% (95% CI -13.7 to -0.4), P = 0.036); severe anaemia (aRD -9.0% (95% CI -16.6 to -1.3), P = 0.021); and participants with clinically suspected tuberculosis (aRD -5.7% (95% CI -10.9 to -0.5), P = 0.033). We will address the impact of testing with LF-LAM on patient outcomes in greater depth in another Cochrane Review (a Cochrane Protocol is underway).

Concerning economic evaluations, we are aware of six studies, all from settings in sub-Saharan Africa (Boyles 2018; Mukora 2018; Orlando 2018; Reddy 2019; Shah 2013; Sun 2013). These study methods and populations were heterogeneous, assessing a range of diagnostic algorithms, and only four of the total studies assessed cost-effectiveness (Orlando 2018; Reddy 2019; Shah 2013; Sun 2013), while Mukora 2018 performed a detailed costing study, including both at clinical and above clinic level costs, and Boyles 2018 assessed whether the addition of LF-LAM in diagnostic algorithms could reduce total program costs. studies employed a range of different willingness to pay thresholds, all four studies

were consistent in concluding LF-LAM could be cost-effective in a population of African adults living with HIV (particularly amongst hospitalized patients). Models found cost-effectiveness of LF-LAM to be robust across a variety of sensitivity analyses, and across different country settings and scenarios investigated. Key parameters that are likely most influential on cost-effectiveness include: tuberculosis prevalence/target population, and LF-LAM specificity, cost of treating tuberculosis and HIV and life expectancy post tuberculosis survival. Detailed empirical costing data are limited for LF-LAM implementation, however one such study published in 2018 (Mukora 2018) estimates unit test costs for LF-LAM implementation several fold higher (~$23-24) than most current models ($2-4), where costs of implementation and staff time are not considered. Underestimation of LF-LAM unit costs could result in overly optimistic cost-effectiveness profiles. However, tuberculosis diagnostics costs likely represent just a small proportion of total costs, indeed inclusion of costs associated with antiretroviral treatment (ART) and HIV care resulted in higher incremental cost-effectiveness ratios. While current evidence is consistent in suggesting LF-LAM is likely cost-effective among HIV-positive patients in sub-Saharan Africa, caution should be used when extrapolating from a small number of studies, and additional evidence from a wider range of populations, settings and diagnostic approaches will be necessary. A systematic review of economic evaluations is underway.

Differences from original Cochrane Review

In comparison to the original Cochrane Review (Shah 2016), this updated review includes 15 published studies (nine new studies since the original review and six studies from the original review). The original review included data from 12 studies, of which six were excluded as they did not meet the eligibility criteria for this updated review. Three used an older threshold for determining test positivity that is no longer recommended and three were abstracts, of which one (Lawn 2014) was added in this review as an updated published manuscript. For comparison, the main findings from the original review (Shah 2016) are listed in Table 4.

In the evaluation of diagnostic accuracy among symptomatic participants, the pooled estimates for sensitivity (42% [CrI 31% to 55%] versus 45% [CrI 29% to 63%]) and specificity (91% [CrI 85% to 95%] versus 92% [CrI 80% to 97%]) remained similar, comparing the current review and prior review, respectively. When stratified by setting, the pooled estimates among inpatients for sensitivity (52% [CrI 40% to 64%] versus 53% [CrI 38% to 70%)]) and specificity (87% [CrI 78% to 93%] versus 90% [CrI 73% to 96%]) did not change substantially when comparing the current review and the prior review, respectively. Pooled estimates among outpatients at the current manufacturer threshold for positivity were not previously available.

In Shah 2016, some studies were classified as 'tuberculosis (TB) screening' if they included participants irrespective of symptoms (i.e. with or without symptoms). Recognizing that these studies may have included a large proportion of symptomatic participants, these studies have been more clearly labelled as studies among 'unselected participants' in the current review. In the prior review, there were insufficient data to perform meta-analysis among unselected participants at the currently recommended manufacturer threshold for test positivity. There were previously insufficient data to investigate heterogeneity due to study setting or CD4 count. In the current review, we report on diagnostic

accuracy among inpatients and outpatients and by CD4. In both the current and prior review, data to assess diagnostic accuracy among asymptomatic participants (without signs or symptoms of tuberculosis) were unavailable.

This updated review did not assess diagnostic accuracy at an older threshold for determining test positivity (grade 1 out of 5, on an older reference card); data on diagnostic accuracy for this threshold can be found in the prior review (Shah 2016). Similarly, in this updated review we did not evaluate accuracy against a composite reference standard, results of which were included in the prior review (Shah 2016). Finally, given the relative lack of data on evaluating the incremental yield of LF-LAM in combination with sputum smear microscopy and Xpert MTB/RIF, we did not include these analyses in the updated review, but a summary of available data is found in the prior review (Shah 2016).

We note that the band intensity of grade 1 in this review corresponds to the current manufacturer threshold for positivity (equivalent to that of grade 2 on the old manufacturer reference card) and all results were evaluated against a microbiological reference standard.

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 people living with HIV when applied to different enrolment criteria - accuracy among individuals with signs and symptoms of tuberculosis, and accuracy when used without assessment of signs and symptoms of tuberculosis (i.e. patients may or may not have had symptoms). This updated review included new studies published since the original review. However, we found considerable heterogeneity across studies with respect to clinical setting, CD4 count and tuberculosis prevalence. For some analyses, few studies and participants contributed data and results should, therefore, be interpreted with caution. The review was further limited by the number of studies that used a lower quality reference standard and the high risk of selection bias in several studies owing to exclusion of patients unable to produce sputum.

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 correspondence with study authors to obtain additional and unpublished data on study characteristics. 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. We only included studies with a microbiological reference standard, which is considered the best currently available reference standard for tuberculosis. We included studies that evaluated LF-LAM for diagnosis of pulmonary tuberculosis, extrapulmonary tuberculosis, or both pulmonary and extrapulmonary tuberculosis. However, we recognize that a substantial number of tuberculosis cases may not be verified by

microbiological testing if only sputum is tested and when patients with advanced HIV are assessed. We acknowledge difficulties in diagnosing HIV-associated tuberculosis with extrapulmonary and disseminated forms of disease and considered a standardized reference standard using two or more specimen types consistently for all participants 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 tuberculosis. A lower quality reference standard may miss some tuberculosis cases and classify some tuberculosis patients as not having tuberculosis. This may make a truly positive LF-LAM result seem like a false-positive result leading to an underestimation of specificity. In this review, only one study (12%) for tuberculosis diagnosis among symptomatic participants and two studies (29%) for tuberculosis diagnosis among unselected participants used a higher quality reference standard for all participants.

We did not assess performance against a composite reference standard that uses microbiological or clinical information to classify tuberculosis. This was done in the original WHO guideline (WHO Lipoarabinomannan Policy Guidance 2015), and Cochrane Review ((Shah 2016), but found little impact on pooled estimates of sensitivity and specificity relative to performance measured against a microbiological reference standard.

We could not determine whether heterogeneity in specificity estimates was fully attributable to misclassification bias. Some studies (Nel 2017; Qvist 2014) have postulated that infection with (disseminated) nontuberculous mycobacteria (NTM) may also result in false-positive results, although this hypothesis is still questioned (Gupta-Wright 2018). Only one study (Thit 2017) was conducted outside of sub-Saharan Africa, and was noted to report the lowest specificity estimates of all included studies. Reasons for potential false-positive results remain unclear and it is unknown if differences in the epidemiology of disseminated NTM and other opportunistic infections across settings could contribute to variation in specificity.

Quality and quality of reporting of the included studies

There were few studies that included participants unable to expectorate sputum and few studies that included extrapulmonary specimens in addition to 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 studies among symptomatic participants, in the patient selection domain, we considered six studies (75%) to be at high risk of bias, and in the reference standard domain, we considered seven studies (88%) to be at high risk of bias. For studies with unselected participants, in the patient selection domain, we considered four studies (57%) to be at high risk of bias and in the reference standard domain, we considered five studies (71%) to be at high risk of bias. Risk of bias was considered low for the index test and flow and timing domains.

Using the GRADE approach, we judged the evidence for LF-LAM sensitivity to be moderate. This means that 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. We judged the evidence for LF-LAM specificity to be low or very low. This means that we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

In general, studies were fairly well reported, though we corresponded with some of the primary study authors to ask for clarification mainly in relation to interpretation of test positivity.

Applicability of findings to the review question

Overall, we had low concern about the applicability of the included studies to our review question as assessed by Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). However, studies of HIV-positive adults irrespective of signs and symptoms had low representation of asymptomatic individuals, as most participants in fact presented with signs and symptoms of tuberculosis. To date, no study has evaluated LF-LAM in an asymptomatic population. This review evaluated LF-LAM at the current manufacturer-recommended cut-off for positivity. Therefore, the findings 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. All studies were performed in low- or middle-income countries, but only a single study was conducted outside sub-Saharan Africa. It is therefore possible that the LF-LAM would perform differently in settings outside Africa. For the reference test domain, most studies had low concern for applicability.

AUTHORS' CONCLUSIONS

Implications for practice

For people living with HIV, this review found overall lower sensitivity of the lateral flow urine lipoarabinomannan assay (LF-LAM) than the internationally suggested target of minimum 65% overall for non-sputum based tuberculosis diagnostic tests (WHO TTP 2014). This was consistent whether the test is used for diagnosis of HIV-associated tuberculosis (without consideration of sites of disease involvement) among symptomatic (sensitivity of 42%) or unselected (i.e. not assessed for symptoms) participants (sensitivity of 35%). The estimated sensitivity suggests that if LF-LAM were to be used alone, more than half of all tuberculosis cases would be missed. Whether diagnostic accuracy differs for specific sites of tuberculosis is unknown (e.g. pulmonary tuberculosis, tuberculosis meningitis, or lymph node tuberculosis). There was a lack of data to evaluate diagnostic accuracy for active tuberculosis in asymptomatic participants, but studies among unselected outpatients that included a combination of symptomatic and asymptomatic participants (i.e. unselected participants) showed lower sensitivity (31%).

Despite the estimated sensitivity, two randomized controlled trials implementing LF-LAM in high-prevalence settings in sub-Saharan Africa have demonstrated reduced mortality and impact on other clinical outcomes when used to guide tuberculosis treatment in hospitalized HIV-positive adults.

LF-LAM is being considered as a diagnostic test that may be used as an add on to clinical judgement and with other tests for the diagnosis of HIV-associated tuberculosis. The test does not have a role as a replacement or triage test. The test may improve important health outcomes among HIV-positive hospitalized patients because it is feasible (i.e. can be done on urine which is easy to obtain in an inpatient) and may yield a positive result in those that are

most likely to die from tuberculosis and are most likely to present atypically (i.e. patients with advanced HIV). Whether these results may be extrapolated to outpatients with advanced HIV disease or those that are seriously ill is unknown.

Findings suggest that sensitivity increases with lower CD4 counts and in inpatient settings compared to outpatient settings. For unselected participants irrespective of symptoms, we found differences in LF-LAM performance based on tuberculosis prevalence of the target population (when stratified at greater or less than 10%), with higher sensitivity when the study population had higher tuberculosis prevalence (≥ 10%). The difference in diagnostic accuracy of LF-LAM by study setting and by degree of immunosuppression was consistent regardless of the approach to patient selection. Taking all findings into account, we think LF-LAM has a role along with other tests among individuals with tuberculosis symptoms or with low CD4 cell count. However, we had limited data and these findings should be interpreted with caution.

The test does not require sputum collection and is not site-specific. Other favourable test characteristics include low-cost, rapidity (< one hour), ease of use (does not require extensive sample preparation), and the fact that the test does not require electricity or special instruments and equipment (WHO TTP 2014). As a simple point-of-care test that does not depend upon sputum evaluation, LF-LAM testing may be the only possible way to confirm a diagnosis when a sputum sample cannot be produced.

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 tuberculosis transmission. The consequences of false-positive results are delayed alternative diagnosis, likelihood of anxiety and morbidity caused by additional testing, unnecessary treatment, and possible adverse events; the possible stigma associated with a diagnosis of tuberculosis. As LF-LAM does not offer information about drug resistance, a culture- or molecular-based diagnosis should be attempted to enable drug susceptibility testing to avoid that patients with unidentified drug-resistant tuberculosis may be inappropriately treated with a regimen appropriate only for drug-susceptible disease.

Ultimately, the use of LF-LAM in a clinical algorithm will not be based on accuracy alone, but also on considerations of cost-effectiveness, user perspectives and preferences, and the overall benefits and harms of the test. This updated review contributed accuracy data for the WHO Guideline Development Group Meeting in May 2019. The updated WHO guidelines on the use of Alere LAM are expected later in 2019.

Implications for research

Future studies that evaluate the diagnostic accuracy of non-sputum-based tests for tuberculosis, such as LF-LAM, in people living with HIV should use a reference standard that includes at least two different specimens (e.g. sputum, and urine), and in addition, for presumed extrapulmonary tuberculosis, appropriate specimens from the suspected sites of involvement. Moreover, future studies should include patients unable to expectorate sputum in the analysis as these are most likely to benefit from the test. Characteristics that possibly drive performance of the test in addition to CD4 status, such as burden of disease, renal co-morbidities or other should be evaluated. More studies should evaluate LF-LAM accuracy in HIV-positive children, as currently there are very limited data available (Kroidl 2015; LaCourse 2018a; Nicol 2014). While some studies enrolled unselected participants, our review suggests that a large proportion were symptomatic, particularly in the inpatient setting. Understanding whether there is value in using LF-LAM for diagnosis of tuberculosis among HIV-positive people who do not report any tuberculosis symptoms (particularly for those with CD4 < 100) remains unknown. Further studies among asymptomatic individuals are needed to better define the scope of recommendations on the use of LF-LAM. The indication of increased sensitivity with use of fresh urine and also morning urine needs further investigations, and studies in settings outside sub-Saharan Africa are lacking. 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

REFERENCES

References to studies included in this review

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 2015a *{published data only}*

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 2016 *{published data only}*

Drain PK, Gounder L, Sahid F, Moosa MY. Rapid urine LAM testing improves diagnosis of expectorated smear-negative pulmonary tuberculosis in an HIV-endemic region. *Scientific Reports* 2016;**6**:19992.

Floridia 2017 *{published data only}*

Floridia M, Ciccacci F, Andreotti M, Hassane A, Sidumo Z, Magid NA, et al. Tuberculosis case finding with combined rapid point-of-care assays (Xpert MTB/RIF and Determine TB LAM) in HIV-positive individuals starting antiretroviral therapy in Mozambique. *Clinical Infectious Diseases* 2017;**65**(11):1878-83.

Hanifa 2016 *{published data only}*

Hanifa Y, Fielding K L, Chihota V N, Adonis L, Charalambous S, Karstaedt A, et al. Diagnostic accuracy of lateral flow urine LAM assay for TB screening of adults with advanced immunosuppression attending routine HIV care in South Africa. *PLOS One* 2016;**11**(6):e0156866. [DOI: 10.1371/journal.pone.0156866]

Huerga 2017 *{published data only}*

Huerga H, Ferlazzo G, Bevilacqua P, Kirubi B, Ardizzoni E, Wanjala S, et al. Incremental yield of including Determine-TB LAM assay in diagnostic algorithms for hospitalized and ambulatory HIV-positive patients in Kenya. *PLOS One* 2017;**12**(1):e0170976. [DOI: 10.1371/journal.pone.0170976]

Juma 2017 *{published data only}*

Juma FO, Yonga G, Waweru P, Revathi G, Nelson M, Shah R. The diagnostic accuracy of Determine TB LAM antigen in detection of extra pulmonary tuberculosis (EPTB). *Journal of Infection* 2017;**75**(6):583-6.

LaCourse 2016 *{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 2017 *{published data only}*

Lawn SD, Kerkhoff AD, Burton R, Schutz C, Boulle A, Vogt M, et al. Diagnostic accuracy, incremental yield and prognostic value of Determine TB-LAM for routine diagnostic testing for tuberculosis in HIV-infected patients requiring acute hospital admission in South Africa: a prospective cohort. *BMC Medicine* 2017;**15**(1):67.

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.

Pandie 2016 *{published data only}*

Pandie S, Peter JG, Kerbelker ZS, Meldau R, Theron G, Govender U, et al. The diagnostic accuracy of pericardial and urinary lipoarabinomannan (LAM) assays in patients with suspected tuberculous pericarditis. *Scientific Reports* 2016;**6**:32924. [DOI: 10.1038/srep32924]

Peter 2012a *{published data only}*

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 data only}*

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.

Peter 2016 *{published data only}*

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.

Thit 2017 *{published data only}*

Thit SS, Aung NM, Htet ZW, Boyd MA, Saw HA, Anstey NM, et al. The clinical utility of the urine-based lateral flow lipoarabinomannan assay in HIV-infected adults in Myanmar: an observational study. *BMC Medicine* 2017;**15**(1):145.

References to studies excluded from this review

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 AB, 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.

Andrews 2014 *{unpublished data only}*

Andrews B, Muchemwa L, Lakhi S, Bwalya M, Mabula C, Chipii G, et al. 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 data only}*

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.

Bisson 2016 *{published data only}*

Bisson GP, Gupta A, Miyahara S, Sung X, Bao J, Fry CL, et al. Urine LAM testing in advanced HIV-infected adults in a trial of empiric TB therapy. Topics in Antiviral Medicine. 2016; Vol. 24, issue E-1:312-13.

Blanc 2018 *{published data only}*

Blanc FX, Badje AD, Bonnet M, Gabillard D, Messou E, Muzoora C, et al. Systematic vs test-guided tuberculosis treatment: Data of the statis randomized trial. Topics in Antiviral Medicine. 2018; Vol. 26, issue Supplement 1:13s-14s.

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.

Broughton 2017 *{published data only}*

Broughton E, Haumba S, Calnan M, Ginindsa S, Jeffries R, Maphalala G, et al. Screening in maternity to ascertain tuberculosis status (SMATS) study. *BMC Infectious Diseases* 2017;**17**(1):191. [DOI: 10.1186/s12879-017-2285-0]

Calligaro 2017 *{published data only}*

Calligaro Gl, Zijenah LS, Peter JG, Theron G, Buser V, McNerney R, et al. Effect of new tuberculosis diagnostic technologies on community-based intensified case finding: a multicentre randomised controlled trial. *Lancet Infectious Diseases* 2017;**17**(4):441-50. [DOI: 10.1016/S1473-3099(16)30384-X]

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 SN, Lee JH, Lee HY, Won HJ, 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.

Cox 2015 *{published data only}*

Cox JA, Lukande RL, Kalungi S, Van Marck E, Van de Vijver K, Kambugu A, et al. Is Urinary Lipoarabinomannan the result of renal tuberculosis? assessment of the renal histology in an autopsy cohort of Ugandan HIV-infected adults. *PLOS One* 2015;**10**(4):e0123323. [DOI: 10.1371/journal.pone.0123323]

d'Elia 2015 *{published data only}*

d'Elia A, Evans D, McNamara L, Berhanu R, Sanne I, Lonnermark E. Predictive and prognostic properties of TB-LAM among HIV-positive patients initiating ART in Johannesburg, South Africa. *Pan African Medical Journal* 2015;**22**:4.

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.

Drain 2014a *{published data only}*

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 *{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.

Drain 2015b *{published data only}*

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. [DOI: http://dx.doi.org/10.1136/bmjopen-2014-006833]

Drain 2016a *{published data only}*

Drain PK, Losina E, Coleman SM, Giddy J, Ross D, Katz JN, et al. Rapid urine lipoarabinomannan assay as a clinic-based screening test for active tuberculosis at HIV diagnosis. *BMC Pulmonary Medicine* 2016;**16**(1):147. [DOI: 10.1186/s12890-016-0316-z]

Drain 2017 *{published data only}*

Drain PK, Losina E, Coleman SM, Giddy J, Ross D, Katz JN, et al. Clinic-based urinary lipoarabinomannan as a biomarker of clinical disease severity and mortality among antiretroviral therapy-naive human immunodeficiency virus-infected adults in South Africa. *Open Forum Infectious Diseases* 2017;**4**(3):ofx167. [DOI: 10.1093/ofid/ofx167]

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.

Gina 2017 *{published data only}*

Gina P, Randall PJ, Muchinga TE, Pooran A, Meldau R, Peter JG, et al. Early morning urine collection to improve urinary lateral flow LAM assay sensitivity in hospitalised patients with HIV-TB co-infection. *BMC Infectious Diseases* 2017;**17**(1):339. [DOI: 10.1186/s12879-017-2313-0]

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.

Grant 2016 *{published data only}*

Grant A, Charalambous S, Tlali M, Johnson S, Dorman S, Hoffmann C, et al. Empirical TB treatment in advanced HIV disease: Results of the TB fast track trial. Topics in Antiviral Medicine. 2016; Vol. 24, issue E-1:61-62.

Gupta-Wright 2016a *{published data only}*

Gupta-Wright A, Peters JA, Flach C, Lawn SD. Detection of lipoarabinomannan (LAM) in urine is an independent predictor of mortality risk in patients receiving treatment for HIV-associated tuberculosis in sub-Saharan Africa: a systematic review and meta-analysis. *BMC Medicine* 2016;**14**:53. [DOI: 10.1186/s12916-016-0603-9]

Gupta-Wright 2016b *{published data only}*

Gupta-Wright A, Fielding Kl, Oosterhout JJ, Wilson DK, Corbett El, Flach C, et al. Rapid urine-based screening for tuberculosis to reduce AIDS-related mortality in hospitalized patients in Africa (the STAMP trial): study protocol for a randomised controlled trial. *BMC Infectious Diseases* 2016;**16**(1):501. [DOI: 10.1186/s12879-016-1837-z]

Gupta-Wright 2018b *{published data only}*

Gupta-Wright A, Corbett EL, Van Oosterhout JJ, Wilson DK, Grint, D, Alufandika-Moyo M, et al. Urine-based screening for tuberculosis: A randomized trial in HIV-positive inpatients. Topics in Antiviral Medicine. 2018; Vol. 26, issue Supplement 1:18s-19s.

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.

Hanifa 2015 *{published data only}*

Hanifa Y, Telisinghe L, Fielding KL, Malden Jl, Churchyard GJ, Grant AD, et al. The diagnostic accuracy of urine lipoarabinomannan test for tuberculosis screening in a South African correctional facility. *PLOS One* 2015;**10**(5):e0127956. [DOI: 10.1371/journal.pone.0127956]

Iskandar 2017 *{published data only}*

Iskandar A, Nursiloningrum E, Arthamin MZ, Olivianto E, Chandrakusuma MS. The diagnostic value of urine lipoarabinomannan (LAM) antigen in childhood tuberculosis. *Journal of Clinical and Diagnostic Research* 2017;**11**(3):Ec32-35. [DOI: 10.7860/jcdr/2017/20909.9542]

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.

Kerkhoff 2017 *{published data only}*

Kerkhoff AD, Barr DA, Schutz C, Burton R, Nicol MP, Lawn SD, et al. Disseminated tuberculosis among hospitalised HIV patients in South Africa: a common condition that can be rapidly diagnosed using urine-based assays. *Scientific Reports* 2017;**7**(1):10931. [DOI: 10.1038/s41598-017-09895-7]

Koralnik 2017 *{published data only}*

Koralnik IJ, Love S, Mubanga E, Mukondya S, Atadzhanov M, Kosloff B, et al. Optimizing rapid CSF diagnostics of tuberculous meningitis in Zambia. Annals of Neurology. 2017; Vol. 82, issue Supplement 21:S74. [DOI: http://dx.doi.org/10.1002/ana.25024]

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.

Kroidl 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.

Kroidl 2015 *{published data only}*

Kroidl I, Clowes P, Reither K, Mtafya B, Rojas-Ponce G, Ntinginya EN, et al. Performance of urine lipoarabinomannan assays for paediatric tuberculosis in Tanzania. *European Respiratory Journal* 2015;**46**(3):761-70. [DOI: 10.1183/09031936.00003315]

LaCourse 2018a *{published data only}*

Lacourse SM, Pavlinac PB, Cranmer LM, Njuguna IN, Mugo C, Gatimu J, et al. Stool Xpert MTB/RIF and urine lipoarabinomannan for the diagnosis of tuberculosis in hospitalized HIV-infected children. *AIDS* 2018;**32**(1):69-78. [DOI: 10.1097/QAD.0000000000001662]

LaCourse 2018b *{published data only}*

LaCourse SM, Cranmer LM, Njuguna IN, Gatimu J, Stern J, Maleche-Obimbo E, et al. Urine TB lipoarabinomannan (LAM) predicts mortality in hospitalized HIV-infected children. *Clinical Infectious Diseases* 2018;**66**(11):1798-801. [DOI: 10.1093/cid/ciy011]

Lawn 2009a *{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 2012a *{published data only}*

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 2012b *{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 2013a *{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 2014 *{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.

Leopold 2017 *{published data only}*

Leopold B, Samuel N, Aimable D, Joannes C. New diagnostic tests for tuberculosis: Performance of LAM and Xpert MTB/RIF in urine of hospitalized patients on intensive phase of TB treatment, and its association with TB dissemination and HIV status. Preliminary results from an observational cohort study in Kigali, Rwanda. Tropical Medicine and International Health. 2017; Vol. 22, issue Supplement 1:35. [DOI: http://dx.doi.org/10.1111/%28ISSN%291365-3156]

Love 2016 *{published data only}*

Love S, Mubanga E, Munkondya S, Atadzhanov M, Kosloff B, Ayles H, et al. Optimizing CSF diagnostics of tuberculous meningitis in Zambia. Neurology. 2016; Vol. 86, issue 16 SUPPL. 1.

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.

Mathabire 2017 *{published data only}*

Mathabire SC, Cossa L, Mpunga J, Manhica I, Quiles IA, Molfino L, et al. Feasibility of using Determine-TB LAM test in HIV-infected adults in programmatic conditions. Journal of the International Aids Society. 2017; Vol. 20:18.

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* 2012;**92**(5):407-16.

Musarurwa 2018 *{published data only}*

Musarurwa C, Zijenah LS, Mhandire DZ, Bandason T, Mhandire K, Chipiti MM, et al. Higher serum 25-hydroxyvitamin D concentrations are associated with active pulmonary tuberculosis in hospitalised HIV infected patients in a low income tropical setting: a cross sectional study. *BMC Pulmonary Medicine* 2018;**18**(1):67. [DOI: 10.1186/s12890-018-0640-6]

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.

Nakiyingi 2015 *{published data only}*

Nakiyingi L, Ssengooba W, Nakanjako D, Armstrong D, Holshouser M, Kirenga BJ, et al. Predictors and outcomes of mycobacteremia among HIV-infected smear- negative presumptive tuberculosis patients in Uganda. *BMC Infectious Diseases* 2015;**15**:62. [DOI: 10.1186/s12879-015-0812-4]

Nakiyingi 2015a *{published data only}*

Nakiyingi L, Nonyane BA, Ssengooba W, Kirenga BJ, Nakanjako D, Lubega G, et al. Predictors for MTB culture-positivity among HIV-infected smear-negative presumptive tuberculosis patients in Uganda: Application of new tuberculosis diagnostic technology. *PLOS One* 2015;**10**(7):e0133756. [DOI: 10.1371/journal.pone.0133756]

Nicol 2014 *{published data only}*

Nicol MP, Allen V, Workman L, Isaacs W, Munro J, Pienaar S, et al. Urine lipoarabinomannan testing for diagnosis of pulmonary tuberculosis in children: a prospective study. *Lancet Global Health* 2014;**2**(5):e278–84.

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, Kasprowicz 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.

Reddy 2018 *{published data only}*

Reddy KP, Gupta-Wright A, Fielding K, Costantini S, Zheng A, Corbett EL, et al. Cost-effectiveness of urine TB screening for hospitalized people with HIV in Africa. Topics in Antiviral Medicine. 2018; Vol. 26, issue Supplement 1:518s.

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. 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 Infectious Diseases* 2009;**9**:141.

Sabur 2017 *{published data only}*

Sabur NF, Esmail A, Brar MS, Dheda K. Diagnosing tuberculosis in hospitalized HIV-infected individuals who cannot produce sputum: is urine lipoarabinomannan testing the answer?. *BMC Infectious Diseases* 2017;**17**(1):803. [DOI: 10.1186/s12879-017-2914-7]

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.

Sahle 2017 *{published data only}*

Sahle SN, Asress DT, Tullu KD, Weldemariam AG, Tola HH, Awas YA, et al. Performance of point-of-care urine test in diagnosing tuberculosis suspects with and without HIV infection in selected peripheral health settings of Addis Ababa, Ethiopia. *BMC Research Notes* 2017;**10**(1):74. [DOI: 10.1186/s13104-017-2404-4]

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.

Suwanpimolkul 2017 *{published data only}*

Suwanpimolkul G, Kawkitinarong K, Manosuthi W, Sophonphan J, Gatechompol S, Ohata PJ, et al. Utility of urine lipoarabinomannan (LAM) in diagnosing tuberculosis and predicting mortality with and without HIV: prospective TB cohort from the Thailand big city TB research network. *International Journal of Infectious Diseases* 2017;**59**:96-102. [DOI: 10.1016/j.ijid.2017.04.017]

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 Infectious Diseases* 2012;**12**:47.

Zijenah 2016 *{published data only}*

Zijenah LS, Kadzirange G, Bandason T, Chipiti MM, Gwambiwa B, Makoga F, et al. Comparative performance characteristics of the urine lipoarabinomannan strip test and sputum smear microscopy in hospitalized HIV-infected patients with suspected tuberculosis in Harare, Zimbabwe. BMC Infectious Diseases 2016; Vol. 16:20. [10.1186/s12879-016-1339-z]

Additional references

Alere 2017

Abbott. Alere DetermineTM TB LAM Ag Product Information. alere.com/en/home/product-details/determine-tb-lam.html (accessed July 18 2017:IN02740004 Rev. 07 2017/08.

Balshem 2011

Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 2011;**64(4)**:401-6.

Batz 2011

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

Boyles 2018

Boyles TH, Griesel R, Stewart A, Mendelson M, Maartens G. Incremental yield and cost of urine Determine TB-LAM and sputum induction in seriously ill adults with HIV. *International Journal of Infectious Diseases* 2018;**75**:67-73.

Brennan 2003

Brennan PJ. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis (Edinb)* 2003;**83**(1-3):91-7.

Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

Briken 2004

Briken V, Porcelli SA, Besra GS, Kremer L. Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. *Molecular Microbiology* 2004;**53**(2):391-403.

Broger 2019

Broger T, Sossen B, du Toit E, Kerkhoff AD, Schutz C, Reipold EI, et al. Novel lipoarabinomannan point-of-care tuberculosis test for people living with HIV: a diagnostic accuracy study. *Lancet Infectious Diseases* 2019;**19**(8):852-61. [DOI: 10.1016/S1473-3099(19)30001-5]

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 2006

Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006;**59**(12):1331–2.

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.

Covidence 2017 [Computer program]

Veritas Health Innovation. Covidence systematic review software. Available at www.covidence.org. Melbourne: Veritas Health Innovation, 2017.

FIND 2018

Foundation for Innovative New Diagnostics (FIND). Call for trial partners now open to advance evaluation of an innovative, sensitive point-of-care test that detects tuberculosis in HIV-positive people. finddx.org/publication/pr-26sep18/ (accessed 14 June 2019).

Flores 2005

Flores LL, Pai M, Colford JM Jr, Riley LW. In-house nucleic acid amplification tests for the detection of Mycobacterium tuberculosis in sputum specimens: meta-analysis and meta-regression. *BMC Microbiology* 2005;**5**:55.

Ford 2016

Ford N, Matteelli A, Shubber Z, Hermans S, Meintjes G, Grinsztejn B, et al. TB as a cause of hospitalization and in-hospital mortality among people living with HIV worldwide: a systematic review and meta-analysis. *Journal of the International AIDS Society* 2016;**19**(1):20714.

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 2012

Gupta A, Wood R, Kaplan R, Bekker LG, Lawn SD. Tuberculosis incidence rates during 8 years of follow-up of an antiretroviral treatment cohort in South Africa: comparison with rates in the community. *PLOS One* 2012;**7**(3):e34156.

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.

Gupta-Wright 2018

Gupta-Wright A, Kerkhoff A D, Meintjes G, Corbett EL. Urinary lipoarabinomannan detection and disseminated nontuberculous mycobacterial disease. *Clinical Infectious Diseases* 2018;**66**(1):158.

Gupta-Wright, 2018a

Gupta-Wright A, Corbett EL, van Oosterhout JJ, Wilson D, Grint D, Alufandika-Moyo M, et al. Rapid urine-based screening for tuberculosis in HIV-positive patients admitted to hospital in Africa (STAMP): a pragmatic, multicentre, parallel-group, double-blind, randomised controlled trial. *Lancet* 2018;**392**:292–301.

Harris 2009

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009;**42**(2):377-81.

Horne 2019

Horne DJ, Kohli M, Zifodya JS, Schiller I, Dendukuri N, Tollefson D, et al. Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* 2019, Issue 6. [DOI: 10.1002/14651858.CD009593.pub4]

Kohli 2018

Kohli M, Schiller I, Dendukuri N, Dheda K, Denkinger CM, Schumacher SG, et al. Xpert® MTB/RIF assay for extrapulmonary tuberculosis and rifampicin resistance. *Cochrane*

Lawn 2009

Lawn SD, Myer L, Edwards D, Bekker LG, Wood R. Short-term and long-term risk of tuberculosis associated with CD4 cell recovery during antiretroviral therapy in South Africa. *AIDS* 2009;**23**(13):1717-25.

Lawn 2012

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**(1):103.

Lawn 2013

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.

Lawn 2016

Lawn SD, Gupta-Wright A. Detection of lipoarabinomannan (LAM) in urine is indicative of disseminated TB with renal involvement in patients living with HIV and advanced immunodeficiency: evidence and implications. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 2016;**110**(3):180-5.

Lewinsohn 2017

Lewinsohn DM, Leonard MK, LoBue PA, Cohn DL, Daley CL, Desmond E, et al. Official american thoracic society/Infectious diseases society of America/Centers for disease control and prevention clinical practice guidelines: Diagnosis of tuberculosis in adults and children. *Clinical Infectious Diseases* 2017;**64**(2):111-5.

Lunn 2009

Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine* 2009;**28**:3049-67.

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://srdta.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.

Mukora 2018

Mukora R, Tlali M, Monkwe S, Charalambous S, Karat AS, Fielding KL, Grant AD, Vassall A. Cost of point-of-care lateral flow urine lipoarabinomannan antigen testing in HIV-positive adults in South Africa. *International Journal of Tuberculosis and Lung Disease* 2018;**22**(9):1082-7.

Nel 2017

Nel JS, Lippincott CK, Berhanu R, Spencer DC, Sanne IM, Ive P. Does disseminated nontuberculous mycobacterial disease cause false-positive Determine TB-LAM lateral flow assay results? A retrospective review. *Clinincal Infectious Diseases* 2017;**65**(7):1226-8.

Orlando 2018

Orlando S, Triulzi I, Ciccacci F, Palla I, Palombi L, Marazzi MC, et al. Delayed diagnosis and treatment of tuberculosis in HIV + patients in Mozambique: A cost-effectiveness analysis of screening protocols based on four symptom screening, smear microscopy, urine LAM test and Xpert MTB/RIF. *PLOS One* 2018;**13**(7):e0200523.

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.

Pai 2016

Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. *Nature Reviews Disease Primers* 2016;**2**:16076.

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.

Qvist 2014

Qvist T, Johansen IS, Pressler T, Hoiby N, Andersen A B, Katzenstein TL, et al. Urine lipoarabinomannan point-of-care testing in patients affected by pulmonary nontuberculous mycobacteria - experiences from the Danish Cystic Fibrosis cohort study. *BMC Infectious Diseases* 2014;**14**:655.

R Statistical Computing 2018 [Computer program]

R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2018.

Reddy 2019

Reddy KP, Gupta-Wright A, Fielding KL, Costantini S, Zheng A, Corbett EL, et al. Cost-effectiveness of urine-based tuberculosis screening in hospitalised patients with HIV in Africa: a

microsimulation modelling study. *Lancet Global Health* 2019;**7**(2):e200-8.

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 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: 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 2016

Schünemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G, et al. GRADE Working Group. GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *Journal of Clinical Epidemiology* 2016 August;**76**:89-98.

SDU Open

Odense Patient data Explorative Network. OPEN. sdu.dk/en/Om_SDU/Institutter_centre/Klinisk_institut/Forskning/Forskningsenheder/open.aspx (accessed 10 June 2018).

Shivakoti 2017

Shivakoti R, Sharma D, Mamoon G, Pham K. Association of HIV infection with extrapulmonary tuberculosis: a systematic review. *Infection* 2017;**45**(1):11-21.

Spiegelhalter 2004

Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health Care Evaluation. New York: Wiley, 2004.

StataCorp 2017 [Computer program]

StataCorp. Stata Statistical Software. Version 13. Texas: StataCorp LP, 2017.

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.

TB CARE I 2014

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

Unitaid 2017

D Boyle. Tuberculosis Diagnostics Technology and Market Landscape. 5th Edition. Vernier: World Health Organization Unitaid Secretariat, 2017.

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.

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 ART Guidelines 2016

World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. 2nd Edition. Geneva: World Health Organization, 2016.

WHO Global Tuberculosis Report 2018

World Health Organization. Global Tuberculosis Report 2018. Licence: CC BY-NC-SA 3.0 IGO. Geneva: World Health Organization, 2018.

WHO ICF 2011

World Health Organization Stop TB Dept Dept of HIV/AIDS. Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings. Geneva: World Health Organization, 2011.

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 Managing Advanced HIV Disease 2017

World Health Organization. Guidelines for managing advanced HIV disease and rapid initiation of antiretroviral therapy. Geneva: World Health Organization, 2017.

WHO TB-LAMP 2016

World Health Organization. The use of Loop-mediated Isothermal Amplification (TB-LAMP) for the diagnosis of pulmonary tuberculosis: Policy guidance. Geneva: World Health Organization, 2016.

WHO TTP 2014

World Health Organization. High-priority target product profiles for new tuberculosis diagnostics, Report of a consensus meeting. Geneva: World Health Organization, 2014:98.

WHO Xpert Ultra 2017

World Health Organization. WHO meeting report of a technical expert consultation: non-inferiority analysis of Xpert MTB/RIF Ultra compared to Xpert MTB/RIF. Geneva: World Health Organization, 2017.

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. Geneva: World Health Organization, 2013.

World Bank 2018

World Bank. World Bank Country Classifications by Income. Washington (District of Columbia): World Bank, 2018.

Yuan 2014

Yuan LY, Li Y, Wang M, Ke ZQ, Xu WZ. Rapid and effective diagnosis of pulmonary tuberculosis with novel and sensitive loop-mediated isothermal amplification (LAMP) assay in clinical samples: a meta-analysis. *Journal of Infection and Chemotherapy* 2014;**20**(2):86-92.

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]

Shah 2016

Shah M, Hanrahan C, Wang ZY, Dendukuri N, Lawn SD, Denkinger CM, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database of Systematic Reviews* 2016, Issue 5. [DOI: 10.1002/14651858.CD011420.pub2]

CHARACTERISTICS OF STUDIES

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

Bjerrum 2015

Study characteristics	
Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults included irrespective of symptoms. A positive WHO symptom screen was reported for 91% of participants.
	Age: mean 39 (SD10)
	Sex, female: 64%
	HIV infection: 100%
	Median CD4 cell count per µL: 127 (IQR 35 to 256)
	History of tuberculosis: 6%
	Sample size: 469
	Clinical setting: outpatient and inpatients
	Country: Ghana
	Tuberculosis incidence rate: 152 per 100,000
	Number (proportion) of tuberculosis cases in the study: 55 (12%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis
	Reference standard: any of sputum mycobacterial culture (solid and liquid) or nucleic acid amplification test; no extrapulmonary specimens tested
Flow and timing	All patients were included in the analysis
Comparative	

Bjerrum 2015 *(Continued)*

Notes	The study authors did not attempt sputum induction and excluded individuals without some respiratory specimen available. There was no attempt to identify extrapulmonary tuberculosis.

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?	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 2015a

Study characteristics

Drain 2015a *(Continued)*

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults included irrespective of symptoms. Proportion of participants that reported with tuberculosis signs and symptoms symptomatic not stated.
	Age: mean 33 (SD 9)
	Sex, female: 49%
	HIV infection: 100%
	Median CD4 cell count per μL: 248 (IQR 107 to 379)
	History of tuberculosis: 9%
	Sample size: 320
	Clinical setting: outpatient
	Country: South Africa
	Tuberculosis incidence rate: 567 per 100,000
	Number (proportion) of tuberculosis cases in the study: 54 (17%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis
	Reference standard: mycobacterial culture of sputum (solid); no extrapulmonary specimens tested
Flow and timing	All patients were included in the analysis
Comparative	
Notes	The study authors performed sputum induction for individuals who could not expectorate, excluded individuals without some respiratory specimen available. They did not attempt to identify extrapulmonary tuberculosis.

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			

Drain 2015a *(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

Drain 2016

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected and HIV-uninfected adults suspected of tuberculosis included if smear microscopy negative x 2; Presenting with at least two of four tuberculosis-related symptoms: cough, fever, weight loss, night sweats for at least two weeks and a Karnofsky performance score >50 Age: mean 37 (SD 9) Sex, female: 49% HIV infection: 93% Median CD4 cell count per µL: 168 (IQR 89 to 256) History of tuberculosis: not stated Sample size: 90 Clinical setting: outpatient Country: South Africa Tuberculosis incidence rate: 567 per 100,000 Number (proportion) of tuberculosis cases in the study: 57 (63%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target: pulmonary tuberculosis Reference standard: any of sputum mycobacterial culture (solid and liquid) or nucleic acid amplification test; no extrapulmonary specimens tested

---

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)

Drain 2016 (Continued)

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 tuberculosis

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	High
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	

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)

45

Floridia 2017

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults included irrespective of symptoms. A positive WHO symptom screen was reported for 34% of participants. Age: median 35 (IQR 30-43) Sex, female: 59% HIV infection: 100% Median CD4 cell count per µL: 278 (IQR 142 to 395) History of tuberculosis: not stated Sample size: 972 Clinical setting: outpatients Country: Mozambique Tuberculosis incidence rate: 551 per 100,000 Number (proportion) of tuberculosis cases in the study: 90 (9%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target: pulmonary tuberculosis Reference standard: nucleic acid amplification test; no extrapulmonary specimens tested
Flow and timing	All patients were included in the analysis
Comparative	
Notes	The study authors did not attempt sputum induction and excluded individuals without some respiratory specimen available. They did not attempt to identify extrapulmonary tuberculosis.

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		

Floridia 2017 *(Continued)*

		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?	Unclear

		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

Hanifa 2016

Study characteristics	
Patient sampling	Cohort, prospective and consecutive enrolment with six months follow-up
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults included irrespective of symptoms. A positive WHO symptom screen was reported for 53% of participants.
	Age: median 39 (IQR 32-45)
	Sex, female: 61%
	HIV infection: 100%
	Median CD4 cell count per µL: 111 (IQR 56 to 161)
	History of tuberculosis: 30%
	Sample size: 424
	Clinical setting: outpatients
	Country: South Africa
	Tuberculosis incidence rate: 567 per 100,000
	Number (proportion) of tuberculosis cases in the study: 40(9%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis, extrapulmonary tuberculosis, mycobacteraemia
	Reference standard: any of mycobacterial culture (liquid) or nucleic acid amplification test; extrapulmonary specimens tested for all
Flow and timing	The study excluded 16 cases defined as 'Clinical tuberculosis' from analysis.
Comparative	

Hanifa 2016 *(Continued)*

Notes	Sputum and non-sputum samples were tested for mycobacteria for all participants; The study authors performed sputum induction for individuals who could not expectorate; The study had specimens collected for index test and reference standard tests greater than seven days apart.

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?	Unclear		
		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?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Huerga 2017

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults suspected of tuberculosis; Presenting with cough for more than 2 weeks or any cough and at least any of loss of weight, night sweats, or fever Age: median 35 (IQR:29-44) Sex, female: 52% HIV infection: 100% Median CD4 cell count per µL: 109 (IQR 43 to 214) History of tuberculosis: 25% Sample size: 474 (275 included in analysis) Clinical setting: outpatient and inpatients Country: Kenya Tuberculosis incidence rate: 319 per 100,000 Number (proportion) of tuberculosis cases in the study: 156 (57%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target: pulmonary and extrapulmonary tuberculosis. Reference standard: any of mycobacterial culture (solid and liquid) or nucleic acid amplification test. Extrapulmonary specimens collected for participants without respiratory samples available
Flow and timing	The study excluded participants with at least one sample with no result or contaminated result (culture) from analysis in the absence of a positive result.
Comparative	
Notes	The study authors performed sputum induction for individuals who could not expectorate. They included individuals without respiratory specimen available and attempted to identify extrapulmonary tuberculosis; Not all participants received the same reference standard and not all participants were included in the analysis.

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		

Huerga 2017 *(Continued)*

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?	Unclear		
		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?	No	
Were all patients included in the analysis?	No	
		High

Juma 2017

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected and HIV-uninfected adults (aged above 14 years) suspected of extrapulmonary tuberculosis. Age: 43 (SD: 14.8) Sex, female: 46% HIV infection: 68% Median CD4 cell count per µL: not stated History of tuberculosis: 6% Sample size: 99 (67 with HIV included) Clinical setting: inpatients Country: Kenya Tuberculosis incidence rate: 319 per 100,000 Number (proportion) of tuberculosis cases in the study: 29 (33%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target: extrapulmonary tuberculosis, any site Reference standard: any of mycobacterial culture (solid and liquid) or nucleic acid amplification test. No sputum specimens tested
Flow and timing	All patients were included in the analysis

Juma 2017 *(Continued)*

	Comparative
Notes	The study excluded participants with suspected pulmonary tuberculosis and the reference standard was restricted to extrapulmonary specimens

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		
		High	**High**
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?	No		
Were all patients included in the analysis?	Yes		
		Low	

LaCourse 2016

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected pregnant women attending antenatal care was included irrespective of symptoms. A positive WHO symptom screen was reported for 19% of participants.
	Other characteristics: pregnant women
	Age: median 25 (IQR 22 to 30)
	Sex, female: 100%
	HIV infection: 100%
	Median CD4 cell count per µL: 437 (IQR 342 to 565)
	History of tuberculosis: 9%
	Sample size: 266
	Clinical setting: outpatient
	Country: Kenya
	Tuberculosis incidence rate: 319 per 100,000
	Number (proportion) of tuberculosis cases in the study: 3 (1%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis
	Reference standard: expectorated sputum mycobacterial culture (liquid)
Flow and timing	All patients were included in the analysis
Comparative	
Notes	The study authors did not attempt sputum induction and excluded individuals without some respiratory specimen available. There was no attempt to identify extrapulmonary tuberculosis.

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			

LaCourse 2016 *(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	

Lawn 2017

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults included irrespective of symptoms. A positive WHO symptom screen was reported for 91% of participants Age: median 36 (IQR 29-42) Sex, female: 61% HIV infection: 100% Median CD4 cell count per µL: 149 (IQR 55 to 312) History of tuberculosis: 6% Sample size: 413 Clinical setting: inpatients Country: South Africa Tuberculosis incidence rate: 567 per 100,000 Number (proportion) of tuberculosis cases in the study: 139 (32.6%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target: pulmonary tuberculosis, extrapulmonary tuberculosis, mycobacteraemia Reference standard: any of mycobacterial culture (liquid) or nucleic acid amplification test; extrapulmonary specimens tested for all

---

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)

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 included individuals without respiratory specimen and attempted to identify extrapulmonary tuberculosis

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?	No		
Were all patients included in the analysis?	Yes		
		Low	

Nakiyingi 2014

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults suspected of tuberculosis; Presented with any of cough, fever, night sweats and weight loss
	Age: median 33 (IQR 29 to 37)
	Sex, female: 63%
	HIV infection: 100%
	Median CD4 cell count per µL: 152 (IQR 41 to 337)
	History of tuberculosis: not stated
	Sample size: 997
	Clinical setting: inpatient and outpatient
	Country: South Africa and Uganda
	Tuberculosis incidence rate: 567 per 100,000 (South Africa); 201per 100,000 (Uganda)
	Number (proportion) of tuberculosis cases in the study: 367 (37%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis, mycobacteraemia
	Reference standard: sputum mycobacterial culture (liquid and solid), mycobacterial blood cultures; extrapulmonary specimens tested for all
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. They attempted to identify extrapulmonary tuberculosis

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			

Nakiyingi 2014 *(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?	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

Pandie 2016

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected and HIV-uninfected adults suspected of pericardial tuberculosis. Presented with pericardial effusion Age: median 34 (IQR 29-42) Sex, female: 38% HIV infection: 74% Median CD4 cell count per µL: 139 (IQR 81 to 249) History of tuberculosis: not stated Sample size: 102 Clinical setting: inpatients Country: South Africa Tuberculosis incidence rate: 567 per 100,000 Number (proportion) of tuberculosis cases in the study: 74 (49%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: extrapulmonary tuberculosis (pericardial tuberculosis) and pulmonary tuberculosis

Pandie 2016 *(Continued)*

	Reference standard: mycobacterial culture (liquid). There was no protocol to ensure a minimum standard of sputum sampling and testing
Flow and timing	Not all patients were included in the analysis
Comparative	
Notes	The study authors excluded participants with disseminated tuberculosis if not having pericardial effusion, and excluded patients with pulmonary tuberculosis if no sign of pericardial effusion. Not all patients received the same reference standard. For some patients sputum induction was attempted and sputum examined for mycobacteria. Not all participants were included in the analysis and it is unclear who was excluded. The true negative values reported conflicts and provides uncertainty about the specificity reported.

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?	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		

Pandie 2016 *(Continued)*

Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
	High

Peter 2012a

Study characteristics

Patient sampling	Cross-sectional, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults suspected of tuberculosis; Presenting with signs and symptoms suggestive of tuberculosis, but these not specified
	Age: median 35 (IQR 29 to 40)
	Sex, female: 60%
	HIV infection: 100%
	Median CD4 cell count per µL: 90 (IQR 47 to 197)
	History of tuberculosis: 35%
	Sample size: 241
	Clinical setting: inpatient
	Country: South Africa
	Tuberculosis incidence rate: 567 per 100,000
	Number (proportion) of tuberculosis cases in the study: 116 (48%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis and extrapulmonary tuberculosis
	Reference standard: mycobacterial culture (liquid); there was no protocol to ensure a minimum standard of respiratory and extrapulmonary specimen testing
Flow and timing	All patients were included in the analysis
Comparative	
Notes	The study authors included all patients with symptoms of tuberculosis 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.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Peter 2012a *(Continued)*

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	**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, prospective and consecutive enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults suspected of tuberculosis; presenting with any of cough, night sweats, fever or weight loss Age: median 36 (IQR 30 to 41) Sex, female: 46%

	HIV infection: 100%
	Median CD4 cell count per µL: 210 (IQR 103 to 375)
	History of tuberculosis: 26%
	Sample size: 569
	Clinical setting: outpatient
	Country: South Africa, Tanzania, Zambia
	Tuberculosis incidence rate: 567 per 100,000 (South Africa); 269 per 100,000 (Tanzania); 361 per 100,000 (Zambia)
	Number (proportion) of tuberculosis cases in the study: 181 (32%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis
	Reference standard: sputum mycobacterial (liquid) culture
Flow and timing	All patients were included in the analysis
Comparative	
Notes	The study authors excluded individuals unable to expectorate sputum and did not attempt sputum induction. They did not attempt to identify extra-pulmonary tuberculosis .

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			

Peter 2015 *(Continued)*

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

Peter 2016

Study characteristics

Patient sampling	Randomized controlled trial, consecutive prospective enrolment
Patient characteristics and setting	Presenting signs and symptoms: HIV-infected adults suspected of tuberculosis; Presenting with any of cough, night sweats, fever, or weight loss; or Illness that necessitate admission
	Age: median 37 (IQR 30 to 44)
	Sex, female: 50%
	HIV infection: 100%
	Median CD4 cell count per µL: 81 (IQR 26 to 198)
	History of tuberculosis: 27%
	Sample size: 1172
	Clinical setting: inpatient
	Country: South Africa, Zimbabwe, Zambia
	Tuberculosis incidence rate: 567 per 100,000 (South Africa); 221 per 100,000 (Zimbabwe); 361 per 100,000 (Zambia)
	Number (proportion) of tuberculosis cases in the study: 342 (29%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis, extrapulmonary tuberculosis, mycobacteraemia
	Reference standard: any of mycobacterial culture (solid and liquid) or nucleic acid amplification test; extrapulmonary specimens collected for those participants where treating clinicians found it indicated

---

Peter 2016 *(Continued)*

Flow and timing	All patients were included in the analysis
Comparative	
Notes	Health providers selected the sites for testing based on their own clinical suspicion. High concern of applicability because of a lack of study directed extrapulmonary testing. Not all participants received the same reference standard as tuberculosis diagnostics performed at the discretion of the clinician and varied across sites.

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?	No		
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	

Thit 2017

Study characteristics	
Patient sampling	Cohort, prospective and consecutive enrolment with six months follow-up
Patient characteristics and setting	Presenting signs and symptoms: irrespective of symptoms. WHO tuberculosis symptom screen positive in 33%. in overall population who could produce sputum signs and symptoms suggestive of tuberculosis, but not specified Age: median 34 (IQR 30-41) Sex, female: 50% HIV infection: 100% Median CD4 cell count per µL: 270 (IQR 128 to 443) History of tuberculosis: not stated Sample size: 517 Clinical setting: outpatient and inpatients Country: Myanmar Tuberculosis incidence rate: 358 per 100.000 Number (proportion) of tuberculosis cases in the study: 54 (12%)
Index tests	LF-LAM
Target condition and reference standard(s)	Target condition: pulmonary tuberculosis Reference standard: any of sputum mycobacterial culture (solid and liquid) or nucleic acid amplification test; no extrapulmonary specimens tested. The study did not report if they speciated mycobacteria isolates
Flow and timing	The study had specimens collected for index and reference standard tests greater than seven days apart.
Comparative	
Notes	The study authors attempted sputum induction and excluded individuals without some respiratory specimen available. They did not attempt to identify extrapulmonary tuberculosis. The study did not report if they speciated mycobacteria isolates.

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			

Thit 2017 *(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?	Unclear		
		High	Unclear

DOMAIN 4: Flow and Timing

Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		High	

Abbreviations: LF-LAM: lateral flow urine lipoarabinomannan assay (Alere Determine™ TB lipoarabinomannan); **IQR:** interquartile range; **SD:** standard deviation; **WHO:** World Health Organization

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

Study	Reason for exclusion
Agha 2013	Index test not studied
Amoudy 1997	Index test not studied
Andrews 2014	Abstract
Balcha 2014	Did not use manufacturer's current recommended positivity threshold
Bisson 2016	Abstract
Blanc 2018	Abstract
Boehme 2005	Index test not studied
Boyles 2012	Index test not studied
Broughton 2017	Study protocol

Study	Reason for exclusion
Calligaro 2017	Insufficient data
Chan 2000	Index test not studied
Cho 1997	Index test not studied
Conesa-Botella 2007	Index test not studied
Cox 2015	Did not use manufacturer's current recommended positivity threshold
d'Elia 2015	Reference standard not satisfied
Daley 2009	Index test not studied
Deng 2011	Index test not studied
Dheda 2009	Index test not studied
Dheda 2010	Index test not studied
Drain 2014a	Did not use manufacturer's current recommended positivity threshold
Drain 2014b	Abstract
Drain 2015b	Impact study (no diagnostic accuracy data)
Drain 2016a	Did not use manufacturer's current recommended positivity threshold
Drain 2017	Impact study (no diagnostic accuracy data)
Elsawy 2012	Index test not studied
Gina 2017	Reference standard not satisfied
Gounder 2011	Index test not studied
Grant 2016	Abstract
Gupta-Wright 2016a	Not a diagnostic accuracy study
Gupta-Wright 2016b	Study protocol, study completed
Gupta-Wright 2018b	Abstract
Hamasur 2001	Index test not studied
Hanifa 2015	Index test not studied
Iskandar 2017	Paediatric population
Kashino 2008	Index test not studied
Kerkhoff 2014a	Insufficient or duplicate data
Kerkhoff 2014b	Insufficient or duplicate data

Study	Reason for exclusion
Kerkhoff 2017	Second reference (complementary data) to Lawn 2017
Koralnik 2017	Abstract
Koulchin 2007	Index test not studied
Kroidi 2014	Index test not studied
Kroidl 2015	Paediatric population
LaCourse 2018a	Paediatric population
LaCourse 2018b	Paediatric population
Lawn 2009a	Index test not studied
Lawn 2012a	Did not use manufacturer's current recommended positivity threshold
Lawn 2012b	Insufficient or duplicate data
Lawn 2013a	Insufficient or duplicate data
Lawn 2014	Abstract
Leopold 2017	Abstract
Love 2016	Abstract
Manabe 2014	Insufficient or duplicate data
Mathabire 2017	Abstract
Mukundan 2012	Index test not studied
Musarurwa 2018	Not a diagnostic accuracy study
Mutetwa 2009	Index test not studied
Nakiyingi 2015	Second reference (complementary data) to Nakiyingi 2014
Nakiyingi 2015a	Second reference (complementary data) to Nakiyingi 2014
Nicol 2014	Paediatric population
Patel 2009	Index test not studied
Patel 2010	Index test not studied
Peter 2011	Insufficient or duplicate data
Peter 2012b	Insufficient or duplicate data
Peter 2012c	Insufficient or duplicate data
Peter 2013	Insufficient or duplicate data

Study	Reason for exclusion
Reddy 2018	Abstract
Reid 2015	Insufficient data
Reither 2009	Index test not studied
Sabur 2017	Reference standard not satisfied
Sada 1992	Index test not studied
Sahle 2017	Did not use manufacturer's current recommended positivity threshold
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	Insufficient or duplicate data
Shah 2014	Insufficient or duplicate data
Singh 2011	Index test not studied
Sun 2013	Insufficient data
Suwanpimolkul 2017	Did not use manufacturer's current recommended positivity threshold
Tessema 2001	Index test not studied
Tessema 2002a	Index test not studied
Tessema 2002b	Index test not studied
Tlali 2014	Insufficient or duplicate data
Van Rie 2013	Insufficient or duplicate data
Wood 2012	Index test not studied
Zijenah 2016	Second reference (complementary data) to Peter 2016

D A T A

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

Table Tests.   Data tables by test

Test	No. of studies	No. of participants
1 Symptomatic adults, all settings	8	3449
2 Symptomatic adults, inpatients	6	2253
3 Symptomatic adults, outpatients	4	1196
4 Symptomatic adults, CD4 > 200, all settings	3	738
5 Symptomatic adults, CD4 ≤ 200, all settings	4	1825
6 Symptomatic adults, CD4 ≤ 200, inpatients	2	1009
7 Symptomatic adults, CD4 ≤ 200, outpatients	1	249
8 Symptomatic adults, CD4 > 100, all settings	4	1519
9 Symptomatic adults, CD4 ≤ 100, all settings	5	1251
10 Symptomatic adults, CD4 ≤ 100, inpatients	3	746
11 Symptomatic adults, CD4 ≤ 100, outpatients	1	121
12 Symptomatic adults, CD4 101-200, all settings	4	586
13 Symptomatic adults, CD4 101-200, inpatients	2	275
14 Symptomatic adults, CD4 101-200, outpatients	1	128
17 Unselected adults, all settings	7	3365
18 Unselected adults, inpatients	3	537
19 Unselected adults, outpatients	6	2828
20 Unselected adults, CD4 > 200, all settings	1	156
21 Unselected adults, CD4 ≤ 200, all settings	2	706
22 Unselected adults, CD4 ≤ 200, inpatients	1	54
23 Unselected adults, CD4 ≤ 200, outpatients	2	652
24 Unselected adults, CD4 > 100, all settings	4	952
25 Unselected adults, CD4 ≤ 100, all settings	3	417
26 Unselected adults, CD4 ≤ 100, inpatients	2	200
27 Unselected adults, CD4 ≤ 100, outpatients	2	217
28 Unselected adults, CD4 101-200, all settings	1	103
29 Unselected adults, CD4 101-200, inpatients	1	9

Test	No. of studies	No. of participants
30 Unselected adults, CD4 101-200, outpatients	1	94

Test 1. Symptomatic adults, all settings.

Test 2. Symptomatic adults, inpatients.

Test 3. Symptomatic adults, outpatients.

Test 4. Symptomatic adults, CD4 > 200, all settings.

Test 5. Symptomatic adults, CD4 ≤ 200, all settings.

Test 6. Symptomatic adults, CD4 ≤ 200, inpatients.

Test 7. Symptomatic adults, CD4 ≤ 200, outpatients.

Test 8. Symptomatic adults, CD4 > 100, all settings.

Test 9. Symptomatic adults, CD4 ≤ 100, all settings.

Test 10. Symptomatic adults, CD4 ≤ 100, inpatients.

Test 11. Symptomatic adults, CD4 ≤ 100, outpatients.

Test 12. Symptomatic adults, CD4 101-200, all settings.

Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

Test 13.   Symptomatic adults, CD4 101-200, inpatients.

Test 14.   Symptomatic adults, CD4 101-200, outpatients.

Test 17.   Unselected adults, all settings.

Test 18.   Unselected adults, inpatients.

Test 19.   Unselected adults, outpatients.

Test 20.   Unselected adults, CD4 > 200, all settings.

Test 21.   Unselected adults, CD4 ≤ 200, all settings.

Test 22.   Unselected adults, CD4 ≤ 200, inpatients.

Test 23.   Unselected adults, CD4 ≤ 200, outpatients.

Test 24.   Unselected adults, CD4 > 100, all settings.

Test 25.   Unselected adults, CD4 ≤ 100, all settings.

Test 26.   Unselected adults, CD4 ≤ 100, inpatients.

Test 27.   Unselected adults, CD4 ≤ 100, outpatients.

Test 28.   Unselected adults, CD4 101-200, all settings.

Test 29.   Unselected adults, CD4 101-200, inpatients.

Test 30.   Unselected adults, CD4 101-200, outpatients.

ADDITIONAL TABLES

Table 1. Summary characteristics of included studies

Study	Participants (% symptom)	Setting	Median CD4 cell count per µL (IQR)	TB prevalence % (n/N)	Did the study avoid inappropriate exclusion?	Specimens collected	High-quality reference standard[a]	Unique study characteristics
HIV-positive adults with signs and symptoms of TB								
Drain 2016	**Symptomatic:** Two of four TB-related symptoms (cough, fever weight loss, night sweat) for > 2 weeks; smear microscopy negative x 2	Outpatient	168 (89 to 256)	63% (57/90)	No	Pulmonary samples	No	Adults (> 18 years); HIV positive (93.2%); targeting a relatively well outpatient population; Karnofsky performance score > 50
Huerga 2017	**Symptomatic:** Cough > 2 weeks or any cough and one of weight loss, night sweats or fever; severely ill; CD4< 200 or BMI below 17	Outpatients (33%); Inpatients (67%)	109 (43 to 214)	57% (156/275)	No	Pulmonary samples; urine Xpert only for patients without sputum available	No	Adults (> 15 years); LAM-guided treatment; excluded many participants from analysis[b]
Juma 2017	**Symptomatic:** Suggestive of extrapulmonary TB, not specified	Inpatients	Not stated	33% (29/67)	No	Extrapulmonary samples only, no sputum samples	No	Adults (> 14 years), HIV-positive (68%); excluded patients with concomitants active pulmonary TB
Nakiyingi 2014	**Symptomatic:** Any of cough, fever weight loss, night sweat	Outpatients (45%); Inpatients (55%)	152 (41 to 337)	37% (367/997)	No	Pulmonary samples; Blood culture for all	Yes	Adults (> 18 years); multisite; large sample size.
Pandie 2016	**Symptomatic:** Presence of a pericardial effusion and suspected of pericardial TB	Inpatients	139 (81 to 249)	95% (36/38)	No	Extrapulmonary samples (pericardial effusion); pulmonary samples for some	No	Adults (> 18 years); HIV-positive (74%); excluded participants from analysis affecting specificity[c]

Table 1. Summary characteristics of included studies *(Continued)*

Peter 2012a	**Symptomatic:** Any of cough, fever weight loss, night sweat	Inpatients	90 (47 to 197)	48% (116/241)	Yes	Clinically relevant pulmonary samples; clinically relevant extrapulmonary samples. No study defined algorithm.	No	Adults (> 18 years); multisite; TB diagnostic work-up was not standardized but up to clinical judgements
Peter 2015	**Symptomatic:** Any of cough, fever weight loss, night sweat	Outpatients	210 (103 to 375)	32% (181/569)	No	Pulmonary samples	No	Adults (> 18 years); multisite; nested within a randomized, parallel-arm trial.
Peter 2016	**Symptomatic:** Any of cough, fever weight loss, night sweat	Inpatients	81 (26 to 198)	29% (342/1172)	Yes	Pulmonary samples; clinically relevant extrapulmonary samples. No study defined algorithm.	No	Adults (> 18 years); multisite; LAM arm of a randomized controlled trial.
HIV-positive adults irrespective of signs and symptoms of TB								
Bjerrum 2015	**Unselected:** 91% symptomatic	Outpatients (85%); Inpatients (15%)	127 (35 to 256)	12% (55/469)	No	Pulmonary samples	No	Adults (> 18 years); majority symptomatic.
Drain 2015a	**Unselected:** proportion symptomatic not stated	Outpatient	248 (107 to 379)	17% (54/320)	No	Pulmonary samples	No	Adults (> 18 years)
Floridia 2017	**Unselected:** 34% symptomatic	Outpatient	278 (142 to 395)	9% (90/972)	No	Pulmonary samples	No	Adults (> 15 years). LAM-guided treatment.
Hanifa 2016	**Unselected:** 53% symptomatic	Outpatient	111 (56 to 161)	9% (40/408)	Yes	Pulmonary samples; blood culture for all	Yes	Adults (> 18 years); CD4 < 200; reference standard included any sample taken within six months from enrolment.
LaCourse 2016	**Unselected:** 19% symptomatic	Outpatient	437 (342 to 565)	1% (3/266)	No	Pulmonary samples	No	Pregnant women (> 16 years) attending ANC; healthy popula-

Table 1.   Summary characteristics of included studies *(Continued)*

								tion; one person with CD4< 400; Few TB cases (n = 3).
Lawn 2017	**Unselected:** 91% symptomatic	Inpatients	149 (55-312)	33% (139/413)	Yes	Pulmonary samples; Blood culture for all; Clinically relevant extrapulmonary samples	Yes	Adults (> 18 years). Included many samples from different sites
Thit 2017	**Unselected:** 33% symptomatic	Outpatients (90%); Inpatients (10%)	270 (128 to 443)	10% (54/517)	Yes	Pulmonary samples	No	Conducted in Myanmar. Adults (median age 34). Reference standard included samples taken within six months from enrolment.

Abbreviations: LF-LAM: lateral flow urine lipoarabinomannan assay (Alere Determine™ TB lipoarabinomannan assay); **ANC:** antenatal clinic; **BMI:** body mass index; **IQR:** interquartile range; **TB:** tuberculosis; **Xpert:** Xpert MTB/RIF

[a]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 in all participants as part of a defined standardized study algorithm.

[b]Huerga 2017 excluded participants from analysis if missing Xpert results or culture contaminated for any of the samples in the absence of a positive result; overall samples size 474 (156 with TB); 275 included in analysis (156 with TB).

[c]Pandie 2016 excluded a large number of non-TB participants from analysis; Overall samples size 102 (36 with TB); 38 included for analysis (36 TB cases).

Table 2.   LF-LAM pooled sensitivity and specificity for TB diagnosis, by study population

Type of analysis	Symptomatic participants				Unselected participants			
	Studies (total participants)	**Participants with TB (%)**	**Pooled sensitivity (95% CrI)**	**Pooled specificity (95% CrI)**	**Studies (total participants)**	**Participants with TB (%)**	**Pooled sensitivity (95% CrI)**	**Pooled specificity (95% CrI)**
Overall accuracy	8 studies (3449)	1277 (37%)	42% (31% to 55%)	91% (85% to 95%)	7 studies (3365)	432 (13%)	35% (22% to 50%)	95% (89% to 98%)
By setting								
Inpatient	6 studies (2253)	868 (39%)	52% (40% to 64%)	87% (78% to 93%)	3 studies (537)	159 (30%)	62% (41% to 83%)	84% (48% to 96%)
Outpatient	4 studies	409	29%	96%	6 studies	273	31%	95%

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)

	(1196)	(34%)	(17% to 47%)	(91% to 99%)	(2828)	(10%)	(18% to 47%)	(87% to 99%)
By CD4 cell								
CD4 > 200	3 studies (738)	163 (22%)	16% (8% to 31%)	94% (81% to 97%)	1 study[a] (156)	11 (7%)	Not applicable	Not applicable
CD4 ≤ 200	4 studies (1825)	722 (40%)	45% (31% to 61%)	89% (77% to 94%)	2 studies (706)	82 (12%)	26% (9% to 56%)	96% (87% to 98%)
CD4 > 100	4 studies (1519)	425 (28%)	17% (10% to 27%)	95% (89% to 98%)	4 studies (952)	115 (12%)	20% (10% to 35%)	98% (95% to 99%)
CD4 ≤ 100	4 studies (1239)	512 (41%)	54% (38% to 69%)	88% (77% to 94%)	3 studies (417)	130 (31%)	47% (40% to 64%)	90% (77% to 96%)
CD4 101-199	4 studies (586)	210 (36%)	24% (14% to 38%)	90% (77 to 96)	1 study (103)[b]	13 (13%)	Not applicable	Not applicable
By CD4 and setting								
CD4 ≤ 200 inpatients	2 studies (1009)	348 (34%)	54% (34% to 73%)	80% (58% to 91%)	1 study[c] (54)	14 (26%)	Not applicable	Not applicable
CD4 ≤ 100 inpatients	2 studies (734)	270 (37%)	61% (40% to 78%)	81% (61% to 91%)	2 studies (200)	84 (42%)	57% (33% to 79%)	90% (69% to 97%)
CD4 101-199 inpatients	2 studies (275)	78 (28%)	32% (16% to 57%)	81% (55% to 92%)	1 study[d] (9)	4 (44%)	Not applicable	Not applicable
CD4 ≤ 200 outpatients	1 study[f] (249)	97 (39%)	Not applicable	Not applicable	2 studies (652)	68 (10%)	21% (8% to 48%)	96% (89% to 99%)
CD4 ≤ 100 outpatients	1 study[g] (121)	48 (40%)	Not applicable	Not applicable	2 studies (217)	46 (21%)	40% (20 to 64)	87% (68 to 94)

CD4 101-199 outpatients	1 study[h] (128)	51 (40%)	Not applicable	Not applicable	1 study[e] (94)	9 (10%)	Not applicable	Not applicable

Abbreviations: LF-LAM: lateral flow urine lipoarabinomannan assay (Alere Determine™ TB lipoarabinomannan assay); **Crl:** credible interval; **TB:** tuberculosis.

[a]Bjerrum 2015, Sensitivity 27% (6% to 61%); Specificity 99% (96% to 100%); [b]Bjerrum 2015, Sensitivity 38% (14% to 68%); Specificity 99% (94% to 100%); [c]Bjerrum 2015, Sensitivity 64% (35% to 87%); Specificity 82% (67% to 93%); [d]Bjerrum 2015, Sensitivity 75% (19% to 99%); Specificity 100% (48% to 100%); [e]Bjerrum 2015, Sensitivity 22% (3% to 60%); Specificity 99% (94% to 100%); [f]Peter 2015, Sensitivity 24% (16% to 33%); Specificity 94% (89% to 97%); [g]Peter 2015, Sensitivity 30% (18% to 46%); Specificity 93% (85% to 98%); [h]Peter 2015, Sensitivity 18% (8% to 31%); Specificity 95% (87% to 99%).

Table 3. Sensitivity analyses, LF-LAM

Type of analysis	Symptomatic participants			Unselected participants		
	Studies partici- pants	Pooled sensitivity (95% CrI)	Pooled specificity (95% CrI)	Studies partici- pants	Pooled sensitivity (95% CrI)	Pooled specificity (95% CrI)
Including studies at low risk of patient se- lection bias	2 studies 1413	48% (29% to 67%)	82% (61% to 92%)	3 studies 1338	39% (17% to 66%)	93% (69% to 98%)
Including studies at low risk of reference standard bias	1 study 997	Insufficient data	Insufficient data	2 studies 821	24% (8% to 53%)	99% (95% to 99%)
Including studies that used only fresh spec- imens	4 studies 2511	52% (38% to 68%)	91% (82% to 95%)	5 studies 2544	41% (26% to 56%)	93% (82% to 97%)
Re-classificiation of Bjerrum 2015; Lawn 2017 as studies among symptomatic par- ticipant (> 80% participants symptomatic)	10 studies 4331	42% (33% to 52%)	93% (88% to 96%)	5 studies 2483	31% (16% to 50%)	94% (84% to 98%)

Abbreviations: LF-LAM: lateral flow urine lipoarabinomannan assay (Alere Determine™ TB lipoarabinomannan assay); **CrI:** Credible Interval

Table 4. Original Alere LAM review, main findings

Type of analysis	Symptomatic participants			Unselected participants		
	Studies (total participants)	Pooled sensitivi- ty (95% CrI)	Pooled specifici- ty (95% CrI)	Studies (total participants)	Pooled sensi- tivity (95% CrI)	Pooled speci- ficity (95% CrI)
Overall ac- curacy	5 studies (2313)	45% (29% to 63%)	92% (80% to 97%)	3 studies[a] (1055)	30% (20% to 43%)	94% (86% to 97%)
By setting						
Inpatient	4 studies (1299)	53% (38% to 70%)	90 % (73% to 96%)	Insufficient data		
Outpatient	Insufficient data			Insufficient data		

Abbreviations: LF-LAM: lateral flow urine lipoarabinomannan assay (Alere Determine™ TB lipoarabinomannan assay); **CrI:** credible interval

Main findings from the original review on LF-LAM accuracy (Shah 2016).

[a]Reported at grade 1 on the previous reference scale card with five bands that is outdated and no longer the recommended positivity threshold.

APPENDICES

Appendix 1. Alere Determine TB LAM Ag test

Figure 10.

Figure 10. (A) Alere™ TB LAM Ag test. 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) Updated Reference Scale Card accompanying test strips to 'grade' the test result and Alere positivity. Copyright © [2019] [Abbott Inc]: reproduced with permission.
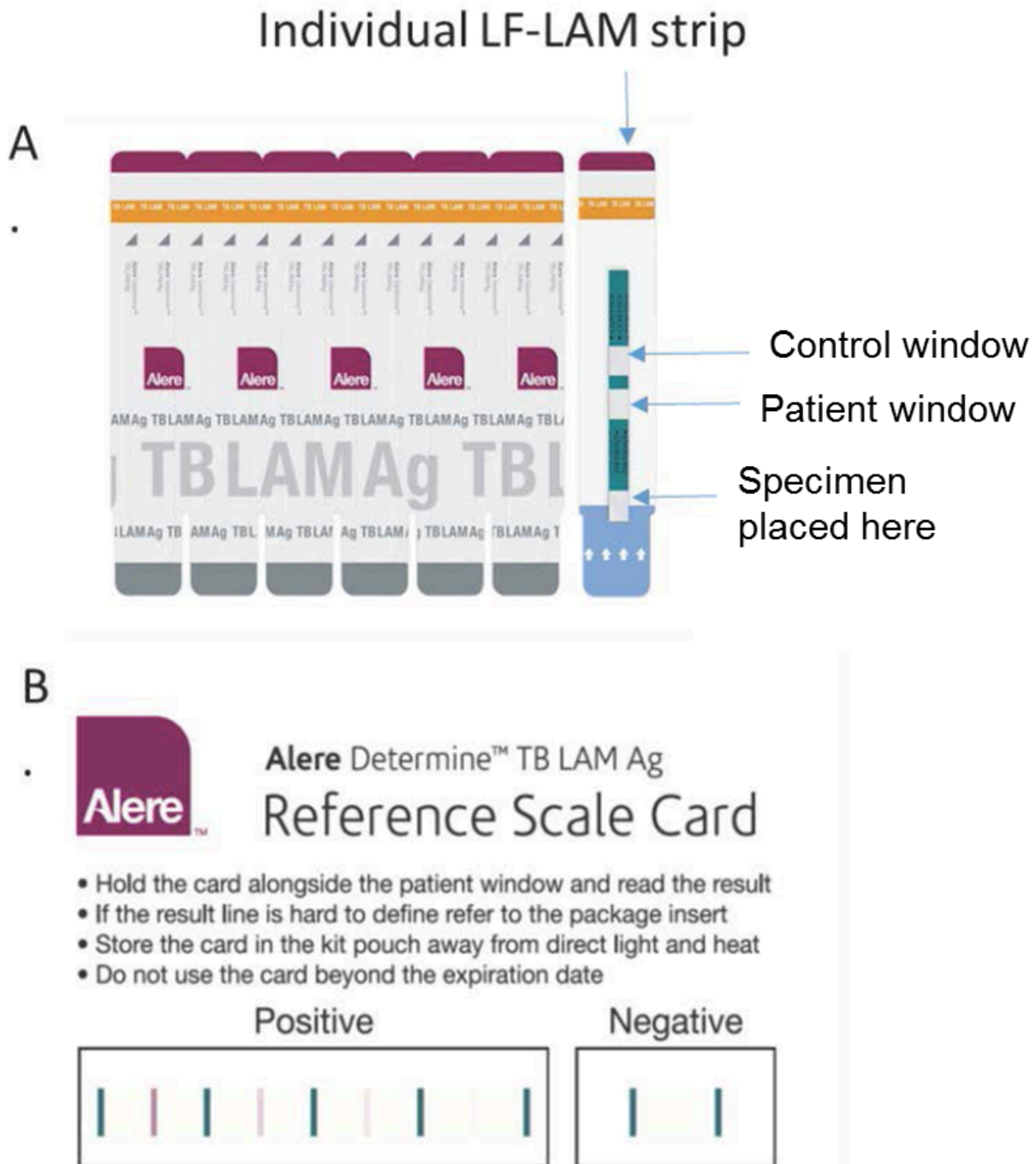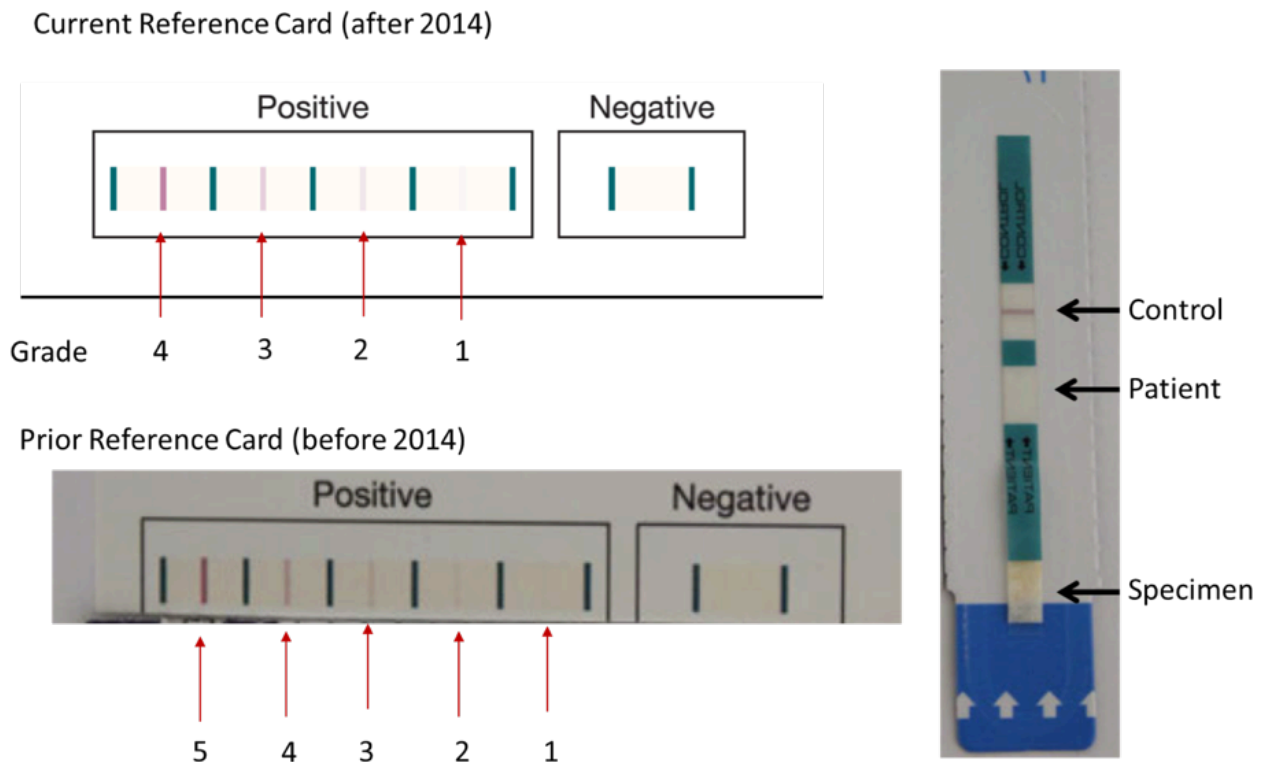

Appendix 2. Reference card grading of Alere Determine TB LAM Ag test

Figure 11.

Figure 11. Reference card grading of Alere Determine™ TB LAM. Comparison of the current reference card (top) with the prior reference card (bottom). Copyright © [2019] [Abbott Inc]: reproduced with permission.

Appendix 3. Minimum specifications for a point-of-care TB diagnostic test

Test specification	Minimum required value
Medical decision	Treatment initiation
Sensitivity: adults (for pulmonary tuberculosis only; regardless of HIV status)	Pulmonary tuberculosis • 95% for smear positive, culture positive • 60% to 80% for smear negative, culture positive (detection of extrapulmonary tuberculosis being a preferred but not minimal requirement)
Sensitivity: children (including extrapulmonary tuberculosis; regardless of HIV status)	• 80% compared to culture of any specimen and • 60% of probable tuberculosis (noting problem of lack of a gold standard)
Specificity: adults	• 95% compared to culture
Specificity: children	• 90% for culture-negative probable tuberculosis (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)

(Continued)

	(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	• 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	• Easy-to-read, unambiguous, simple "yes", "no", or "invalid" answer
	• Readable for at least one hour
Waste disposal	• Simple burning or sharps disposal; no glass component
	• Environmentally acceptable disposal
Controls	• Positive control included in test kit
	• Quality control simpler and easier than with sputum smear microscopy
Reagents	• All reagents in self-contained kit
	• Kit contains sample collection device and water (if needed)
Storage/stability required	• 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	• 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	• 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: USD, United States Dollar.

Appendix 4. Detailed search strategies

MEDLINE (PubMed)

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

Embase 1947-Present, updated daily

Search strategy:

----

1 tuberculosis.mp. or tuberculosis/ or Mycobacterium tuberculosis/ (115438)

2 limit 1 to yr="2014 -Current" (8833)

3 lipoarabinomannan.mp. or lipoarabinomannan/ (775)

4 LAM.mp. (4928)

5 limit 4 to yr="2014 -Current" (500)

6 3 or 5 (1252)

7 2 and 6 (79)

8 (test or assay or antigen or Ag or lateral flow assay* or urine antigen or point of care).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] (2733052)

9 limit 8 to yr="2014 -Current" (223771)

10 7 and 9 (46)

Cochrane Central Register of Controlled Trials : Issue 4 of 12, April 2018

ID Search

#1 tuberculosis:ti,ab,kw (Word variations have been searched)

#2 TB:ti, ab, kw

#3 MeSH descriptor: [Mycobacterium tuberculosis] explode all trees

#4 MeSH descriptor: [Tuberculosis] explode all trees

#5 #1 or #2 or #3 or #4

#6 LAM:ti,ab,kw

#7 lipoarabinomannan:ti,ab,kw

#8 #6 or #7

#9 #5 and #8

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

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 (test OR diagnos* OR urine OR assay)

CIDG Specialized Register, LILACS, Proquest dissertations, Current Controlled trials, WHO trials register

Tuberculosis AND (lipoarabinomannan OR LAM)

Appendix 5. Data collection form

LF-LAM- Lateral flow urine lipoarabinomannan assay for diagnosing active tuberculosis in people living with HIV		
Data form		
I. STUDY IDENTIFICATION		
1	First author	
2	Corresponding author and email	
3	Title of study	
4	Year of publication	
5	Year of study start	
6	Language if other than English	
II. STUDY DETAILS		
7	Population	1.Adults (15 years of age) 2.Children and adolescents 3.Both adults, children and adolescents 4.Other If other, describe
8	In which country or countries was the study conducted?	List all countries:
9	Country World Bank Classification (income) (World Bank 2018)	1.Low income 2.Lower-middle income 3.Upper-middle-income 4.High income 7. Other combination 9.Unknown/Not reported If other, describe:

(Continued)

10	Country WHO classification for high TB burden country (WHO Global Tuberculosis Report 2018)	1.Yes, part of the High TB/HIV burden list 2.No, not part of the High TB/HIV burden list
11	Study design	1.Randomized controlled trial 2.Cross-sectional 3.Cohort 7.Other, specify 9.Could not tell If other, describe:
12	Was a case-control design avoided?	1.Yes 2.No 9.Unclear

III. PATIENT SELECTION

13	What was the manner of participant selection into the study?	1.Consecutive 2.Random 3.Convenience 7.Other, specify 9.Unknown/Not Reported/Unclear If other, describe:
14	Direction of study data collection	1.Prospective 2.Retrospective 9.Unknown/Not reported
15	Please select the statement that best describes the selection of participants into the study.	1.HIV-positive participants with signs or symptoms suggestive of active tuberculosis were tested using LF-LAM. Please provide study definition of 'signs and symptoms': 2.A preAlered target population of HIV-positive individuals, irrespective of signs and symptoms of tuberculosis, were tested using LF-LAM. Please specify target population: 3.Both 1 and 2 4.Neither 1 nor 2. This is what was done:
16	Sample size	1._____ 9.Unknown/Not reported
17	Did the study avoid inappropriate exclusions?	1.Yes 2.No 9.Unknown/Not reported/Unclear
17a	Could the selection of patients have introduced bias?	1.High risk 2.Low risk 9.Unclear risk
18	NOTES ON PATIENT SELECTION	

IV. PATIENT CHARACTERISTICS AND SETTING

19	Presenting signs and symptoms	List
20	Age (years)	_____ If age is reported in median indicate IQR

(Continued)

		If age is reported in mean indicate SD
21	Age of all study participants, Range	Upper Lower
22	HIV infection (%)	
23	Participants included of female sex (%)	
24	CD4	_____ If CD4 is reported in median indicate IQR If CD4 is reported in mean indicate SD
25	Number (percent) of tuberculosis cases in the study (%):	
26	What was the target condition?	1.Pulmonary tuberculosis 2.Extra pulmonary tuberculosis 3.Mycobacteraemia 4.Both 1 and 2 5.Any of 1,2,3 7.Other, specify
27	Did the study include patients with prior tuberculosis history?	1.Yes 2.No 9.Unknown/Not reported If yes, what is the % _____ Specify the numerator/denominator _____/_____
28	What was the clinical setting of the study?	1.Outpatient 2.Inpatient 3.Both out-patient and in-patient 7.Other, describe: 9.Unknown/Not reported
29	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 7.Other, describe: 9.Unknown/Not reported
30	Are there concerns that the included patients and setting do not match the review question?	1.High concern 2.Low concern 9.Unclear concern
31	NOTES ON CHARACTERISTICS	

V. INDEX TEST

32	Was a LF-LAM threshold used to define positivity that was pre-specified in the primary analysis?	1.Yes, Grade 1/5 2.Yes, Grade 2/5 3.Yes, Grade 1/4 4.Yes, Grade 2/4 5.No 7.Other, specify:
33	What LF-LAM threshold was used to define positivity for data extraction?	1.Grade 2/5 2.Grade 1/4

(Continued)

		7. Other, specify
34	Are their concerns about index test conduct or interpretation differing from review question?	1.High concern 2.Low concern 9.Unclear concern
35	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 9.Unknown/Not reported
36	Was LF-LAM result interpreted without knowledge of the result of the reference standard result?	1.Yes 2.No 9.Unknown/Not reported/Unclear
37	Were there any LF-LAM results that were invalid (no bar in control window)?	1.Yes a.Specify number of invalid tests: _____ b.Were invalid tests repeated (yes/no): _____ 2.No 9. Unknown/Not reported
38	Could the conduct or interpretation of the index test have introduced bias?	1.High risk 2.Low risk 9.Unclear risk
39	NOTES ON INDEX TEST	

VI. REFERENCE STANDARD

40	For the diagnosis of pulmonary tuberculosis, what reference standard was used to identify tuberculosis and not tuberculosis?	1.Sputum: solid culture 2.Sputum: liquid culture 3.Sputum: both solid and liquid culture 4.Nucleic acid amplification test, specify 5.Any of culture or nucleic amplification test, specify 7.Other, specify
41	Was sputum induction performed for individuals unable to produce expectorated sputum?	1.Yes Specify N/% requiring sputum induction ____ 2.No
42	Were patients without sputum specimens (for example, no expectorated, no induced sputum) included in this study?	1.Yes Specify N/% included without sputum ____ 2.No Specify N/% excluded due to lack of sputum____
43	Were non-pulmonary specimens evaluated to allow diagnosis of extrapulmonary tuberculosis?	1.All participants received testing of non-pulmonary specimens, please specify sites/fluids: 2.Some participants received testing of non-pulmonary specimens, please specify which patients were tested, and sites/fluids: 3.Extrapulmonary tuberculosis was not evaluated 7.Other, please specify:
44	For the diagnosis of extrapulmonary tuberculosis, what tests were used to identify tuberculosis and not tuberculosis (circle all that apply)?	1.Solid culture 2.Liquid culture 3.Both solid and liquid culture 4.Nucleic acid amplification test, specify 7.Other, specify: _____

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)

(Continued)

		8.Not applicable, extrapulmonary tuberculosis was not evaluated
45	Did the study speciate mycobacteria isolated in culture?	1.Yes 2.No 9.Unknown/Not reported
46	Was the reference standard likely to correctly classify the target condition	1.Yes 2.No 9. Unclear
47	Was the reference standard result interpreted without knowledge of the result of LF-LAM?	1.Yes 2.No 9.Unclear
48	How many sputum specimens were obtained in order to detect pulmonary tuberculosis?	1. Single 2. Multiple 8.Not applicable
49	How many specimens from fluid (sites) other than sputum were obtained to detect extrapulmonary tuberculosis?	1. Single 2. Multiple 8.Not applicable
50	Could the reference standard, its conduct, or its interpretation have introduced bias?	1.High risk 2.Low risk 9.Unclear risk
51	Are there concerns that the target condition as defined by the reference standard does not match the question?	1.High concern 2.Low concern 9.Unclear concern
52	NOTES ON REFERENCE STANDARD	

VII. FLOW AND TIMING

53	Was there appropriate interval between index test and reference standard	1.Yes, specimens collected at the same time. 2.No, specimens collected greater than 7 days apart 9.Unclear
54	Did all patients receive a reference standard?	1.Yes 2.No 9.Unclear
55	Did all patients receive the same reference standard?	1.Yes 2.No (answer no if clinicians chose sample types, or other differences in reference standards between patients) 9.Unclear
56	Were all participants included in the analysis?	1.Yes 2.No 9.Unclear
57	Could the patient flow have introduced bias?	1.High risk 2.Low risk 9.Unclear risk
58	NOTES ON FLOW AND TIMING	

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

VIII. TABLES: Tuberculosis detection against a microbiological reference standard

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

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

(Table example; provide additional tables for each of the applicable questions)

LAM result		Tuberculosis	Not tuberculosis	Total
	Positive			
	Negative			
	Total			

Provide additional tables for each of the questions:

What is the diagnostic accuracy of LF-LAM for the diagnosis of tuberculosis in all HIV-positive adults with advanced HIV disease and signs and symptoms of tuberculosis?
a. in inpatient setting CD4 ≤ 200
b. in outpatient setting CD4 ≤ 200
c. in all settings CD4 ≤ 200
d. in all settings CD4 > 200
e. in inpatient setting CD4 ≤ 100
f. in outpatient setting CD4 ≤ 100
g. in all settings CD4 ≤ 100
h. in all settings CD4 > 100
i. in inpatients settings CD4 101 to 200
j. in outpatient settings CD4 101 to 200
k. in all settings CD4 101 to 199

What is the diagnostic accuracy of LF-LAM for the diagnosis of tuberculosis in adults with advanced HIV disease irrespective of signs and symptoms of tuberculosis?
a. in all settings CD4 > 200
b. in all settings CD4 > 100
c. in inpatients settings CD4 101 to 200
d. in outpatient settings CD4 101 to 200
e. in all settings CD4 101 to 199

Appendix 6. QUADAS-2

Domain 1: patient selection

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

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

We answered '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.

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

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

Signaling question 3: Did the study avoid inappropriate exclusion?

We answered 'yes' to studies which included all HIV-positive participants and participants who were unable to produce sputum (expectorated or induced). We answered 'no' if studies excluded participants who could not produce sputum (i.e. there were no attempts

at sputum induction or patients could not produce sputum despite sputum induction and were excluded). We also answered 'no' if studies excluded patients presumed to have extrapulmonary tuberculosis. 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 judge 'low concern' for most studies since we planned to determine test accuracy both for patients with signs and symptoms of tuberculosis and patients investigated for tuberculosis irrespective of signs and symptoms for tuberculosis.

For LF-LAM used as a tuberculosis diagnostic test among patients with signs and symptoms of tuberculosis, we judged 'high concern' if the study participants did not resemble people with presumed HIV/tuberculosis; 'low concern' if the study population did resemble a population with presumed HIV/tuberculosis, and 'unclear concern', if we could not tell.

For LF-LAM used as a tuberculosis diagnostic test among patients that were investigated for tuberculosis irrespective of signs and symptoms of tuberculosis, we judged '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. We judged 'high concern' if the study participants did not resemble people with presumed HIV/tuberculosis coinfection.

Domain 2: index test

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

Signaling 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 and we could not tell if the index test results were interpreted without knowledge of the reference standard results.

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

We answered 'yes' if the threshold was prespecified in the study or by the authors, '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. In cases where the primary study defined grade 1 of 5 as the positivity threshold, but where we were able to extract data at the manufacturer's currently recommended positivity threshold, we judged 'low concern' for applicability.

Domain 3: reference standard

Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias?

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

HIV-infected tuberculosis patients may have pulmonary tuberculosis, extrapulmonary tuberculosis, or both pulmonary and extrapulmonary tuberculosis. A microbiological reference standard, primarily culture, is considered the gold standard for tuberculosis. Owing to the difficulties in diagnosing HIV-associated tuberculosis, 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 tuberculosis. For presumed pulmonary tuberculosis, 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 tuberculosis, specimens should be consistent with Standard 4 of the International Standards for Tuberculosis 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 tuberculosis. An Xpert® MTB/RIF test is recommended as the preferred initial microbiological test for suspected tuberculosis 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 tuberculosis lymphadenitis receive lymph node tissue sampling). We answered 'unclear' if we could not tell.

Signaling 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?

Signaling 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 tuberculosis or natural progression of tuberculosis 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.

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

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

Signaling 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.

Judgements for 'Risk of bias' assessments:

If we answered all signalling questions for a domain "yes", then we judged risk of bias as "low".
If we answered all or most signalling questions for a domain "no", then we judged risk of bias as "high".
If we answered only one signalling question for a domain "no", we discussed further the "risk of bias" judgement.
If we answered all or most signalling questions for a domain "unclear", then we judged risk of bias as "unclear".
If we answered only one signalling question for a domain "unclear", we discussed further the "risk of bias" judgement for the domain.

Appendix 7. Statistical appendix

We list here the OpenBUGS program used to fit the bivariate meta-analysis models for estimating the accuracy of the index test. In the subsections below, we first describe the likelihood and prior distribution for the model followed by the OpenBUGS program.

As is usual with Bayesian models, initial values must be provided for all unknown parameters. We selected three independent sets of initial values for the parameters using the in-built ModelGenInits() function within OpenBUGS. The Gelman-Rubin statistic within the OpenBUGS program was used to assess convergence. We did not observe any convergence problems for the analyses presented. We treated the first 10,000 iterations as burn-in iterations and dropped them. We obtained summary statistics based on a total of 150,000 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 Binomial(p,N), the Bivariate Normal probability distribution with mean vector μ and variance-covariance matrix TAU as BVN(mu, TAU), the univariate Normal distribution with mean m and variance tau^2 by N(m, tau^2) and the Uniform probability distribution between a and b by Uniform(a,b). Note that logit refers to log odds.

Likelihood:

Within studies:

TP_i ~ Binomial(TPR_i, TP_i + FN_i), and

FP_i ~ Binomial(FPR_i, TN_i + FP_i)

Between studies:

The bivariate vector (logit(TPR_i), logit(FPR_i)) ~ BVN(mu = (mu_1, mu_2), TAU) where

TAU is a 2 X 2 matrix with entries

TAU[1,1] = variance of logit(TPR_i) = tau_1^2,

TAU[2,2] = variance of logit(FPR_i) = tau_2^2 and

TAU[1,2] = TAU[2,1] = covariance between logit(TPR_i) and logit(FPR_i) = rho × tau_1 × tau_2

and rho is the correlation between logit(TPR_i) and logit(FPR_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:

mu_1 and mu_2 ~ N(m=0, tau^2=4),

rho ~ Uniform(-1, 1)

$(1/tau_1^2)$ and $(1/tau_2^2)$ ~ Gamma(shape=2, rate=0.5)

A.1 OpenBUGS 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

# Likelihoood

pos[i]<-TP[i]+FN[i]

neg[i]<-TN[i]+FP[i]

TP[i] ~ dbin(TPR[i],pos[i])

FP[i] ~ dbin(FPR[i],neg[i])logit(TPR[i]) <- l[i,1]

logit(FPR[i]) <- -l[i,2]se[i] <- TPR[i]

sp[i] <- 1-FPR[i]

l[i,1:2] ~ dmnorm(mu[1:2], T[1:2, 1:2])

}

# Prior Distributions
```

![Cochrane Library logo] Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

```
mu[1] ~ dnorm(0,0.25)

mu[2] ~ dnorm(0,0.25)

T[1:2,1:2]<-inverse(TAU[1:2,1:2])

# Between-study variance-covariance matrix

TAU[1,1] <- tau[1]*tau[1]

TAU[2,2] <- tau[2]*tau[2]

TAU[1,2] <- rho*tau[1]*tau[2]

TAU[2,1] <- rho*tau[1]*tau[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)

tau[1]<-pow(prec[1],-0.5)

tau[2]<-pow(prec[2],-0.5)

# Pooled sensitivity and specificity

Pooled_S<-1/(1+exp(-mu[1]))

Pooled_C<-1/(1+exp(-mu[2]))

}

# Predicted sensitivity and specificity in a new study

l.new[1:2] ~ dmnorm(mu[],T[,])

sens.new <- 1/(1+exp(-l.new[1]))

spec.new <- 1/(1+exp(-l.new[2]))

}

# An R function (SROC_BUGS) for creating a joint summary plot of sensitivity and specificity of the index test
# using results from the OpenBUGS program above can be found in the DTAplots R package
# available for download at https://www.nandinidendukuri.com/software/diagnostic-meta-analysis
```

Appendix 8. Summary plot of LF-LAM sensitivity and specificity for tuberculosis detection in symptomatic participants.
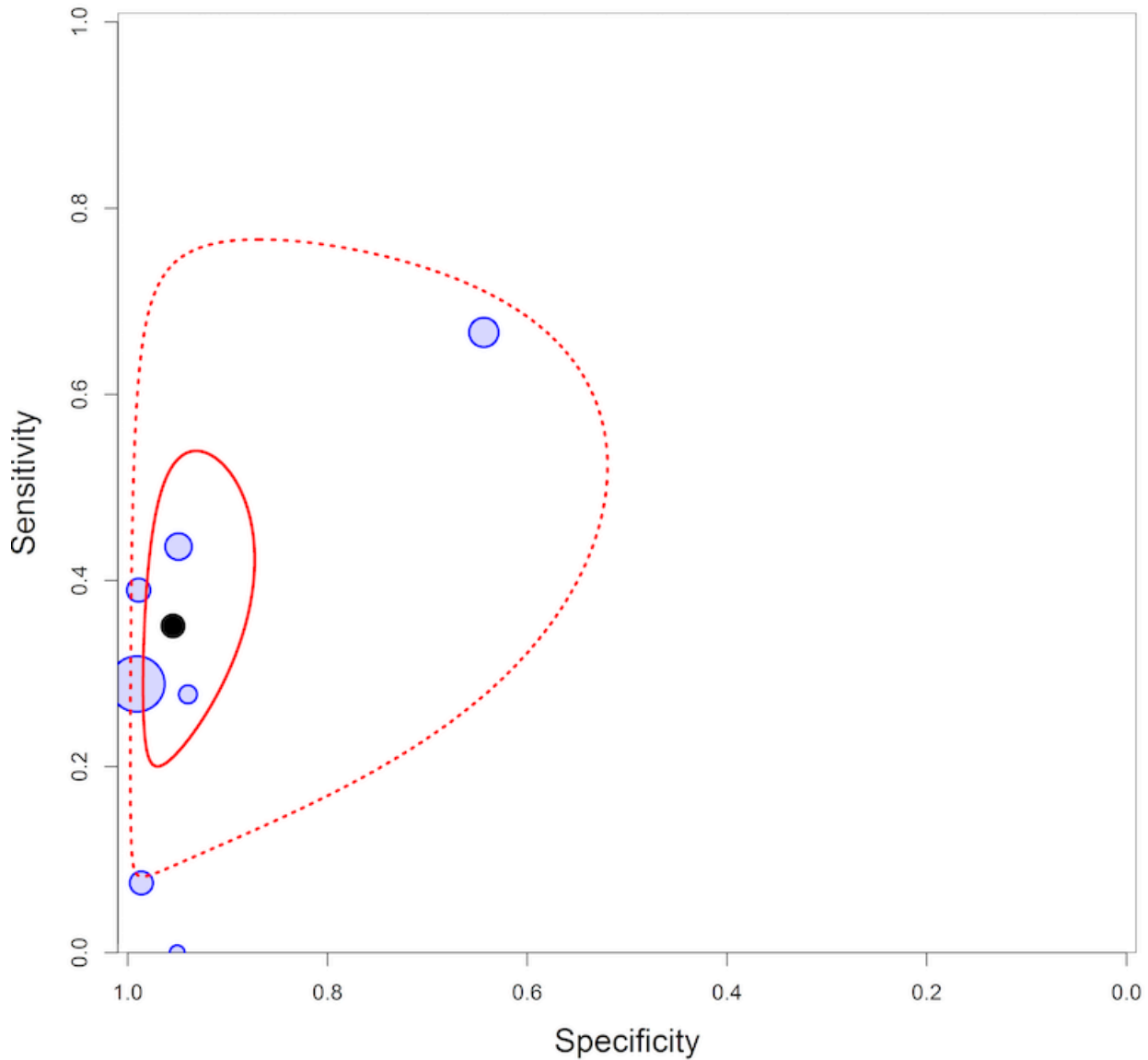
Figure 12.

Figure 12. Summary plot of LF-LAM sensitivity and specificity for tuberculosis detection in symptomatic participants. The blue circles represent individual study estimates for sensitivity and specificity for the studies among symptomatic participants. The size of the circle is proportional to the sample size of the study. The filled black circle is the pooled estimate for sensitivity and specificity. The solid red line marks the 95% credible region around the summary estimate, the dashed red line marks the 95% prediction region.
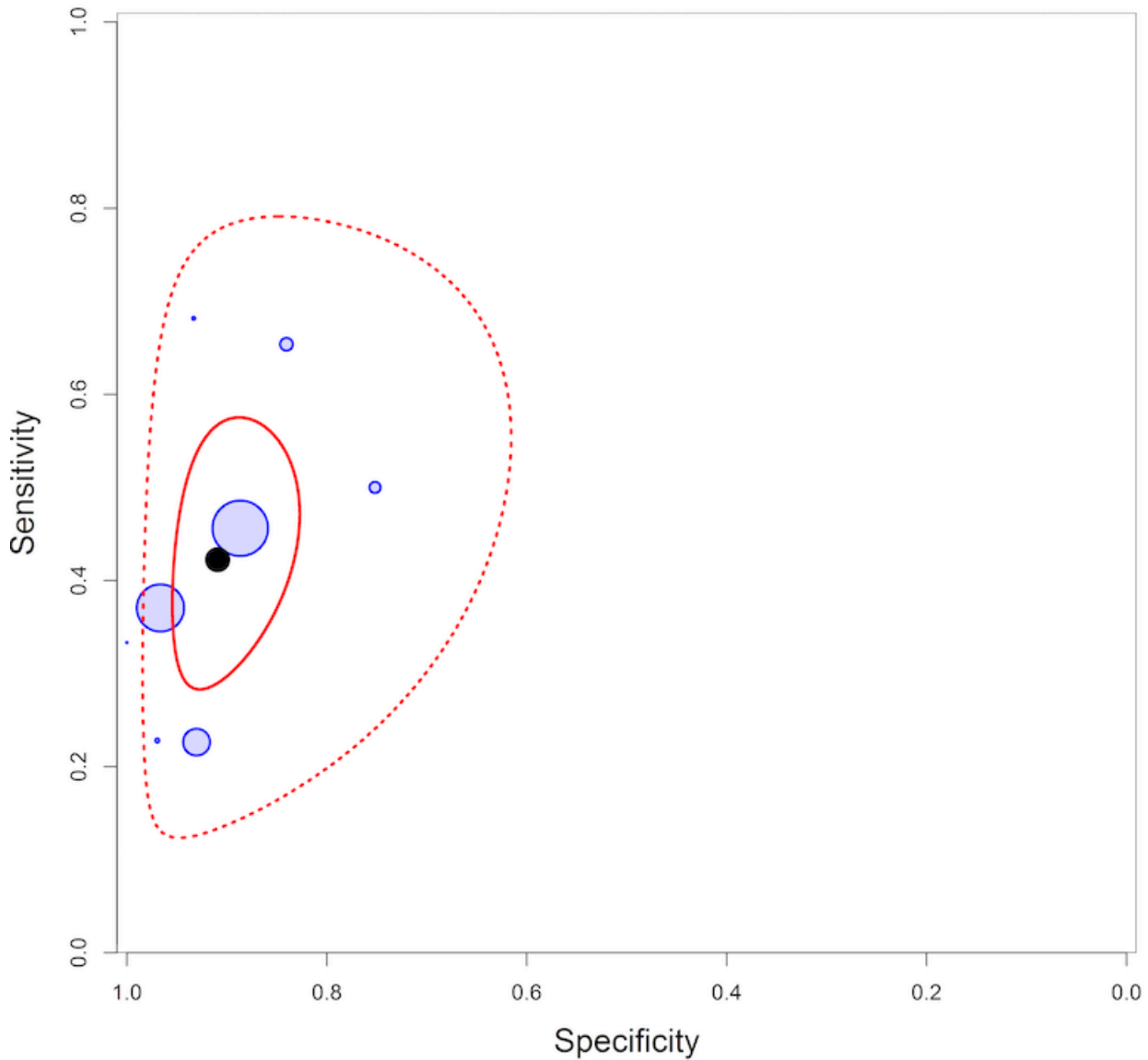
Appendix 9. Summary plot of LF-LAM sensitivity and specificity for tuberculosis detection in unselected participants.

Figure 13.

Figure 13. **Summary plot of LF-LAM sensitivity and specificity for tuberculosis detection in unselected participants. The blue circles represent individual study estimates for sensitivity and specificity for the studies among unselected participants. The size of the circle is proportional to the sample size of the study. The filled black circle is the pooled estimate for sensitivity and specificity. The solid red line marks the 95% credible region around the summary estimate, the dashed red line marks the 95% prediction region.**

Appendix 10. Unselected participants with CD4 > 200 and CD4 ≤ 200

Figure 14.

Figure 14.   Forest plots of LF-LAM sensitivity and specificity for tuberculosis against a microbiological reference standard for studies among unselected participants, by CD4 strata (CD4 > 200 and CD4 ≤ 200) and health setting. The individual studies are ordered by decreasing sensitivity. 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).

## Appendix 11. Unselected participants with CD4 > 100 and CD4 ≤ 100

Figure 15.

Figure 15.   Forest plots of LF-LAM sensitivity and specificity for tuberculosis against a microbiological reference standard for studies among unselected participants, by CD4 strata (CD4 > 100 and CD4 ≤ 100) and health setting. 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)
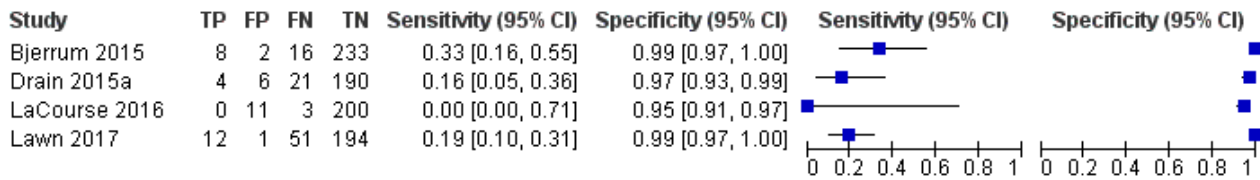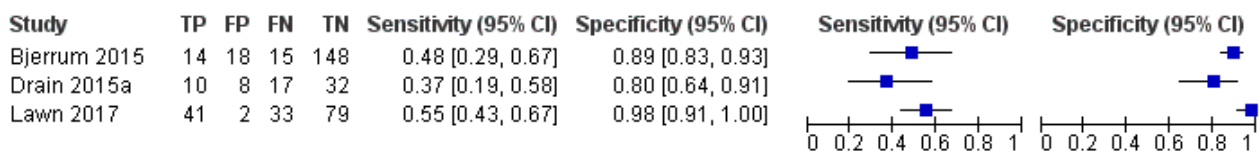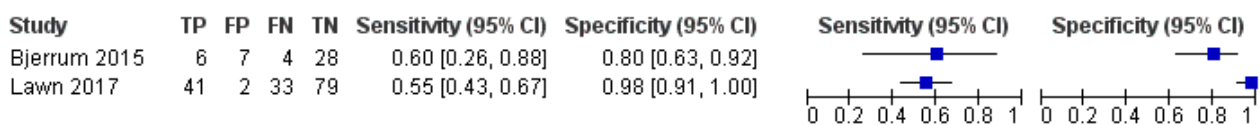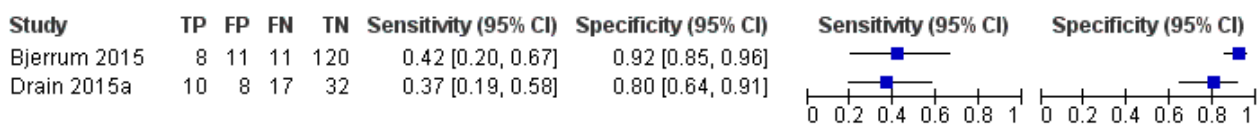

Appendix 12. Characteristics of studies among children with HIV

Study	Country	Age of enrolment	Presence of tuberculosis symptoms?	Setting	Tuberculosis prevalence % (n/N)	Percent children with advanced or severe immunosuppression
Kroidl 2015	Tanzania	6 weeks to 14 years, median 8.8 years (IQR 3.9 to 9.5)a	Tuberculosis symptoms	Outpatient	40% (12/30	65%
LaCourse 2018a	Kenya	≤ 12 years, median 24 months (IQR 13 to 58)	Irrespective of tuberculosis symptoms	Inpatient	7% (9/130)	70%
Nicol 2014	South Africa	≤ 15 years, median 42.5 months (IQR 19.1 to 66.3)*	Tuberculosis symptoms	Inpatient and outpatient	22% (23/160)	53%

For all participants, including HIV-positive and HIV-negative children.
IQR: interquartile range.

WHAT'S NEW

Date	Event	Description
9 October 2019	New citation required but conclusions have not changed	We included 15 studies in this review update: nine new studies and six studies from the original review (Shah 2016) that met the inclusion criteria. The findings in this update are consistent with those reported previously (Shah 2016).
9 October 2019	New search has been performed	A new search was performed. The review authors identified 15 unique studies, integrating nine new studies to six of those included in the previous Cochrane Review (Shah 2016). We used the terminology of lateral flow urine lipoarabinomannan assay (LF-LAM). The Alere Determine™ TB LAM Ag assay was the only LF-LAM test commercially available at the time of the review. New tests coming into the market, such as the FujiLAM, was not commercialized nor with any published data as of May 2018). We evaluated performance of LF-LAM only at grade 1 on the update reference scale card (scale of 1 to 4) equivalent to grade 2 on the old reference scale card (scale 1 to 5). We evaluated LF-LAM accuracy against a microbiological reference standard and did not include analysis against a composite reference standard. We grouped the analysis by studies among 'Symptomatic' and 'Unselected' participants instead of the previous terminology of 'Diagnosis' and 'Screening'. We changed the QUADAS judgements for some of the previously included studies.

CONTRIBUTIONS OF AUTHORS

MS and SB reviewed articles for inclusion and extracted data. MS, SB, IS, ND, and KRS analysed the data. MS, SB, IS, ND, CMD, and KRS interpreted the analyses. SB, MS, and KRS drafted the manuscript in adults. RRN reviewed articles with impact data and drafted the summary sections on patient outcome data. MK, SB, MS, and KRS drafted the GRADE tables. ND and IS drafted the statistical analysis section and the statistical appendix (Appendix 7). All authors 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 Tuberculosis Programme, Switzerland. KRS, MS, ND, and SB received funding to carry out the review from USAID.

KRS has also received financial support for the preparation of systematic reviews and educational materials, consultancy fees from FIND (for the preparation of systematic reviews), honoraria, and travel support to attend WHO guideline meetings.

RN is a board member of TB Proof and has participated on an advisory board for Insmed, but these are not thought to represent a conflict of interest.

CMD is employed by FIND, a Swiss non-profit organization and WHO Collaborating Centre for Diagnosis. FIND provided funding for an initial assessment of data available for the review.

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.

Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV (Review)
Copyright © 2019 The Authors. Cochrane Database of Systematic Reviews published by John Wiley & Sons, Ltd. on behalf of The Cochrane Collaboration.

96

SOURCES OF SUPPORT

Internal sources

- Liverpool School of Tropical Medicine, UK.

External sources

- Department for International Development, UK.

  Project number 300342-104
- 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
- The Odense University Hospital Internationalisation Fund, Denmark.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Differences between review and review update

To further investigate for heterogeneity in data, we carried out additional subgroup analyses for the following CD4 strata: CD4 > 200; CD4 > 100 and CD4 101 to 200 cells/μL. We also performed a subgroup analysis by tuberculosis prevalence in the study population (pre-test probability). In the review, we also estimated the predicted sensitivity and specificity in a future study together with their 95% credible intervals (CrIs), which we did not mention in the protocol.

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