Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV

Policy update (2019)

Web Annex A. LF-LAM assay for detecting active tuberculosis in people living with HIV: an updated systematic review

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see http://www.who.int/about/licensing.

Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers’ products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use. The named authors alone are responsible for the views expressed in this publication.

This publication forms part of the WHO guideline entitled Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV: policy update.
It is being made publicly available for transparency purposes and information, in accordance with the *WHO handbook for guideline development*, 2nd edition (2014).
ACKNOWLEDGEMENTS

Stephanie Bjerrum1,2, Ian Schiller3, Nandini Dendukuri3, Michael Eisenhut4, Mikashmi Kohli5, Ruvandhi R Nathavitharana6, Alice A Zwerling7, Claudia M Denkinger8,9, Karen R Steingart10, Maunank Shah11

1Department of Clinical Research, Research Unit of Infectious Diseases, University of Southern Denmark, Odense, Denmark
2OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark
3McGill Global Health Programs and McGill International Tuberculosis Centre and Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
4Paediatric Department, Luton & Dunstable University Hospital NHS Foundation Trust, Luton, UK
5Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
6Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA
7School of Epidemiology & Public Health, University of Ottawa, Ottawa, Canada
8Foundation for Innovative New Diagnostics (FIND), Geneva, Switzerland
9Center of Infectious Diseases, University of Heidelberg, Germany
10Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK
11Division of Infectious Diseases, Center for Tuberculosis Research, Johns Hopkins University, Baltimore, Maryland, USA

Corresponding author
Stephanie Bjerrum, MD PhD
Department of Clinical Research, Research Unit of Infectious Diseases
University of Southern Denmark
Winsløwparken 19, 3.sal, DK-5000 C Odense C
Denmark
sbjerrum@health.sdu.dk
## Table of Contents

**EXECUTIVE SUMMARY** .......................................................................................................................... 7

**BACKGROUND** ...................................................................................................................................... 12

**PICO QUESTIONS** ................................................................................................................................. 15

**METHODS** .............................................................................................................................................. 17

- Criteria for considering studies for this review .................................................................................. 17
- Search methods for identification of studies ....................................................................................... 19
- Selection of studies ................................................................................................................................ 19
- Data extraction and management ....................................................................................................... 19
- Assessment of methodological quality ............................................................................................... 20
- Statistical analysis and data synthesis ............................................................................................... 20
- Approach to uninterpretable ALERE LAM results .......................................................................... 22
- Investigations of heterogeneity ........................................................................................................ 22
- Sensitivity analyses ........................................................................................................................... 22
- Additional analyses ............................................................................................................................ 23
- Assessment of reporting bias ........................................................................................................... 24
- Assessment of the certainty of the evidence .................................................................................... 25

**RESULTS** ............................................................................................................................................... 26

- Results of the search ............................................................................................................................ 26
- Methodological quality of included studies ...................................................................................... 28
- Findings .................................................................................................................................................. 31

  **PICO 1:** What is the diagnostic accuracy of ALERE LAM for TB diagnosis in HIV-positive adults with signs and symptoms of TB? ........................................................................... 33

  **PICO 2:** What is the diagnostic accuracy of ALERE LAM for TB diagnosis in HIV-positive adults irrespective of signs and symptoms for TB? .................................................................................. 39

  **PICO 3:** What is the diagnostic accuracy of ALERE LAM for diagnosis of TB in adults with advanced HIV disease irrespective of signs and symptoms of TB? .......................................................... 42

  **PICO 4:** Can the use of ALERE LAM in HIV-positive adults reduce mortality associated with advanced HIV disease? ......................................................................................................................... 49

- Association between ALERE LAM positivity with mortality ............................................................ 53

**DISCUSSION** .......................................................................................................................................... 59

- Summary of main results .................................................................................................................... 59

  **ALERE LAM for TB diagnosis in symptomatic participants** .......................................................... 60

  **ALERE LAM for TB diagnosis in unsel ected participants** ............................................................. 60

  **ALERE LAM for TB diagnosis, overall** ......................................................................................... 61

- Impact on mortality ............................................................................................................................. 63

- Strengths and weaknesses of the review ............................................................................................. 65

**AUTHORS’ CONCLUSIONS** .................................................................................................................. 65

- Implications for practice ..................................................................................................................... 65

- Implications for research ..................................................................................................................... 66

**ACKNOWLEDGEMENTS** ...................................................................................................................... 68

**CONTRIBUTIONS OF AUTHORS** ....................................................................................................... 68
DECLARATIONS OF INTEREST........................................................................................................68
REFERENCES ....................................................................................................................................69
INCLUDED STUDIES .........................................................................................................................69
ADDITIONAL STUDIES FOR IMPACT AND ASSOCIATION WITH MORTALITY..............................71
ADDITIONAL REFERENCES ................................................................................................................71
OTHER PUBLISHED VERSIONS OF THIS REVIEW ...........................................................................77
APPENDICES ...............................................................................................................................78
APPENDIX 1. REFERENCE CARD GRADING OF ALERE DETERMINE™ TB LAM .........................78
APPENDIX 2. PICO QUESTIONS ........................................................................................................79
APPENDIX 3. DETAILED SEARCH STRATEGIES ............................................................................80
APPENDIX 4. DATA COLLECTION FORM, DIAGNOSTIC ACCURACY ........................................82
APPENDIX 5. DATA COLLECTION FORM, IMPACT DATA ...............................................................89
APPENDIX 6. QUADAS-2 .................................................................................................................92
APPENDIX 7. STATISTICAL APPROACH ........................................................................................96
APPENDIX 8: CHARACTERISTICS OF INCLUDED STUDIES .......................................................99
APPENDIX 9: DIAGNOSTIC ACCURACY OF ALERE LAM AMONG HIV-POSITIVE CHILDREN, SUMMARY.............................................................................................................101
Executive summary

The lateral flow urine lipoarabinomannan assay Alere Determine™ TB LAM Ag ‘AlereLAM’ is a commercially available point-of-care test that detects lipoarabinomannan, a lipopolysaccharide present in mycobacterial cell walls, in people with active tuberculosis (TB), including both pulmonary and extrapulmonary forms of disease. This systematic review summarizes the current literature on the accuracy of the AlereLAM for diagnosis of TB in people living with HIV as part of a World Health Organization process to develop updated guidelines for use of AlereLAM. AlereLAM is being considered as a diagnostic test that may be used in combination with existing tests for the diagnosis of HIV-associated TB. We report data on children separately from adults.

We identified 15 unique published studies that assessed the accuracy of AlereLAM in adults and integrated nine new studies identified since the original WHO and Cochrane reviews in 2015 and 2016. We classified studies that evaluated AlereLAM in participants with signs and symptoms of TB as ‘studies with symptomatic participants’ and studies that included both individuals with symptoms of TB and individuals without symptoms of TB (i.e. enrolled irrespective of symptoms) as ‘studies with unselected participants’. All studies were performed in TB/HIV high burden countries. For this review, we report positive AlereLAM results in accordance with the manufacturer’s updated recommendations for test interpretation (graded 1 to 4 based on band intensity). We estimated sensitivity and specificity at the grade 1 cut-off for positivity on the updated reference scale card, corresponding to grade 2 on the prior reference scale card with band intensities graded on a scale of 1 to 5. We performed all analyses with respect to a microbiological reference standard.

AlereLAM for TB diagnosis in HIV-positive adults with signs and symptoms of TB

Of the 15 included studies, eight studies reported accuracy data on AlereLAM for TB diagnosis among adults that presented with signs and symptoms of TB. Six of the studies contributed data partially or exclusively for inpatient settings. As assessed by QUADAS-2, six studies (75%) had high risk of bias in the patient selection domain, and seven studies (88%) had high risk of bias in the reference standard domain. Regarding applicability, we scored low concern for most studies in all domains. AlereLAM sensitivity and specificity varied with setting and CD4 cell count.

For all settings, AlereLAM pooled sensitivity and specificity (95% credible interval (CrI)) were 42% (31% to 55%) and 91% (85% to 95%), respectively (eight studies, 3449 participants (37% with TB); moderate-certainty evidence for sensitivity and low-certainty evidence for specificity).

Results of these studies indicate that in theory, for a population of 1000 people where 300 have microbiologically-confirmed TB, 189 would be AlereLAM-positive: of these, 63 (33%) would not have TB (false-positives); and 811 would be AlereLAM-negative: of these, 174 (21%) would have TB (false-negatives).
Stratified by setting, pooled sensitivity was 52% (40% to 64%) among inpatients versus 29% (17% to 47%) among outpatients. Pooled specificity was lower among inpatients, 87% (78% to 93%), versus 96% (91% to 99%) among outpatients.

Stratified by CD4 cell count, pooled sensitivity increased and specificity decreased with lower CD4 cell count. For all settings, in participants with CD4 ≤ 200 cells per µL pooled sensitivity was 45% (31% to 61%) versus 16% (8% to 31%) in participants with CD4 > 200 cells per µL. 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. In participants with a CD4 ≤ 100 cells per µL pooled sensitivity was 54% (38% to 69%) versus 17% (10% to 27%) in participants with CD4 > 100 cells per µL. Pooled specificity was 88% (77% to 94%) in participants with a CD4 ≤ 100 cells per µL and 95% (89% to 98%) in participants with CD4 > 100 cells per µL. Pooled sensitivity in participants with CD4 between 101-199 cells per µL was 24% (14% to 38%).

AlereLAM for TB diagnosis in HIV-positive adults irrespective of signs and symptoms of TB

Of the 15 included studies, seven studies reported accuracy data on AlereLAM for TB diagnosis among unselected adults who may or may not have presented with TB symptoms at enrolment (i.e. enrolled irrespective of signs and symptoms of TB). Studies were predominantly conducted in outpatient settings among patients with higher CD4 cell counts and lower TB prevalence compared with studies evaluating the test for TB diagnosis among exclusively symptomatic patients. Studies ranged from including 19% of participants with symptoms to 91%. As assessed by QUADAS-2, four studies (57%) had high risk of bias in the patient selection domain, and five studies (71%) in the reference standard domain. Regarding applicability, we scored low concern for all studies in all domains.

For all settings, AlereLAM pooled sensitivity and specificity were 35% (22% to 50%) and 95% (89% to 96%), respectively (seven studies, 3365 participants (13% with TB); moderate-certainty evidence for sensitivity and low-certainty evidence for specificity).

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

Stratified by setting, pooled sensitivity was 62% (41% to 83%) among inpatients versus 31% (18% to 47%) among outpatients. Pooled specificity was lower among inpatients, 84% (48% to 96%) versus 95% (87% to 99%) for outpatients.

For all settings, stratified by CD4 cell count, in unselected participants with CD4 ≤ 200 cells per µL, AlereLAM pooled sensitivity and specificity were 26% (9% to 56%) and 96% (87% to 98%) (two studies). Pooled sensitivity in participants with a CD4 ≤ 100 cells per µL was 47% (30% to 64%) versus 20% (10% to 35%) in participants with CD4 > 100 cells per µL. Specificity was 90% (77% to
96%) in participants with a CD4 ≤ 100 cells per µL and 98% (95% to 99%) in participants with CD4 > 100 cells per µL. For other CD4 strata we had limited data.

Impact of AlereLAM on mortality
We identified two multi-site randomized controlled trials that included data on the impact of AlereLAM on mortality and other patient outcomes. Both trials were conducted in sub-Saharan Africa, with each including a study site in South Africa. Both trials involved hospitalized, HIV-positive patients, used the results of AlereLAM to guide therapy, and assessed all-cause mortality at eight weeks. We note that data stratified by CD4 count were limited. In the meta-analysis, the pooled risk ratio for mortality was 0.85 (0.76 to 0.94), that is, study participants undergoing AlereLAM testing had a 15% lower risk of mortality than participants undergoing routine TB diagnostic testing without AlereLAM; the absolute effect was 35 fewer deaths per 1,000 (from 14 fewer to 55 fewer deaths) (high-certainty evidence).

Association of AlereLAM and mortality
We identified 12 studies that had data on the association between AlereLAM positivity and mortality as part of post-hoc analyses within diagnostic accuracy studies (in which AlereLAM was not used for clinical decision making). The timing of mortality analysis, setting, use of TB therapy, and outcome measures to compare AlereLAM positive and AlereLAM negative patients differed across studies. In a descriptive analysis, 11 out of 12 of these studies suggested that there was an association of AlereLAM test positivity and mortality. Of importance, we note that these studies did not use results of AlereLAM to guide therapy.

AlereLAM studies in children
We identified three published studies of AlereLAM in children as the result of a broader search for studies in adults and children using the same inclusion criteria. The three studies involved a total of 266 HIV-positive children. One study enrolled children aged 14 years and less; one study enrolled children aged 12 years and less; and one study enrolled children aged 15 years and less. For the three studies, median age ranged from 24 months to 6.8 years. Two studies included HIV-positive children presumed to have TB with symptoms and one included HIV-positive children irrespective of TB signs and symptoms. One study was conducted in an outpatient setting, one in an inpatient setting, and one in both an inpatient and an outpatient setting. All three studies took place in high TB/HIV burden countries in Africa. The prevalence of microbiologically-confirmed TB ranged from 7% to 40% in the studies.

Given the differences in population and setting, we did not perform meta-analyses and provide sensitivity and specificity estimates for individual studies. In all settings, including all children, sensitivity and specificity (95% CI) were 42% (15% to 72%) and 94% (73% to 100%), (30 participants, outpatient); 56% (21% to 86%) and 95% (90% to 98%), (130 participants, inpatient); and 43% (23% to 66%) and 80% (69% to 88%), (106 participants, both inpatient and outpatient).
Two studies provided data stratified by age group. In adolescents, AlereLAM sensitivities were 100% (3% to 100%) (four participants, inpatient) and 60% (15% to 95%) (nine participants, outpatient); in both studies, specificity was 100%. In children ≤ 5 years, sensitivities were 50% (7% to 93%) (95 participants, inpatient), and 25% (1% to 81%) (13 participants, outpatient); corresponding specificities were 93% (86% to 98%) and 89% (52% to 100%).

**Authors' conclusions**

We found that AlereLAM has lower sensitivity to detect TB in adults living with HIV than the internationally suggested target of minimum 65% overall for non-sputum based TB tests (WHO TTP 2014). This finding was consistent whether the test is used for diagnosis of TB among symptomatic participants (sensitivity of 42%) or unselected participants (sensitivity of 35%). The estimated sensitivity suggests that if AlereLAM were to be used alone, more than half of all TB cases would be missed. Although the estimated sensitivity is lower than the WHO target for non-sputum based TB tests, two randomised controlled trials implementing AlereLAM have demonstrated a mortality reduction and impact on other patient health outcomes when used in hospitalized HIV-positive adults.

The proposed role for the AlereLAM test is to be used in combination with other existing TB tests to assist TB diagnosis and possibly improve important outcomes among HIV-positive patients with advanced disease. The test does not require sputum collection and is not site-specific. Other favorable test characteristics include low-cost, rapidity (less than 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).

Findings suggest that sensitivity increases with lower CD4 cell count and among inpatients regardless of the approach to enrolment of study participants (symptomatic versus unselected), but with a decrease in specificity. Overall estimates of specificity were approaching the internationally suggested target of 98% for non-sputum-based TB tests (WHO TTP 2014). Whether lower specificity among inpatients and individuals with lower CD4 can be attributed to misclassification of true positives as false positives due to an imperfect reference standard, or is due to other biological or environmental factors is unclear.

An increased number of studies were included in this updated review compared to the original review on LAM from 2015. However, we found considerable heterogeneity across studies and there were limited data for some sub-group analyses with respect to setting and CD4 count. Most studies used a lower quality reference standard where only sputum was microbiologically tested, and this may have led to misclassification of true-positive results as false-positive results (i.e. reduced specificity estimates). Many studies excluded participants unable to produce sputum, the target population expected to benefit the most from urine-based testing as they cannot have other sputum-based diagnostic testing.

All studies except one were conducted in sub-Saharan Africa, and we wish to underscore a concern about the applicability of the results on the whole outside of sub-Saharan Africa. We further consider
the impact of AlereLAM to be affected by a number of factors, including the health care infrastructure and access to other diagnostic tests, prevalence of MDR-TB (which AlereLAM misses), and rates of empiric TB treatment. The results should, therefore, be interpreted with caution.

Concerning the accuracy of AlereLAM for TB in children living with HIV, there were too few studies and participants to draw conclusions.
Background

Tuberculosis (TB) remains the leading cause of hospitalization and in-hospital deaths among people living with HIV despite the increased access to antiretroviral treatment (ART) (Ford 2016). A systematic review of the prevalence of TB identified at autopsy suggests that, in resource-limited settings, TB is responsible for around 40% of all HIV-related deaths and that TB often was disseminated and undiagnosed at the time of death (Gupta 2015). Globally in 2017, only 51% of the estimated 10.0 million TB cases were notified among people living with HIV (WHO Global Report 2018). However, most death from TB is preventable if TB is detected early and effectively treated.

To improve TB case detection, new diagnostic tools and strategies for systematic screening of people living with HIV is a key component of the World Health Organization’s (WHO) “End TB strategy” (WHO End TB 2014). Non-sputum-based point-of-care TB diagnostic tests are highly desired to narrow the diagnostic gap 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 TB diagnosis that is non-site specific. Urine is further easy to collect and store, and lacks the infection control risks associated with sputum collection. Multiple platforms have been developed to detect lipoarabinomannan (LAM), initially as enzyme-linked immunosorbent (ELISA) assays that were evaluated in several clinical settings (Minion 2011). Later, the lateral flow assay, Alere Determine™ TB LAM Ag assay ‘AlereLAM’, was developed as a simple point-of-care test for diagnosis of active TB in people living with HIV. AlereLAM 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 AlereLAM is recommended in the WHO Policy Guidance, “The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV”, published in 2015 (WHO Lipoarabinomannan Policy Guidance 2015). The guidance was informed by a review of evidence that was subsequently published in the original Cochrane Review of the AlereLAM (Shah 2016). The guidelines recommend that AlereLAM “may be used to assist in the diagnosis of TB in HIV-positive adult inpatients with signs and symptoms of TB (pulmonary and/or extrapulmonary) and a CD4 cell count less than or equal to 100 cells per 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 TB (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 AlereLAM should not be used for general TB screening “owing to suboptimal sensitivity” (WHO Lipoarabinomannan Policy Guidance 2015). The guidelines further suggest that
AlereLAM should be used in combination with existing tests, and not as a replacement test (to existing tests).

Of note, in 2018, preliminary performance characteristics of a second commercially developed lateral flow assay to detect LAM for the diagnosis of TB was announced based on data from frozen biobank specimens (Fujifilm SILVAMP TB LAM, Japan; FujiLAM) (Broger 2018). The test is projected to become commercially available in 2020.

The current systematic review includes published studies evaluating the commercially available AlereLAM assay for diagnosis of active TB disease (pulmonary and extrapulmonary TB) in people living with HIV. Since 2015, additional evidence for the use of AlereLAM has emerged. This updated systematic review will inform the WHO Guideline Development Group if there is evidence to update or modify recommendations on the use of AlereLAM.

**Index test**

The urine-based lateral flow lipoarabinomannan assay AlereLAM is a commercially available point-of-care test for active TB (Alere Determine™ TB LAM Ag, Abbott, Palatine, IL, USA, previous Alere Inc., Waltham, MA, USA). AlereLAM is an immunocapture assay that detects LAM antigen in urine. LAM is a lipopolysaccharide present in mycobacterial cell walls (Brennan 2003), which is released from metabolically active or degenerating bacterial cells during TB disease (Briken 2004). LAM is detectable in urine of people with active TB disease and evaluated for both LAM ELISA and the lateral flow AlereLAM testing platforms (Peter 2010; Lawn 2012; Minion 2011; Shah 2016). The original Cochrane Review of AlereLAM (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 TB involvement, and greater glomerular permeability to allow increased antigen levels in urine (Minion 2011; Lawn 2016). Based on current WHO guidelines, the role of the test can be characterized as a test to be used in combination with existing TB tests.

AlereLAM is performed manually by applying 60 µL of urine to the test strip and incubating at room temperature for 25 minutes (Alere 2017). See Figure 1. The strip is then inspected by eye. The intensity of any visible band on the test strip is graded by comparing it with the intensities of the bands on a manufacturer-supplied reference 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). Some studies prior to January 2014 utilized grade 1 as the threshold for test positivity, while other studies utilized grade 2 as the positivity threshold. After January 2014, the manufacturer revised the reference 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. Under the current manufacturer recommendations
(using the current 4 bands reference card), only bands that are grade 1 or higher are considered positive (Alere 2017). See Appendix 1. Reference card grading of Alere Determine™ TB LAM.

Figure 1. Alere Determine™ TB LAM Ag tests, ‘AlereLAM’
(A) Alere Determine™ TB LAM Ag tests. To the sample pad (white pad marked by the arrow symbols) 60 µL of urine is applied and visualized bands are read 25 minutes later. (B) Reference card accompanying test strips to 'grade' the test result and determine positivity.

Footnote: The reference scale card was changed in 2014. See Appendix 1. Reference card grading of Alere Determine™ TB LAM
PICO questions

We addressed the following PICO questions also listed in Appendix 2. PICO questions.

1. **What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in all HIV-positive adults and children with signs and symptoms of TB?**
   a. in inpatient settings (adults, adolescents and older children)
   b. in outpatient settings (adults, adolescents and older children)
   c. all settings (adults, adolescents and older children)
   d. in inpatient settings (children ≤ 5 years)
   e. in outpatient settings (children ≤ 5 years)
   f. all settings (children ≤ 5 years)

2. **What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in all HIV-positive adults and children irrespective of signs and symptoms of TB?**
   a. in inpatient settings (adults, adolescents and older children)
   b. in outpatient settings (adults, adolescents and older children)
   c. all settings (adults, adolescents and older children)
   d. in inpatient settings (children ≤ 5 years)
   e. in outpatient settings (children ≤ 5 years)
   f. all settings (children ≤ 5 years)

3. **What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in adults with advanced HIV disease irrespective of signs and symptoms of TB?**
   a. in inpatient setting CD4 ≤ 200
   b. in outpatient setting CD4 ≤ 200
   c. in all settings CD4 ≤ 200
   d. in inpatient setting CD4 ≤ 100
   e. in outpatient setting CD4 ≤ 100
   f. in all settings CD4 ≤ 100

4. **Can the use of LF-LAM in HIV-positive adults reduce mortality associated with advanced HIV disease?**
   a. in all settings
   b. in inpatient settings
   c. in outpatient settings
   d. in individuals with CD4 ≤ 200
   e. in inpatient setting CD4 ≤ 200
   f. in outpatient setting CD4 ≤ 200
   g. in individuals with CD4 ≤ 100
   h. in inpatient setting CD4 ≤ 100
   i. in outpatient setting CD4 ≤ 100

**Other questions:** What is the cost and cost-effectiveness of LF-LAM implementation for TB diagnosis, based on review of the published literature?
People living with HIV are at increased risk of TB and may present with symptoms of TB but may also be asymptomatic or have symptoms not routinely associated with TB disease. To estimate accuracy in HIV-positive individuals with signs and symptoms of TB (PICO 1), we combined studies in which presentation with signs and symptoms suggestive of TB 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 TB (PICO 2 and PICO 3), we combined studies that considered all HIV-positive individuals eligible to participate, including both individuals with and individuals without symptoms of TB and refer to these as ‘Studies with unselected participants’.

We reviewed data related to patient-important outcomes, in particular, the impact of AlereLAM implementation on mortality (PICO 4).

A priori we wanted to investigate heterogeneity by clinical setting (inpatient versus outpatient) and by CD4 cell count (CD4 ≤ 200; CD4 ≤ 100) (PICO 1a-f; PICO 2a-f; PICO 3a-f; PICO 4a-i).

Throughout the report, we presented the diagnostic accuracy for AlereLAM in children and adolescents separately from adults.

Economic evaluations of AlereLAM for TB are reported in another document.
Methods

Criteria for considering studies for this review

Types of studies
We included primary studies evaluating the diagnostic accuracy of urine AlereLAM assay for the detection of active TB 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 TB are largely cross-sectional in design but may include some clinical follow-up as part of patient classification. We included randomized controlled studies, cross-sectional studies and observational cohort studies and excluded case-control studies or other study designs. We excluded data reported only in abstracts, reviews, commentaries and editorial notes. We did not include unpublished data.

Participants
We included participants who were adults (15 years and older is considered 'adult' for purpose of TB surveillance) and HIV positive. We included studies in which there was a suspicion of TB among study participants based on the presence of signs and symptoms compatible with TB (studies with symptomatic participants), as well as studies that included participants who presented for medical care irrespective of signs and symptoms of TB (studies with unselected participants). Signs and symptoms of TB include cough, fever, weight loss, and night sweats. Participants who were known with active TB or taking anti-TB drug were not included.

Index tests
We included studies that evaluated Alere Determine™ TB LAM Ag test (Abbott, Palatine, IL, USA, previous Alere Inc., Waltham, MA, USA) ‘AlereLAM’ on urine samples. As of December 2018, AlereLAM was the only commercial lateral flow urine LAM assay available 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 TB disease among people living with HIV, which includes pulmonary and extrapulmonary TB.

**Reference standards**

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

'TB' is defined as a positive *M. tuberculosis* culture or Nucleic Acid Amplification Test (NAAT). 'Not TB ' is defined as a negative *M. tuberculosis* culture and NAAT (if performed).

NAAT tests included: Enhanced Amplified Mycobacterium Tuberculosis Direct Test (E-MTD, GenProbe, San Diego, USA); Amplicor *Mycobacterium tuberculosis* Test (Amplicor, Roche Diagnostics, Basel, Switzerland); COBAS® TaqMan® MTB Test (Roche Diagnostics); GenoType MTBDRplus (HAIN Lifesciences, Nehren, Germany); Xpert® MTB/RIF assay (Cepheid, Sunnyvale, USA); and Xpert® MTB/RIF Ultra.

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 standardised study algorithm. We considered a lower quality reference standard to be one in which only one specimen type was evaluated for TB diagnosis or if there was no algorithm defined to ensure a standardised 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 TB 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 TB cannot produce sputum at all ([Lawn 2013a](#)) or have extrapulmonary TB without pulmonary TB. This means that index test TPs may be misclassified as FPs by sputum culture. Therefore, when evaluating AlereLAM 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 AlereLAM specificity may be underestimated ([Lawn 2015](#)). This misclassification may also lead to underestimation of sensitivity. Increasing the sensitivity of the reference standard by evaluating multiple specimens, including evaluating specimens from sites of disease for extrapulmonary TB, may reduce the number of cases of TB disease incorrectly classified as 'not TB' by culture 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 TB in patients with TB disease. However, our original review found relatively little data using a composite reference standard; found heterogeneity in defining and applying composite reference standards; and found 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).

**Search methods for identification of studies**

We performed literature searches up to 11 May 2018 in the following databases using the search terms reported in Appendix 3. Detailed search strategies: 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](http://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. We further examined reference lists of relevant reviews and studies and searched the WHO websites.

**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.

**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. See Appendix 4. Data collection form. Then two review authors (MS and SB) independently extracted data from each included study on the following characteristics.

- Author, publication year, study design, country(ies), clinical setting (outpatient or inpatient).
- Participants: age, gender, HIV-status, CD4 count, TB 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) and the number of TB cases in the study.
- Number of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) values for the index test.
- 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 2018/2019). In addition, we classified a country as being high burden or not high burden for TB/HIV according to the post-2015 era classification by the WHO (WHO Global Report 2018). We contacted study authors for clarifications on the AlereLAM 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.

**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). QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing (flow and timing domain includes differential verification of TB status for study participants). We assessed all domains for risk of bias and the first three domains for concerns regarding applicability. As recommended, we first developed 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 judgements. We resolved disagreements through discussion or consulted a third review author (KRS).

**Statistical analysis and data synthesis**

We performed descriptive analyses of the characteristics of the included studies using Stata 15 (StataCorp 2017). We used the number of TPs, FPs, FNs, and TNs to calculate the individual study estimates of sensitivity and specificity and their 95% confidence intervals (CI). We presented individual study results graphically by plotting the estimates of sensitivity and specificity (and their 95% CIs) in forest plots using Review Manager (RevMan) (Review Manager).

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. See Appendix 1. Reference card grading of Alere Determine™ TB LAM. 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 AlereLAM for: (I) diagnosis of TB in HIV-positive people with signs and symptoms of TB i.e. 'Studies with symptomatic participants' and (II) diagnosis of TB in HIV-positive people irrespective of signs and symptoms of TB i.e. 'Studies with unselected participants'.

When data were sufficient, we carried out meta-analyses to estimate AlereLAM pooled sensitivity and specificity with a bivariate random-effects model (Chu 2009; Reitsma 2005). This approach allowed us to calculate pooled sensitivity and specificity while dealing with potential sources of variation caused by: (1) imprecision of sensitivity and specificity estimates within individual studies; (2) correlation between sensitivity and specificity across studies; and (3) variation in sensitivity and specificity between studies.

We 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. For the between-study precision we used a gamma distribution with a shape parameter of two and rate parameter of 0.5. This corresponds to a 95% prior credible interval (CrI) for the between-study SD in the log odds of sensitivity or log odds of specificity ranging from roughly 0.29 to 1.44, corresponding to moderate to high values of between-study heterogeneity. Covariance terms followed a uniform prior distribution whose upper and lower limits were determined by the sensitivity of the two tests. We have summarized the models we used (including the prior distributions) and the OpenBUGS programs we used to estimate them in Appendix 7. Statistical approach.

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% 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.

In our original review we evaluated the incremental change in sensitivity and specificity when combining AlereLAM 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 was limited.

**Approach to uninterpretable AlereLAM results**

We excluded uninterpretable test results from the analyses for determination of sensitivity and specificity, but these were very few in numbers across studies.

**Investigations of heterogeneity**

Several PICO questions specifically sought to assess diagnostic accuracy among subgroups. A priori and when data were sufficient, we performed subgroup analyses with the following categorical covariates: clinical setting (inpatient versus outpatient) and CD4 count (CD4 ≤ 200, CD4 ≤ 100).

To further investigate heterogeneity, we performed additional subgroup analyses for CD4 strata CD4 101- 200; CD4 > 200 and; CD4 > 100 as well as by TB prevalence. We investigated heterogeneity for the group of ‘studies with symptomatic participants’ separately from the group of ‘studies with unselected participants’.

**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 TB symptoms.

Additional analyses

Investigations of heterogeneity, all studies combined

We performed several additional post-hoc analyses to inform interpretation of findings. We assessed performance for the two groups of studies combined i.e. studies with symptomatic participants combined with studies with unselected participants. We did this analysis overall and stratified by inpatients, outpatients, and CD4 strata (CD4 $\leq 200$; CD4 $\leq 100$; CD4 $> 200$ and; CD4 $> 100$ cells per $\mu$L.)

Impact on mortality and other patient-important outcomes

Data that directly address the impact of test implementation on patient-important outcomes, such as mortality, are important for patients, decision makers, and the wider TB community. Our diagnostic test accuracy systematic review was not designed to answer questions on the impact of the test on patient outcomes, since a different methodology and separate search strategy would have been required. Nonetheless, we carried out additional efforts as follows. The primary reviewer authors (MS and SB) identified full text articles that included data on health impact. Another review author (RRN) examined all 15 articles included in this updated review, in addition to the other articles excluded during full text screening, and articles included in the original Cochrane Review (Shah 2016), to determine whether studies reported data on impact. We also looked for information on the association of test positivity and patient outcomes.

To evaluate impact data, we developed a standardized data extraction form and piloted this on two of the included studies. Based on the pilot, we finalized the form. See Appendix 5. Data collection form, impact data. Subsequently, one review author (RRN) extracted data from each included study on the following characteristics using Excel to collect and manage study data.

- Author, publication year, study design, country(ies), clinical setting (outpatient or inpatient), number enrolled and analysed.
- Participants: age, HIV-status, presence of symptoms (symptoms versus unselected).
- Mode of mortality assessment, type of mortality (all-cause versus TB-related), timing of mortality assessment.
- AlereLAM grade and use of old versus new reference card, timing of AlereLAM.
- Mortality analysis metrics used (absolute risk reduction (ARR), (adjusted) Hazard Ratio (HR) or Kaplan Meier, (adjusted) odds ratio).
- Comparator groups analysed.
- Mortality in the intervention group, mortality in the control group (for randomized controlled trials).
- Mortality in AlereLAM positive, mortality in AlereLAM negative.
- Mortality in AlereLAM positive patients with confirmed TB, mortality in AlereLAM negative patients with confirmed TB.
- Mortality in AlereLAM positive patients with inconclusive or non-TB diagnosis, mortality in AlereLAM negative patients with inconclusive or non-TB diagnosis.
- Mortality data stratified by CD4 count.
- Time to diagnosis, time to treatment.
- Other outcomes assessed in the study.

For two randomized controlled trials, we assessed risk of bias using the Cochrane tool in RevMan (Review Manager). Then we narratively described the effect of AlereLAM implementation (that is, AlereLAM used versus AlereLAM not used to guide treatment) on patient-important outcomes including time to diagnosis and treatment, disease severity, and mortality. For mortality as a critical outcome, we performed a fixed-effect meta-analysis, including data from the two randomized trials, to estimate the pooled risk ratio (95% CI) for mortality. We thought it appropriate to use a fixed-effect approach because the estimates of the effect of the intervention in the different studies appeared similar, the differences between them being small enough to be explained by chance. However, a fixed-effect approach does not enable a measure of between-study heterogeneity.

**Association between AlereLAM test positivity with patient-important outcomes**

For diagnostic accuracy studies identified as having data on patient-important outcomes, we recorded whether there was an association between AlereLAM results and patient-important outcomes, including time to diagnosis and treatment, disease severity, and mortality. Of note, these studies did not use AlereLAM to guide treatment and are therefore not considered as impact studies.

**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 for intervention studies as recommended using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Balshem 2011). Although, the approach is similar for diagnostic studies, we describe it in detail here (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, index test, and outcomes 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 TB and made judgements on imprecision from these calculations. We also considered whether the number of participants included in the analysis was less than the number generated by a conventional sample size calculation for a single adequately powered study (optimal information size).
- Publication bias: we rated publication bias as undetected (not serious) for several reasons including the comprehensiveness of the literature search and extensive outreach to TB researchers to identify studies.
Results

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 and Cochrane Review (WHO Lipoarabinomannan Policy Guidance 2015; Shah 2016) that met the refined inclusion criteria, and nine new studies identified in the updated search. Of six previously included studies, three were excluded because they did not use the currently recommended threshold for test positivity (Lawn 2012a; Balcha 2014; Drain 2014a); one abstract was included as an updated published manuscript (Lawn 2014a), one abstract remained unpublished (Andrews 2014), and one abstract was published but did not provide diagnostic accuracy data (Drain 2014e). Eight studies evaluated the accuracy of AlereLAM for TB diagnosis in participants with signs and symptoms suggestive of TB. Seven studies evaluated the accuracy of AlereLAM for diagnosis of unselected participants that may or may not have had TB signs and symptoms at enrolment. See Figure 2.
**Figure 2. Flow of studies in the review.**
Methodological quality of included studies

Figure 3 and Figure 4 show the quality assessment of the 15 included studies.

Studies with symptomatic participants

Eight studies were included that evaluated AlereLAM for TB diagnosis among symptomatic participants suspected of TB. 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 if they did not have a full set of complement reference standard results i.e. had any sample with a missing Xpert MTB/RIF result or a contaminated culture result in the absence of a positive result (Huerga 2017); (4) the study only included patients suspected of extrapulmonary TB and excluded patients suspected of pulmonary TB (Juma 2017); (5) the study only included participants with pericardial effusion and suspected TB and excluded participants suspected of other forms of TB (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 study participants did not resemble people with presumed HIV/TB 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 with a high concern of applicability as the study used grade 2 (on the current reference scale card) as the test positivity threshold, as opposed to the current manufacturer recommendation to use grade 1 (on the updated reference card) to define test positivity (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 2012; Peter 2016). We deemed three studies at high concern for applicability as they lacked a study or protocol directed testing (Peter 2012; Peter 2016; Pandie 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 TB.

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 2012) or because not
all participants were included in the two-by-two tables (Pandie 2016; Huerga 2017). We judged the remaining studies to be at low risk of bias because all participants received the index test and the same reference standard, and none of the participants enrolled in the studies were excluded from analysis.

**Studies with unselected participants**

Seven studies contributed data for the purpose of evaluating AlereLAM for TB diagnosis among unselected participants that may or may not have TB symptoms (Figure 4). 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; Floridia 2017 LaCourse 2016; Drain 2015). 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 AlereLAM, 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 2015; Floridia 2017; LaCourse 2016; Thit 2017). We judged these studies to be of low concern in terms of applicability. In one study it was unclear if the reference standard results was interpreted without knowledge of the index test result and we judge an unclear concern of applicability.

In the flow and timing domain, we considered two studies (29%) to be at high risk of bias because the study collected specimens for index and reference standard tests up to six months apart (Hanifa 2016; Thit 2017) and Hanifa 2016 excluded clinical TB cases from analysis. We considered the remaining five studies (71%) to be of 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.
**Figure 3. Risk of bias and applicability concerns graph.**
Review authors' judgements about each domain presented as percentages across included studies.

![Risk of Bias and Applicability Concerns Graph](image)

**Figure 4. Risk of bias and applicability concerns summary.**
Review authors' judgements about each domain for each included study. (A) Studies with symptomatic participants. (B) Studies with unselected participants.

![Risk of Bias and Applicability Concerns Summary](image)
Findings

The 15 included studies involved 6814 participants, 1761 (26%) with TB. Eight of the studies evaluated the accuracy of AlereLAM for TB diagnosis in participants with signs and symptoms suggestive of TB involving 3449 participants, 1277 (37%) with TB. Seven studies evaluated the accuracy of AlereLAM for diagnosis of unselected participants that may or may not have had TB signs and symptoms at enrolment involving 3365 participants, 439 (13%) with TB.

All studies were performed in high TB/HIV burden countries and classified as low-income or middle-income countries. We noted substantial differences in the studies for the following characteristics: type of study ('studies with symptomatic participants' and 'studies with unselected participants'); setting (inpatients versus outpatients); median CD4 cell count; TB prevalence; inclusion and exclusion of participants based on whether or not they could produce sputa; and whether patients were evaluated for pulmonary TB, extrapulmonary TB or both. The key study characteristics are summarised by study in Appendix 8: Characteristics.

Most studies reported that a valid AlereLAM result was obtained on the first attempt for all tests. Few uninterpretable test results (< 1%) were reported in three studies (Peter 2012; Peter 2015; Peter 2016).

Table 1 presents pooled sensitivity and specificity results for AlereLAM against a microbiological reference standard grouped by the type of study 'TB diagnosis among symptomatic participants' and 'TB diagnosis among unselected participants'.
<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Studies (total participants)</th>
<th>Participants with TB (%)</th>
<th>Sensitivity (95% CrI)</th>
<th>Specificity (95% CrI)</th>
<th>Sensitivity (95% CrI)</th>
<th>Specificity (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall accuracy</td>
<td>8 studies (3449)</td>
<td>1277 (37%)</td>
<td>42% (31 to 55)</td>
<td>91% (85 to 95)</td>
<td>7 studies (3365)</td>
<td>432 (13%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>35% (22 to 50)</td>
</tr>
<tr>
<td>By setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient</td>
<td>6 studies (2253)</td>
<td>868 (39%)</td>
<td>52% (40 to 64)</td>
<td>87% (78 to 93)</td>
<td>3 studies (537)</td>
<td>159 (30%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>62% (41 to 83)</td>
</tr>
<tr>
<td>Outpatient</td>
<td>4 studies (1196)</td>
<td>409 (34%)</td>
<td>29% (17 to 47)</td>
<td>96% (91 to 99)</td>
<td>6 studies (2828)</td>
<td>273 (10%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31% (18 to 47)</td>
</tr>
<tr>
<td>By CD4 cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 &gt; 200</td>
<td>3 studies (738)</td>
<td>163 (22%)</td>
<td>16% (8 to 31)</td>
<td>94% (81 to 97)</td>
<td>1 study (156)</td>
<td>11 (7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>CD4 ≤ 200</td>
<td>4 studies (1825)</td>
<td>722 (40%)</td>
<td>45% (31 to 61)</td>
<td>89% (77 to 94)</td>
<td>2 studies (706)</td>
<td>82 (12%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26% (9 to 56)</td>
</tr>
<tr>
<td>CD4 &gt; 100</td>
<td>4 studies (1519)</td>
<td>425 (28%)</td>
<td>17% (10 to 27)</td>
<td>95% (89 to 98)</td>
<td>4 studies (952)</td>
<td>115 (12%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20% (10 to 35)</td>
</tr>
<tr>
<td>CD4 ≤ 100</td>
<td>4 studies (1239)</td>
<td>512 (41%)</td>
<td>54% (38 to 69)</td>
<td>88% (77 to 94)</td>
<td>3 studies (417)</td>
<td>130 (31%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47% (40 to 64)</td>
</tr>
<tr>
<td>CD4 101-200</td>
<td>4 studies (586)</td>
<td>210 (36%)</td>
<td>24% (14 to 38)</td>
<td>90% (77 to 96)</td>
<td>1 study (103)</td>
<td>13 (13%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>By CD4 and setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 ≤ 200</td>
<td>2 studies (1009)</td>
<td>348 (34%)</td>
<td>54% (34 to 73)</td>
<td>80% (58 to 91)</td>
<td>1 study (54)</td>
<td>14 (26%)</td>
</tr>
<tr>
<td>inpatients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>CD4 ≤ 100</td>
<td>2 studies (734)</td>
<td>270 (37%)</td>
<td>61% (40 to 78)</td>
<td>81% (61 to 91)</td>
<td>2 studies (200)</td>
<td>84 (42%)</td>
</tr>
<tr>
<td>inpatients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>57% (33 to 79)</td>
</tr>
<tr>
<td>CD4 101-200</td>
<td>2 studies (275)</td>
<td>78 (28%)</td>
<td>32% (16 to 57)</td>
<td>81% (55 to 92)</td>
<td>1 study (9)</td>
<td>4 (44%)</td>
</tr>
<tr>
<td>inpatients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>CD4 ≤ 200</td>
<td>1 study (249)</td>
<td>97 (39%)</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>2 studies (652)</td>
<td>68 (10%)</td>
</tr>
<tr>
<td>outpatients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21% (8 to 48)</td>
</tr>
<tr>
<td>CD4 ≤ 100</td>
<td>1 study (121)</td>
<td>48 (40%)</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>2 studies (217)</td>
<td>46 (21%)</td>
</tr>
<tr>
<td>outpatients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40% (20 to 64)</td>
</tr>
<tr>
<td>CD4 101-200</td>
<td>1 study (128)</td>
<td>51 (40%)</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>1 study (94)</td>
<td>9 (10%)</td>
</tr>
<tr>
<td>outpatients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

Abbreviations: CrI: credible interval; AlereLAM: Alere Determine™ TB lipoarabinomannan assay; TB: tuberculosis.

*Bjerrum 2015, Sensitivity 27% (6% to 61%); Specificity 99% (96% to 100%); *Bjerrum 2015, Sensitivity 38% (14% to 68%); Specificity 99% (94% to 100%); *Bjerrum 2015, Sensitivity 64% (35% to 87%); Specificity 82% (67% to 93%); *Bjerrum 2015, Sensitivity 75% (19% to 99%); Specificity 100% (48% to 100%); *Bjerrum 2015, Sensitivity 22% (3% to 60%); Specificity 99% (94% to 100%); *Peter 2015, Sensitivity 24% (16% to 33%); Specificity 94% (89% to 97%); *Peter 2015, Sensitivity 30% (18% to 46%); Specificity 93% (85% to 98%); *Peter 2015, Sensitivity 18% (8% to 31%); Specificity 95% (87% to 99%).
PICO 1: What is the diagnostic accuracy of AlereLAM for TB diagnosis in HIV-positive adults with signs and symptoms of TB?

Of the 15 included studies, eight evaluated the accuracy of AlereLAM for TB diagnosis in participants with signs and symptoms suggestive of TB. The suggestive signs and symptoms of TB varied from study to study, but were often based on any of cough, fever, weight loss, or night sweats. The TB prevalence ranged from 29% to 63%. Two studies were conducted exclusively among patients with presumed extrapulmonary TB: Juma 2017 and Pandie 2016. Four studies were conducted exclusively in an inpatient setting; two studies exclusively in an outpatient setting, and two studies in both inpatient and outpatient settings. 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 81-139 cells per µL) compared to studies evaluating outpatients (median CD4 was 168-210 cells per µL). See Appendix 8. Characteristics of Included Studies.

Results for children are presented in Appendix 9: Diagnostic accuracy of AlereLAM among HIV-positive children, summary.

PICO 1a: Accuracy, in inpatient settings

Six studies were conducted among inpatients involving 2253 participants, 868 (39%) with TB (Huerga 2017; Juma 2017; Nakijingi 2014; Pandie 2016; Peter 2012; Peter 2016). Sensitivity estimates ranged from 33% to 69% and specificity estimates ranged from 75% to 100%. The pooled sensitivity and specificity (95% CrI) among inpatients were 52% (40% to 64%) and 87% (78% to 93%). See Figure 5. The highest sensitivity (69%) was reported by Huerga 2017 with a relatively low specificity (78%). This study included inpatients who were severely ill or with CD4 < 200 cell per µL (median CD4 109) and low BMI. The study did not include microbiological or histological evaluation of extrapulmonary specimens for TB in their reference standard, which may have led to LAM-positive participants with extrapulmonary TB being misclassified as ‘false-positive’ and lowered specificity. Pandie 2016 reported the lowest sensitivity (33%) and a specificity of 100%. This study differed from others by evaluating accuracy only for pericardial TB. The authors excluded a number of participants in the analysis for unknown reasons and reported TN as ‘2’ and FP as ‘0’ that may have inflated specificity. In the original review, the pooled sensitivity and specificity among inpatients were 53% (38% to 70%) and 90% (73% to 96%) (four studies, 1299 participants). (Shah 2016).

PICO 1b: Accuracy, in outpatient settings

Four studies were conducted among outpatients involving 1196 participants, 409 (34%) with TB (Drain 2016; Huerga 2017; Nakijingi 2014; Peter 2015). Sensitivity estimates ranged from 18% to 58% and specificity estimates ranged from 93% to 99%. The highest sensitivity (58%) was reported by Huerga 2017 where outpatients included were severely ill, or had a CD4 < 200 cell per µL, or a
low Body Mass Index below 17Kg/m². Pooled sensitivity and specificity were 29% (17% to 47%) and 96% (91% to 99%). See Figure 5.

Pooled estimates were not previously calculated for outpatients due to lack of data in the original Cochrane Review (Shah 2016).

**Figure 5. Diagnostic accuracy in adults with signs and symptoms, by health care setting**

Forest plots and meta-analysis of AlereLAM sensitivity and specificity for TB against a microbiological reference standard, by health care setting.

### Pooled estimates were not previously calculated for outpatients due to lack of data in the original Cochrane Review (Shah 2016).

#### Symptomatic adults, inpatients

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huerga 2017</td>
<td>73</td>
<td>17</td>
<td>33</td>
<td>60</td>
<td>0.69 [0.59, 0.78]</td>
<td>0.78 [0.67, 0.87]</td>
<td>0.78 [0.67, 0.87]</td>
<td>0.78 [0.67, 0.87]</td>
</tr>
<tr>
<td>Juma 2017</td>
<td>15</td>
<td>3</td>
<td>7</td>
<td>42</td>
<td>0.68 [0.45, 0.86]</td>
<td>0.93 [0.82, 0.99]</td>
<td>0.93 [0.82, 0.99]</td>
<td>0.93 [0.82, 0.99]</td>
</tr>
<tr>
<td>Peter 2012</td>
<td>58</td>
<td>31</td>
<td>58</td>
<td>94</td>
<td>0.50 [0.41, 0.59]</td>
<td>0.75 [0.67, 0.82]</td>
<td>0.75 [0.67, 0.82]</td>
<td>0.75 [0.67, 0.82]</td>
</tr>
<tr>
<td>Nakinyingi 2014</td>
<td>134</td>
<td>19</td>
<td>132</td>
<td>287</td>
<td>0.46 [0.40, 0.53]</td>
<td>0.94 [0.90, 0.96]</td>
<td>0.94 [0.90, 0.96]</td>
<td>0.94 [0.90, 0.96]</td>
</tr>
<tr>
<td>Peter 2016</td>
<td>156</td>
<td>94</td>
<td>186</td>
<td>736</td>
<td>0.46 [0.40, 0.51]</td>
<td>0.89 [0.86, 0.91]</td>
<td>0.89 [0.86, 0.91]</td>
<td>0.89 [0.86, 0.91]</td>
</tr>
<tr>
<td>Pandie 2016</td>
<td>12</td>
<td>0</td>
<td>24</td>
<td>2</td>
<td>0.33 [0.19, 0.51]</td>
<td>1.00 [0.16, 1.00]</td>
<td>1.00 [0.16, 1.00]</td>
<td>1.00 [0.16, 1.00]</td>
</tr>
</tbody>
</table>

#### Symptomatic adults, outpatients

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huerga 2017</td>
<td>29</td>
<td>2</td>
<td>23</td>
<td>40</td>
<td>0.58 [0.43, 0.72]</td>
<td>0.95 [0.88, 0.99]</td>
<td>0.95 [0.88, 0.99]</td>
<td>0.95 [0.88, 0.99]</td>
</tr>
<tr>
<td>Drain 2016</td>
<td>13</td>
<td>1</td>
<td>44</td>
<td>32</td>
<td>0.23 [0.13, 0.36]</td>
<td>0.97 [0.84, 1.00]</td>
<td>0.97 [0.84, 1.00]</td>
<td>0.97 [0.84, 1.00]</td>
</tr>
<tr>
<td>Peter 2015</td>
<td>42</td>
<td>27</td>
<td>140</td>
<td>361</td>
<td>0.23 [0.17, 0.29]</td>
<td>0.93 [0.90, 0.95]</td>
<td>0.93 [0.90, 0.95]</td>
<td>0.93 [0.90, 0.95]</td>
</tr>
<tr>
<td>Nakinyingi 2014</td>
<td>22</td>
<td>2</td>
<td>99</td>
<td>322</td>
<td>0.18 [0.12, 0.26]</td>
<td>0.99 [0.98, 1.00]</td>
<td>0.99 [0.98, 1.00]</td>
<td>0.99 [0.98, 1.00]</td>
</tr>
</tbody>
</table>

#### Type of analysis

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Symptomatic participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Studies (total participants)</td>
</tr>
<tr>
<td>Inpatient</td>
<td>6 studies (2253)</td>
</tr>
<tr>
<td>Outpatient</td>
<td>4 studies (1196)</td>
</tr>
</tbody>
</table>

**PICO 1c: Overall accuracy, all settings**

For the analysis of the overall accuracy of AlereLAM in HIV-positive adults with signs and symptoms of TB, eight studies provided data for 3449 participants, including 1277 (37%) TB patients, (Drain 2016; Huerga 2017; Juma 2017; Pandie 2016; Peter 2016; Nakinyingi 2014; Peter 2012; Peter 2015). Sensitivity estimates ranged from 23% to 68%, and specificity estimates from 75% to 100%. The pooled sensitivity and specificity (95% CrI) were 42% (31% to 55%) and 91% (85% to 95%).

See Figure 6. Juma 2017 evaluated diagnostic accuracy for extrapulmonary TB (all forms) exclusively and had sensitivity of 68%. Pandie 2016 evaluated accuracy for pericardial TB and found sensitivity of 33%. Sensitivity was lowest in the studies by Peter 2015 and Drain 2016. Differences between these studies and the other studies in this analysis were the setting (outpatient only), focus on pulmonary TB (no extrapulmonary samples were taken), and exclusion of participants unable to produce sputum. In particular, Drain 2016 included smear-negative participants with presumed TB and a small number of HIV-negative participants. In addition, this study excluded participants with a
low Karnofsky score in order to target relatively well outpatients, where smear-negative TB disease is often seen. Specificity was lowest for Peter 2012, 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 a specificity of 100%, but excluded participants from specificity analysis as mentioned above. For comparison, the pooled sensitivity and specificity in the original review were 45% (29% to 63%) and 92% (80% to 97%) based on five studies and 2313 participants (Shah 2016).

Figure 6. Diagnostic accuracy in adults with signs and symptoms, all settings
Forest plots and meta-analysis of AlereLAM sensitivity and specificity for TB against a microbiological reference standard, overall.

![Forest plots and meta-analysis of AlereLAM sensitivity and specificity for TB against a microbiological reference standard, overall.](image)

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juma 2017</td>
<td>15</td>
<td>3</td>
<td>7</td>
<td>42</td>
<td>0.68 [0.45, 0.86]</td>
<td>0.83 [0.82, 0.99]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haerga 2017</td>
<td>102</td>
<td>19</td>
<td>54</td>
<td>100</td>
<td>0.65 [0.57, 0.73]</td>
<td>0.84 [0.76, 0.90]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter 2012</td>
<td>58</td>
<td>31</td>
<td>58</td>
<td>94</td>
<td>0.50 [0.41, 0.59]</td>
<td>0.75 [0.67, 0.82]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter 2016</td>
<td>156</td>
<td>94</td>
<td>186</td>
<td>736</td>
<td>0.46 [0.40, 0.51]</td>
<td>0.89 [0.86, 0.91]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakifyngi 2014</td>
<td>136</td>
<td>21</td>
<td>231</td>
<td>609</td>
<td>0.37 [0.32, 0.42]</td>
<td>0.97 [0.95, 0.98]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pandie 2016</td>
<td>12</td>
<td>0</td>
<td>24</td>
<td>2</td>
<td>0.33 [0.19, 0.51]</td>
<td>1.00 [0.16, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dlam 2016</td>
<td>13</td>
<td>1</td>
<td>44</td>
<td>32</td>
<td>0.23 [0.13, 0.36]</td>
<td>0.97 [0.84, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter 2015</td>
<td>41</td>
<td>27</td>
<td>140</td>
<td>361</td>
<td>0.23 [0.17, 0.29]</td>
<td>0.92 [0.90, 0.95]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Additional investigations of heterogeneity**

**CD4 count**

Accuracy stratified by CD4 > 200 cells per µL and ≤ 200 cells per µL

Three studies evaluated participants with CD4 > 200 cells per µL. (Nakifyngi 2014; Peter 2012; Peter 2016). Sensitivity estimates ranged from 9% to 27% and specificity estimates ranged from 83% to 99%. 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% (Nakifyngi 2014; Peter 2012; Peter 2015; Peter 2016). See Figure 7. The pooled sensitivity (95% CrI) was higher among participants with CD4 ≤ 200 cells per µL at 45% (31% to 61%) (1825 participants; 40% with TB) versus 16% (8% to 31%) among those with CD4 > 200 cells per µL (738 participants; 22% with TB). 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. 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%) (two studies, 1009 participants; 34% with TB (Peter 2012; Peter 2016). Only one study reported data for outpatients with CD4 ≤ 200 cells per µL Peter 2015.
Figure 7. Diagnostic accuracy in adults with signs and symptoms, by CD4 count.

Symptomatic adults, CD4 > 200, all settings

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter 2012</td>
<td>7</td>
<td>5</td>
<td>19</td>
<td>24</td>
<td>0.27 [0.12, 0.48]</td>
<td>0.83 [0.64, 0.94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter 2016</td>
<td>7</td>
<td>14</td>
<td>42</td>
<td>197</td>
<td>0.14 [0.06, 0.27]</td>
<td>0.93 [0.89, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakinyingi 2014</td>
<td>8</td>
<td>5</td>
<td>80</td>
<td>330</td>
<td>0.06 [0.04, 0.17]</td>
<td>0.99 [0.97, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Symptomatic adults, CD4 ≤ 200, all settings

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter 2012</td>
<td>136</td>
<td>74</td>
<td>131</td>
<td>495</td>
<td>0.51 [0.45, 0.57]</td>
<td>0.87 [0.84, 0.90]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakinyingi 2014</td>
<td>128</td>
<td>15</td>
<td>149</td>
<td>275</td>
<td>0.46 [0.40, 0.52]</td>
<td>0.95 [0.92, 0.97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter 2015</td>
<td>23</td>
<td>9</td>
<td>74</td>
<td>143</td>
<td>0.24 [0.16, 0.33]</td>
<td>0.94 [0.89, 0.97]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Symptomatic adults, CD4 > 100, all settings

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter 2016</td>
<td>20</td>
<td>30</td>
<td>81</td>
<td>349</td>
<td>0.19 [0.12, 0.28]</td>
<td>0.92 [0.89, 0.95]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pandie 2016</td>
<td>5</td>
<td>0</td>
<td>22</td>
<td>4</td>
<td>0.19 [0.06, 0.38]</td>
<td>1.00 [0.40, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter 2015</td>
<td>23</td>
<td>17</td>
<td>103</td>
<td>257</td>
<td>0.18 [0.12, 0.26]</td>
<td>0.94 [0.90, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakinyingi 2014</td>
<td>20</td>
<td>8</td>
<td>149</td>
<td>429</td>
<td>0.12 [0.07, 0.18]</td>
<td>0.98 [0.96, 0.99]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Accuracy stratified by CD4 > 100 cells per µL and ≤ 100 cells per µL

Four studies evaluated participants with CD4 > 100 cells per µL, (Nakinyingi 2014; Pandie 2016; Peter 2015; Peter 2016). Sensitivity estimates ranged from 12% to 19% and specificity estimates ranged from 92% to 100%. See Figure 7. 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% (Nakinyingi 2014; Pandie 2016; Peter 2012; Peter 2015; Peter 2016). One study (Pandie 2016)
had no estimable specificity, as they reported zero TN. The pooled sensitivity (95% CrI) was higher among participants with CD4 ≤ 100 cells per µL at 54% (38% to 69%) (1239 participants; 41% with TB) versus 17% (10% to 27%), (1519 participants; 28% with TB) among those with CD4 > 100 cells per µL. 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. When we limited the analysis to studies involving inpatients with CD4 ≤ 100 cells per µL (Peter 2012; Peter 2016), the pooled sensitivity and specificity were 61% (40% to 78%) and 81% (61% to 91%). Pandie 2016 reported a sensitivity of 50% (95%CI, 21% to 79%) among inpatients with CD4 ≤ 100 cells per µL, but specificity was not estimable and therefore not included in the meta-analysis. Only one study reported data for outpatients with CD4 ≤ 100 cells per µL Peter 2015.

We observed that AlereLAM 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 101-199 to 54% (38% to 69%) in patients with CD4 ≤ 100. See Figure 8. Also, we observed that a majority of participants contributing data for the CD4 ≤ 200 cells per µL stratum (1825 participants) were participants with CD4 ≤ 100 cells per µL (1239 participants). See Table 1.

**TB prevalence**

The median prevalence of TB in studies with symptomatic participants was 43% (IQR 32% to 60%). In secondary analysis by TB prevalence, we found that pooled sensitivity and specificity for symptomatic participants in settings with TB prevalence of greater than 43% were 44% (27% to 62%) and 86% (73% to 94%) and 39% (21% to 63%) and 95% (89% to 97%) in settings with a TB prevalence less than 43%. We note that no studies had a TB prevalence of less than 29%.

**Sensitivity analyses**

When we included all studies with more than 80% symptomatic participants, two studies were re-assigned from 'studies of unselected adults' to 'studies of symptomatic participants' Bjerrum 2015; Lawn 2017. In comparison with estimates without re-classification, pooled sensitivity remained at 42% (33% to 52%) and specificity changed to 93% (88% to 96%), (10 studies, 4331 participants). When we limited the studies to those with low risk of bias for patient selection pooled sensitivity increased to 48% (29% to 67%) and specificity dropped to 82% (61% to 92%), (Peter 2012; Peter 2016, 1413 participants). We did not have enough studies to do sensitivity analysis including only studies with low risk of bias in the reference standard domain. Limiting studies to those that used fresh urine samples (four studies) rather than stored urine sample increased sensitivity to 52% (38% to 68%) with specificity remaining at 91%.
Figure 8. Plot by CD4 of diagnostic accuracy in adults with signs and symptoms. (A) Sensitivity by CD4 strata; (B) Specificity by CD4 strata. The circle represents the pooled estimates (median), with bars representing 95% credible intervals.
PICO 2: What is the diagnostic accuracy of AlereLAM for TB diagnosis in HIV-positive adults irrespective of signs and symptoms for TB?

Of the 15 studies included, seven studies evaluated the accuracy of AlereLAM for diagnosis in participants irrespective of sign and symptoms (‘unselected participants’). The TB prevalence varied from 1% to 33%. 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). Four studies were carried out in an inpatient setting, one study exclusively in an inpatient setting and two studies in both inpatient and outpatient settings. The median CD4 cell count across studies of unselected adults ranged from 111 to 437 cells per µL across studies. See Appendix 8. Characteristics of Included Studies.

PICO 2a: Accuracy, in inpatient settings

We identified three studies that included inpatients involving 537 participants, 159 (30%) with TB (Bjerrum 2015; Lawn 2017; Thit 2017). Sensitivity estimates ranged from 39% to 88% and specificity estimates ranged from 39% to 99%. The pooled sensitivity and specificity (95% CrI) among inpatients were 62% (41% to 83%) and 84% (48% to 96%). See Figure 9. 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 TB. They reported that 41 of the inpatients (76%) were symptomatic at enrolment with a median CD4 of 96 (IQR 37-277) cells per µL, which is comparable to evaluation of AlereLAM in a population with advanced HIV disease. The study differed from other studies for reasons listed under PICO 2c.

Pooled estimates were not calculated in the original review (Shah 2016) for studies with unselected inpatients due to lack of data.

PICO 2b: Accuracy, in outpatient settings

Outpatients

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

Pooled estimates were not calculated in the original review (Shah 2016) for outpatients due to lack of data.
Figure 9. Diagnostic accuracy in adults irrespective of signs and symptoms, by setting
Forest plots and meta-analysis of AlereLAM sensitivity and specificity for TB against a microbiological reference standard, by clinical setting.

Unselected adults, inpatients

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thit 2017</td>
<td>7</td>
<td>28</td>
<td>1</td>
<td>18</td>
<td>0.88 [0.47, 1.00]</td>
<td>0.39 [0.25, 0.55]</td>
</tr>
<tr>
<td>Bjerrum 2015</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td>47</td>
<td>0.67 [0.38, 0.88]</td>
<td>0.85 [0.73, 0.94]</td>
</tr>
<tr>
<td>Lawn 2017</td>
<td>53</td>
<td>3</td>
<td>83</td>
<td>274</td>
<td>0.39 [0.31, 0.48]</td>
<td>0.99 [0.97, 1.00]</td>
</tr>
</tbody>
</table>

Unselected adults, outpatients

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thit 2017</td>
<td>29</td>
<td>137</td>
<td>17</td>
<td>280</td>
<td>0.63 [0.48, 0.77]</td>
<td>0.67 [0.62, 0.72]</td>
</tr>
<tr>
<td>Bjerrum 2015</td>
<td>14</td>
<td>13</td>
<td>26</td>
<td>346</td>
<td>0.35 [0.21, 0.52]</td>
<td>0.96 [0.94, 0.98]</td>
</tr>
<tr>
<td>Florida 2017</td>
<td>26</td>
<td>8</td>
<td>64</td>
<td>874</td>
<td>0.29 [0.20, 0.39]</td>
<td>0.99 [0.98, 1.00]</td>
</tr>
<tr>
<td>Drain 2015</td>
<td>15</td>
<td>16</td>
<td>39</td>
<td>250</td>
<td>0.28 [0.16, 0.42]</td>
<td>0.94 [0.90, 0.97]</td>
</tr>
<tr>
<td>Hanifa 2016</td>
<td>3</td>
<td>5</td>
<td>37</td>
<td>363</td>
<td>0.07 [0.02, 0.20]</td>
<td>0.99 [0.97, 1.00]</td>
</tr>
<tr>
<td>LaCourse 2016</td>
<td>0</td>
<td>13</td>
<td>3</td>
<td>250</td>
<td>0.00 [0.00, 0.48]</td>
<td>0.95 [0.92, 0.97]</td>
</tr>
</tbody>
</table>

PICO 2c: Overall accuracy, all settings

For the analysis of the overall accuracy of AlereLAM in HIV-positive adults irrespective of signs and symptoms of TB seven studies provided data for 3365 participants; 439 (13%) with TB (Bjerrum 2015; Drain 2015; Florida 2017; Hanifa 2016; LaCourse 2016; Lawn 2017; Thit 2017). The pooled sensitivity and specificity (95% CrI) were 35% (22% to 50%) and 95% (89% to 98%). Sensitivity estimates ranged from 0% to 67%, and specificity estimates from 64% to 99%. See Figure 10. 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 TB prevalence (1%), and d) a high median CD4 cell count (437 cells per µL). Specificity was lowest (64%) for Thit 2017 that also reported the highest sensitivity (67%). This study used the new reference scale card with 4 bands and 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 mm³ and 33% were symptomatic at enrolment. The study evaluated sputum samples only and allowed a follow-up for 6 months from AlereLAM testing at enrolment to final classification of participants as ‘TB’ or
‘Not TB’ cases. Thit 2017 differed from the other studies by being conducted in Myanmar, and is the only study included in this review that evaluated AlereLAM in a setting outside sub-Saharan Africa.

The pooled sensitivity and specificity of the original review were 30% (20% to 43%) and 94% (86% to 97%) based on three studies and 1055 participants (reported at grade 1 on the old reference scale card with five bands) (Shah 2016).

**Figure 10. Diagnostic accuracy in adults irrespective of signs and symptoms, all settings**

Forest plots and meta-analysis of AlereLAM sensitivity and specificity for TB against a microbiological reference standard, overall.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thit 2017</td>
<td>36</td>
<td>165</td>
<td>18</td>
<td>298</td>
<td>0.67 [0.53, 0.79]</td>
<td>0.64 [0.60, 0.69]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bjerrum 2015</td>
<td>24</td>
<td>21</td>
<td>31</td>
<td>393</td>
<td>0.44 [0.30, 0.58]</td>
<td>0.95 [0.92, 0.97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawn 2017</td>
<td>53</td>
<td>3</td>
<td>83</td>
<td>274</td>
<td>0.39 [0.31, 0.48]</td>
<td>0.99 [0.97, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida 2017</td>
<td>26</td>
<td>8</td>
<td>64</td>
<td>874</td>
<td>0.29 [0.20, 0.39]</td>
<td>0.99 [0.98, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drain 2015</td>
<td>15</td>
<td>16</td>
<td>39</td>
<td>250</td>
<td>0.28 [0.16, 0.42]</td>
<td>0.94 [0.90, 0.97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haifa 2016</td>
<td>3</td>
<td>5</td>
<td>37</td>
<td>363</td>
<td>0.07 [0.02, 0.20]</td>
<td>0.99 [0.97, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LaCourse 2016</td>
<td>0</td>
<td>13</td>
<td>3</td>
<td>250</td>
<td>0.00 [0.00, 0.71]</td>
<td>0.95 [0.92, 0.97]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Footnote: The proportion of symptomatic participants included ranged from 19% in LaCourse 2016 to 91% in Bjerrum 2015 and in Lawn 2017. See Appendix 8. Characteristics of Included Studies.
PICO 3: What is the diagnostic accuracy of AlereLAM for diagnosis of TB in adults with advanced HIV disease irrespective of signs and symptoms of TB?

There were limited data to evaluate AlereLAM by CD4 threshold for unselected participants irrespective of signs and symptoms of TB.

PICO 3a: Accuracy, in inpatient setting and CD4 ≤ 200

For inpatients with CD4 ≤ 200 cells per µL, only one study contributed data and found a sensitivity (95% CI) of 64% (35% to 87%) and specificity of 82% (67% to 93%) (54 participants; 26% with TB) (Bjerrum 2015). See Figure 11.

PICO 3b: Accuracy, in outpatient settings and CD4 ≤ 200

For outpatients with a CD4≤ 200 cells per µL, two studies contributed data (652 participants; 10% with TB. Sensitivity and specificity were 36% and 94% for Bjerrum 2015, and 7% and 99% for Hanifa 2016. Pooled sensitivity and specificity were 21% (8% to 48%) and 96% (89% to 99%). See Figure 11.

PICO 3c: Accuracy, in all settings and CD4 ≤ 200

Two studies evaluated AlereLAM 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. Pooled sensitivity and specificity were 26% (9% to 56%) and 96% (87% to 98%) (706 participants; 12% with TB). See Figure 11.
Figure 11. Diagnostic accuracy in adults irrespective of signs and symptoms, ≤ 200, by setting
Forest plots and meta-analysis of AlereLAM sensitivity and specificity for TB against a microbiological reference standard, CD4 ≤ 200, by setting.

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Unselected participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Studies (total participants)</td>
</tr>
<tr>
<td>CD4 ≤ 200, All settings</td>
<td>2 studies (706)</td>
</tr>
<tr>
<td>CD4 ≤ 200, Inpatients</td>
<td>1 study (54)</td>
</tr>
<tr>
<td>CD4 ≤ 200, Outpatients</td>
<td>2 studies (652)</td>
</tr>
</tbody>
</table>

Footnote: The proportion of symptomatic participants included was 91% in Bjerrum 2015 and 53% in Hanifa 2016. See Appendix 8. Characteristics of Included Studies.

PICO 3d: Accuracy, in inpatient setting and CD4 ≤ 100

For inpatients with CD4 ≤ 100 cells per μL, two studies contributed data, (200 participants; 42% with TB) (Bjerrum 2015; Lawn 2017). Sensitivity and specificity were 60% and 80% for Bjerrum 2015, and 55% and 98% for Lawn 2017. The pooled sensitivity and specificity were 57% (33% to 79%) and 90% (69% to 97%). See Figure 12.

PICO 3e: Accuracy, in outpatient settings and CD4 ≤ 100

Two studies evaluated outpatients with CD4 ≤ 100 cells per μL, the pooled sensitivity and specificity were 40% (20% to 64%) and 87% (68% to 94%) (217 participants; 21% with TB) (Bjerrum 2015; Drain 2015). See Figure 12.
PICO 3f: Accuracy, in all settings and CD4 ≤ 100

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%. See Figure 12. 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%) for participants with CD4 ≤ 100 cells per µL (417 participants; 31% with TB) (Bjerrum 2015; Drain 2015; Lawn 2017).

Figure 12. Diagnostic accuracy in adults irrespective of signs and symptoms, CD4 ≤ 100, by setting.
Forest plots and meta-analysis of AlereLAM sensitivity and specificity for TB against a microbiological reference standard, CD4 ≤ 100, by setting.

### Unselected adults, CD4 ≤ 100, all settings

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bjerrum 2015</td>
<td>14</td>
<td>18</td>
<td>15</td>
<td>148</td>
<td>0.48 [0.29, 0.67]</td>
<td>0.89 [0.83, 0.93]</td>
</tr>
<tr>
<td>Drain 2015</td>
<td>10</td>
<td>8</td>
<td>17</td>
<td>32</td>
<td>0.37 [0.19, 0.58]</td>
<td>0.80 [0.64, 0.91]</td>
</tr>
<tr>
<td>Lawn 2017</td>
<td>41</td>
<td>2</td>
<td>33</td>
<td>79</td>
<td>0.55 [0.43, 0.67]</td>
<td>0.98 [0.91, 1.00]</td>
</tr>
</tbody>
</table>

### Unselected adults, CD4 ≤ 100, inpatients

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bjerrum 2015</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>28</td>
<td>0.60 [0.26, 0.88]</td>
<td>0.80 [0.63, 0.92]</td>
</tr>
<tr>
<td>Lawn 2017</td>
<td>41</td>
<td>2</td>
<td>33</td>
<td>79</td>
<td>0.55 [0.43, 0.67]</td>
<td>0.98 [0.91, 1.00]</td>
</tr>
</tbody>
</table>

### Unselected adults, CD4 ≤ 100, outpatients

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bjerrum 2015</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>120</td>
<td>0.42 [0.20, 0.67]</td>
<td>0.92 [0.85, 0.96]</td>
</tr>
<tr>
<td>Drain 2015</td>
<td>10</td>
<td>8</td>
<td>17</td>
<td>32</td>
<td>0.37 [0.19, 0.58]</td>
<td>0.80 [0.64, 0.91]</td>
</tr>
</tbody>
</table>

Footnote: The proportion of symptomatic participants included was not stated for Drain 2015 and was 91% in both Bjerrum 2015 and Lawn 2017. See Appendix 8. Characteristics of Included Studies.

### Additional investigations of heterogeneity

**CD4 count**

For comparison to studies evaluating diagnostic accuracy at lower CD4 counts, we assessed diagnostic accuracy among participants with CD4 > 200 cells per µL and CD4 > 100 cells per µL.
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). Four studies evaluated AlereLAM in participants with CD4 > 100 cells per µL where sensitivity estimates ranged from 0% to 33% and specificity estimates ranged from 95% to 99%. Pooled sensitivity and specificity (95% CrI) were 20% (10% to 35%) and 98% (95% to 99%), (952 participants, 12% with TB). See Figure 13.

**Figure 13. Diagnostic accuracy in adults irrespective of signs and symptoms, by CD4 count.**

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Studies (total participants)</th>
<th>Participants with TB (%)</th>
<th>Pooled sensitivity (95% CI)</th>
<th>Pooled specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4 &gt; 200</td>
<td>1 study a (156)</td>
<td>11 (7%)</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>CD4 ≤ 200</td>
<td>2 studies (706)</td>
<td>82 (12%)</td>
<td>26% (9 to 56)</td>
<td>96% (87 to 98)</td>
</tr>
<tr>
<td>CD4 &gt; 100</td>
<td>4 studies (952)</td>
<td>115 (12%)</td>
<td>20% (10 to 35)</td>
<td>98% (95 to 99)</td>
</tr>
<tr>
<td>CD4 ≤ 100</td>
<td>3 studies (417)</td>
<td>130 (31%)</td>
<td>47% (40 to 64)</td>
<td>90% (77 to 96)</td>
</tr>
</tbody>
</table>

Footnote: The proportion of symptomatic participants included ranged from 19% in LaCourse 2016 to 91% in both Bjerrum 2015 and Lawn 2017. See Appendix 8. Characteristics of Included Studies.

*Bjerrum 2015*, Sensitivity 27% (6% to 61%); Specificity 99% (96% to 100%).
**TB prevalence**

The median prevalence of TB in studies with unselected participants was 10% (IQR 9% to 17%). In a secondary analysis by TB prevalence, we found that pooled sensitivity and specificity for unselected participants in settings with TB 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 TB prevalence less than 10% (three studies). In general, TB prevalence increased in studies with a higher proportion of symptomatic participants.

**Sensitivity analyses**

When we reclassified studies with more than 80% of participants being symptomatic at inclusion as ‘studies of symptomatic adults’ ([Bjerrum 2015; Lawn 2017](#)) pooled sensitivity and specificity changed slightly to 31% (16% to 50%) and 95% (84% to 98%) (five studies, 2483 participants).

Limiting analysis to studies with low risk of bias for patient selection pooled sensitivity increased to 39% (17% to 66%) and specificity dropped to 93% (61% to 92%) (three studies, 1338 participants).

Limiting analysis to studies with a low risk of bias in the reference standard domain (two studies), increased pooled specificity to 99% (95% to 99%) while pooled sensitivity decreased to 24% (8% to 53%). As for studies with symptomatic individuals, sensitivity increased in studies evaluating AlereLAM on fresh urine rather than stored urine sample with sensitivity at 41% and specificity at 93% (five studies).

**Additional analyses**

**Investigations of heterogeneity, all studies combined**

As we found a similar pattern among studies with symptomatic and studies with unselected participants in regard to performance when stratified by clinical setting and CD4, we investigated heterogeneity for these variables across all 15 studies combined. We present pooled sensitivity and specificity of AlereLAM for all studies combined stratified by setting and by CD4 cell count in
Table 2.

**Setting**
We identified a total of nine studies evaluating AlereLAM among inpatients and 10 studies among outpatients. The pooled sensitivity (95% CrI) among inpatients was 54% (44% to 67%) (2790 participants) versus 30% (21% to 41%) (3772 participants) among outpatients. Pooled specificity among inpatients was lower at 87% (75% to 94%) versus 95% (86% to 98%) among outpatients.

**CD4 cell count**
In the combined analysis of participants with CD4 ≤ 100 cells per µL, pooled sensitivity (95% CrI) was 52% (41% to 63%) (7 studies, 1656 participants) versus 18% (12% to 25%) (eight studies, 2471 participants) among those with CD4 > 100 cells per µL. The pooled specificity was 89% (82% to 94%) for participants with CD4 ≤ 100 cells per µL and 97% (94% to 98%) for those with CD4 > 100 cells per µL.

In the combined analysis of participants with CD4 ≤ 200 cells per µL, pooled sensitivity (95% CrI) was 39% (25% to 54%) (six studies, 2531 participants) versus 17% (9% to 30%) (four studies, 894 participants) among those with CD4 > 200 cells per µL. The pooled specificity was 92% (85% to 96%) for participants with CD4 ≤ 200 cells per µL and 96% (87% to 98%) for those with CD4 > 200 cells per µL. When stratified by both setting and CD4 for all studies combined the sensitivity remained higher (and specificity lower among inpatients compared to outpatient in all CD4 strata.)
Table 2. AlereLAM pooled sensitivity and specificity for TB diagnosis, combined analysis of studies among symptomatic and unselected participants.

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Studies (total participants)</th>
<th>Participants with TB (%)</th>
<th>Pooled sensitivity (95% CrI)</th>
<th>Pooled specificity (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>By setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient</td>
<td>9 studies (2790)</td>
<td>1027 (37%)</td>
<td>54% (44 to 67)</td>
<td>87% (74 to 94)</td>
</tr>
<tr>
<td>Outpatient</td>
<td>10 studies (4024)</td>
<td>682 (17%)</td>
<td>30% (20 to 41)</td>
<td>96% (92 to 98)</td>
</tr>
<tr>
<td><strong>By CD4 cell</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 &gt; 200</td>
<td>4 studies (894)</td>
<td>174 (19%)</td>
<td>17% (9 to 30)</td>
<td>96% (87 to 98)</td>
</tr>
<tr>
<td>CD4 ≤ 200</td>
<td>6 studies (2531)</td>
<td>804 (32%)</td>
<td>39% (25 to 54)</td>
<td>92% (85 to 96)</td>
</tr>
<tr>
<td>CD4 &gt; 100</td>
<td>8 studies (2471)</td>
<td>540 (22%)</td>
<td>18% (12 to 25)</td>
<td>97% (94 to 98)</td>
</tr>
<tr>
<td>CD4 ≤ 100</td>
<td>7 studies (1656)</td>
<td>642 (39%)</td>
<td>52% (41 to 63)</td>
<td>89% (82 to 94)</td>
</tr>
<tr>
<td><strong>By CD4 and setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 ≤ 200 inpatients</td>
<td>3 studies (1063)</td>
<td>362 (34%)</td>
<td>56% (39 to 72)</td>
<td>81% (66 to 90)</td>
</tr>
<tr>
<td>CD4 ≤ 100 inpatients</td>
<td>4 studies (934)</td>
<td>354 (38%)</td>
<td>60% (46 to 72)</td>
<td>86% (72 to 93)</td>
</tr>
<tr>
<td>CD4 101-200 inpatients</td>
<td>3 studies (284)</td>
<td>82 (29%)</td>
<td>37% (20 to 62)</td>
<td>83% (63 to 93)</td>
</tr>
<tr>
<td>CD4 &gt; 200 inpatients</td>
<td>2 studies (324)</td>
<td>75 (23%)</td>
<td>21% (9 to 42)</td>
<td>89% (72 to 96)</td>
</tr>
<tr>
<td>CD4 &gt; 100 inpatients</td>
<td>4 studies (789)</td>
<td>197 (25%)</td>
<td>23% (4 to 789)</td>
<td>97% (89 to 99)</td>
</tr>
<tr>
<td>CD4 ≤ 200 outpatients</td>
<td>3 studies (901)</td>
<td>165 (18%)</td>
<td>22% (11 to 40)</td>
<td>96% (90 to 98)</td>
</tr>
<tr>
<td>CD4 ≤ 100 outpatients</td>
<td>3 studies (338)</td>
<td>92 (27%)</td>
<td>36% (21 to 55)</td>
<td>89% (77 to 95)</td>
</tr>
<tr>
<td>CD4 101-200 Outpatients</td>
<td>2 studies (222)</td>
<td>60 (27%)</td>
<td>20% (8 to 43)</td>
<td>96% (88 to 99)</td>
</tr>
<tr>
<td>CD4 &gt; 200 outpatients</td>
<td>1 study (147)</td>
<td>0 (0%)</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>CD4 &gt; 100 patients</td>
<td>4 studies (1120)</td>
<td>234 (21%)</td>
<td>19% (11 to 31)</td>
<td>97% (94 to 99)</td>
</tr>
</tbody>
</table>

Abbreviations: CrI: credible interval; AlereLAM: Alere Determine™ TB lipoarabinomannan assay; TB: tuberculosis.
PICO 4: Can the use of AlereLAM in HIV-positive adults reduce mortality associated with advanced HIV disease?

For PICO 4, data were available for the inpatient setting only. We therefore report results for PICO 4b) Inpatient setting; PICO 4e) Inpatient setting with CD4 ≤ 200; and PICO 4h) Inpatient setting with CD4 ≤ 100.

We identified two studies that assessed the impact of AlereLAM on mortality when the test was used for clinical decision-making (Peter 2016; Gupta-Wright 2018a) Both studies were multi-site randomized controlled trials that evaluated the impact of using AlereLAM as a TB diagnostic test to guide treatment initiation in HIV-positive adult inpatients, comparing all-cause mortality at 56 days between the AlereLAM intervention arm and standard-of-care control arm.

Figure 14 shows the risk of bias assessment for the two studies.

Figure 14. Risk of bias summary: review authors’ judgements about each risk of bias item for each included study.
PICO 4b: Impact of AlereLAM on mortality in inpatient settings

Both Peter 2016 and Gupta-Wright 2018a demonstrated that the use of AlereLAM was associated with reduced eight-week mortality, although in Gupta-Wright 2018a, this was only demonstrated in three subgroups (patients with presumed TB, patients with CD4 counts less than 100 cells per μL, and patients with severe anaemia) rather than the overall trial cohort (Table 3 and Table 4).

Peter 2016 found that, in randomly assigned HIV-positive inpatients, AlereLAM in combination with routine diagnostic tests (smear microscopy, Xpert MTB/RIF, and culture) to guide the rapid initiation of TB treatment in HIV-positive adults with at least one TB symptom and illness severity that warranted admission to hospitals in South Africa, Tanzania, Zambia, and Zimbabwe, was associated with a relative risk reduction of 17% (95% CI 4% to 28%) in eight-week mortality compared with routine diagnostic tests alone (no AlereLAM) (Table 3 and Table 4).

Gupta-Wright 2018a randomly assigned HIV-positive inpatients from two hospitals in Malawi and South Africa, to either the standard of care (sputum Xpert MTB/RIF, with the option of sending additional samples for routine TB investigations such as smear microscopy or culture) or intervention (which included urine testing for AlereLAM and Xpert MTB/RIF in addition to sputum Xpert MTB/RIF) irrespective of clinical presentation or TB status. Mortality at 56 days was 21% in the standard-of-care group versus 18% in the intervention group, [adjusted risk reduction (aRD) -2.8% (95% CI -5.8 to 0.3), P = 0.074]. However, in three of the twelve prespecified, but underpowered, subgroups, mortality was lower in the intervention group than in the standard-of-care group for CD4 counts less than 100 cells per μL [aRD -7.1% (95% CI -13.7 to -0.4), P = 0.036]; severe anaemia [-9.0% (95% CI -16.6 to -1.3), P = 0.021]; and patients with clinically suspected TB [aRD -5.7% (95% CI -10.9 to -0.5), P = 0.033] (Table 3 and Table 4).

In the meta-analysis involving both trials, the pooled risk ratio was 0.85 (95% CI 0.76 to 0.94) i.e. study participants undergoing AlereLAM testing had 0.85 times the risk or 15% lower risk of mortality than participants undergoing routine TB diagnostic testing without AlereLAM (Figure 15). The absolute effect was 35 fewer deaths per 1,000 (from 14 fewer to 55 fewer) (high-certainty evidence).
Figure 15. Impact of AlereLAM on mortality in HIV-positive adult inpatients.
Forest plots and meta-analysis of the impact of AlereLAM on mortality, compared to the control study arms that did not include AlereLAM testing.

PICO 4e: Impact of AlereLAM on mortality in inpatients with CD4 ≤ 200

Peter 2016, reported that in their trial of HIV-positive adult inpatients with at least one TB symptom with CD4 count of ≤ 200 cells per µL (1725 patients), the use of AlereLAM testing (intervention) was associated with a HR of 0.87 (0.72 to 1.04) for mortality (i.e. 13% reduction in mortality) compared to the study arm without AlereLAM testing (Table 3 and Table 4). Gupta-Wright 2018a found that in their trial of unselected HIV-positive adult inpatients, rapid urine-based screening (which included AlereLAM) was associated with an adjusted risk difference of -0.1 (-3.3 to -3.1, P = 0.96) in patients with a CD4 count ≥ 100 cells per µL (Table 3 and Table 4).

PICO 4h: Impact of AlereLAM on mortality in inpatients with CD4 ≤ 100

Peter 2016, reported that in their trial of HIV positive adult inpatients with at least one TB symptom with a CD4 count ≤ 100 cells per µL (1272 patients), the use of AlereLAM testing (intervention) was associated with a HR of 0.88 (0.72 to 1.08) for mortality (i.e. 12% reduction in mortality) compared to the study arm without AlereLAM testing (Tables 3 and 4). The greatest reduction in mortality (29%) occurred in the 867 patients with a CD4 count ≤ 50 cells per µL (HR 0.71, 0.56 to 0.90). Gupta-Wright 2018a found that in their trial of unselected HIV-positive adult inpatients, rapid urine-based screening (which included AlereLAM) was associated with an adjusted risk difference of -7.1 (-13.7 to -0.4; P = 0.036) in patients with a CD4 count < 100 cells per µL [adjusted odds ratio of 0.72 (0.53 to 0.98)] (Table 3 and Table 4).
Table 3. Comparison of mortality in randomized trials that evaluated a diagnostic or screening intervention using AlereLAM in HIV-positive participants

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Illness severity metrics</th>
<th>Design</th>
<th>Time of mortality assessment</th>
<th>Mortality analysis</th>
<th>Mortality in intervention</th>
<th>Mortality in control</th>
<th>Other outcomes assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gupta-Wright 2018a</td>
<td>2574 HIV-positive adults, inpatients (unselected)</td>
<td>Median CD4 222 cells per μL, Karnofsky score 60, BMI 21.7, median haemoglobin 10.4 g/dL</td>
<td>Randomized controlled trial</td>
<td>56 days</td>
<td>aRD -2.8%, 95% CI: 5.8 to 0.3; P = 0.074</td>
<td>18% (235/1287)</td>
<td>21% (272/1287)</td>
<td>Significant mortality reduction in three subgroups - severe anaemia (aRD -9.0%, 95% CI -16.6 to -1.3; P = 0.021); CD4 &lt; 100 cells per μL (aRD -7.1%, 95% CI -13.7 to -0.4; P = 0.036); clinically suspected TB (-5.7% 95% CI -10.9 to -0.5; P = 0.033). More patients in LAM arm were started on treatment (aHR 1.56, 95% CI 1.29 to 1.88; P &lt; 0.0001).</td>
</tr>
<tr>
<td>Peter 2016</td>
<td>2528 HIV-positive adults, inpatients (symptoms)</td>
<td>Median CD4 84 cells per μL, Karnofsky score 50, BMI 18.8, median haemoglobin 9.2 g/dL</td>
<td>Randomized controlled trial</td>
<td>8 weeks</td>
<td>ARR 4% (1% to 7%) aRR 0.83, 95% CI 0.73 to 0.96, P = 0.012</td>
<td>20.8% (261/1257)</td>
<td>24.9% (317/1271)</td>
<td>Greatest mortality reduction in those with CD4 &lt; 50 cells per μL (HR 0.71, 0.56 to 0.90). More patients in LAM arm were started on treatment (52% versus 47%; P = 0.024).</td>
</tr>
</tbody>
</table>

Abbreviations: aRD: adjusted risk difference; ARR: absolute risk reduction; aRR: adjusted risk ratio.
Table 4. Effect of using AlereLAM on mortality, stratified by CD4 group

<table>
<thead>
<tr>
<th>Study</th>
<th>CD4 group</th>
<th>Effect (95% CI)</th>
<th>CD4 group</th>
<th>Effect (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gupta-Wright 2018a</td>
<td>≥ 100</td>
<td>aOR 0.96 (0.74 to1.25)</td>
<td>&lt; 100</td>
<td>aOR 0.72 (0.53 to 0.98)</td>
</tr>
<tr>
<td>Peter 2016</td>
<td>&gt; 100</td>
<td>HR 0.71 (0.53 to 0.96)</td>
<td>≤ 100</td>
<td>HR 0.88 (0.72 to 1.08)</td>
</tr>
<tr>
<td>Peter 2016</td>
<td>&gt; 200</td>
<td>HR 0.65 (0.44 to 0.97)</td>
<td>≤ 200</td>
<td>HR 0.87 (0.72 to 1.04)</td>
</tr>
</tbody>
</table>

aOR: adjusted odds ratio; CI: confidence interval; HR: hazard ratio

Important differences between the trials evaluating the impact of AlereLAM on mortality

There were several differences between the two trials. The median CD4 count was lower in Peter 2016 compared to Gupta-Wright 2018a (84 cells per µL versus 227 cells per µL). This, in addition to a lower BMI and Karnofsky score, suggests that the population evaluated in the Peter 2016 study may have been sicker. Overall severity of illness was higher in Peter 2016 (mortality 21% in AlereLAM and in 25% in no AlereLAM arms) compared to Gupta-Wright 2018a (mortality 18% in AlereLAM and 21% in no AlereLAM arms). The percentage of patients on antiretroviral therapy was lower in Peter 2016 than in Gupta-Wright 2018a (48% versus 72%). A greater proportion of participants were started on TB treatment (52% versus 21% in Peter 2016 compared to Gupta-Wright 2018a, reflecting different exclusion criteria (clinical suspicion of TB compared with an unselected population irrespective of symptoms).

Impact of AlereLAM on other patient-important outcomes

Although mortality was the primary patient-important outcome of interest for our data extraction, we also recorded data on other patient-important outcomes. Peter 2016 found that the overall percentage of patients started on TB treatment was higher in the intervention group that received AlereLAM (52% versus 47%; P = 0.024), including a higher proportion that was started on days 0-3 (79% versus 69%; P < 0.0001). Gupta-Wright 2018a found that time to diagnosis was marginally shorter in the AlereLAM intervention group compared to the standard-of-care group [median 0 days (IQR 0 to 1) versus 1 day (0 to 6)]. They reported that increases in TB diagnoses in the intervention group that received AlereLAM were not confined to high-risk subgroups, unlike mortality, with an adjusted absolute risk increase of 7.0% (95% CI 4.1 to 10.0) in TB diagnoses in patients with CD4 counts of 100 cells per µL. Time from diagnosis to treatment was short (median of 1 day, IQR 0 to 1) and did not differ between the group that received AlereLAM and the standard-of-care. However, more patients were started on TB treatment during admission in the group that received AlereLAM (268/1287) compared to the standard-of-care group (182/1287) (aHR 1.56, 95% CI 1.29 to 1.88; P < 0.0001).

Association between AlereLAM positivity with mortality

Association between AlereLAM test positivity with mortality in all settings

We additionally identified 12 studies that had data on the association between AlereLAM positivity and mortality (Table 5) as part of diagnostic accuracy studies (in which AlereLAM was not used for...
clinical decision making). Three studies evaluated the diagnostic accuracy of AlereLAM (without using results for clinical decision making) in inpatients (LaCourse 2018a; Lawn 2017; Manabe 2014) and six studies evaluated the diagnostic accuracy of AlereLAM (without using results for clinical decision making) in outpatients (Balcha 2014; Drain 2015a; Drain 2017; Hanifa 2016; Lawn 2012b; Peter 2015) and three studies evaluated its use in both inpatients and outpatients (Bjerrum 2015; Huerga 2017; Thit 2017). All studies included only HIV-positive participants except one (Drain 2015a), in which the study population consisted of adults with ≥ two TB symptoms for ≥ two weeks being initiated on TB therapy, of whom 93% were HIV-positive. All studies evaluated adults aside from one (LaCourse 2018a) that evaluated children. All were prospective cohorts or nested prospective cohorts within trials or cross-sectional diagnostic accuracy studies (Table 5). The timing of mortality assessment was highly variable and ranged from 56 days to 12 months. The type of mortality analysis also varied although the majority of prospective cohort studies used hazard ratios.

When considering the association of AlereLAM and mortality (not used for clinical decision making), all prospective studies compared patients who had a positive AlereLAM test with those who had a negative AlereLAM test, with some studies providing additional data on the AlereLAM test status stratified by those with a confirmed diagnosis of TB, those who did not have TB and those with an inconclusive evaluation for TB. Data on patient outcomes were largely restricted to post-hoc analyses. However, all prospective cohort studies aside from one (Thit 2017) demonstrated a significant association between AlereLAM test positivity and mortality, despite considerable variability in the method of TB diagnosis, provision of treatment and length of follow-up. We note that these investigators did not use the results of AlereLAM to guide treatment initiation.

**Association between AlereLAM test positivity with mortality in inpatient settings**

LaCourse 2018a reported higher mortality at six months (134/100 person years versus 32/100 person years [AHR 4.61 (95% CI 1.63 to 12.96), P = 0.004] in AlereLAM positive than AlereLAM negative hospitalized HIV-positive children who were evaluated for TB irrespective of clinical suspicion. Lawn 2017 reported higher mortality at 90 days [24.5% versus 7.2%, aOR 4.2 (95% CI 1.50 to 11.75)] in unselected HIV-positive adult inpatient. Of note, AlereLAM was performed on frozen urine specimens that were obtained at the time of enrolment. Manabe 2014 reported higher mortality at six months (40% versus 28%, P = 0.016) in AlereLAM positive than AlereLAM negative hospitalized HIV-positive adults with at least one TB symptom (secondary analysis, Nakifyingr 2014). Manabe 2014 additionally reported higher mortality in AlereLAM positive study participants with confirmed TB (39% versus 20%), those with possible TB (22% versus 17%) and those without evidence of TB (49% versus 31%), compared to those in each of these groups respectively that were AlereLAM negative. In contrast, Thit 2017, which was a hospital-based study in Myanmar with inpatients and outpatients that represents the only study with impact data that was conducted outside Africa, reported that AlereLAM had limited potential clinical utility in their cohort. Four out of the six inpatients who died had a positive AlereLAM test but three received anti-TB therapy prior to death and the fourth had cryptococcal meningitis. Bjerrum 2015 did not report mortality data for inpatients and outpatients separately but found that AlereLAM positive participants had a significantly higher probability of death compared to AlereLAM negative in the overall population (49% versus 14%, P < 0.001) and
among those with confirmed TB (54% versus 16%, P = 0.002). Bjerrum 2015 reported that among TB participants who received TB treatment, 31% of those who were AlereLAM positive died compared to only 4% of those who were AlereLAM negative. Among TB participants who did not receive treatment at the time of assessment in the study, 100% of those who were AlereLAM positive died compared to 33% of those who were AlereLAM negative. Huerga 2017 also did not report mortality data for inpatients and outpatients separately but found that mortality was higher in AlereLAM positive compared to AlereLAM negative patients (22.8% versus 8.1%, P < 0.0001), although this difference was not statistically significant amongst confirmed TB patients (confirmed TB patients: 22.8% vs 11.1%, P = 0.130). In a post-hoc analysis, Peter 2013 reported that among inpatients, AlereLAM positive TB participants missed by empirical early treatment had lower CD4 counts and higher median illness severity scores, compared to participants who received early treatment based on clinical decision making.

**Association between AlereLAM test positivity with mortality in outpatient settings**

We identified six studies that presented results on the association of LAM positivity and mortality in the outpatient setting. Balcha 2014 reported higher mortality (20.0% versus 2.7%, P < 0.001) in AlereLAM positive than AlereLAM negative participants. Drain 2015a reported AlereLAM responses over time. They reported that among participants receiving TB therapy, having a positive AlereLAM test at the two-month visit was associated with an adjusted hazard ratio (HR) of 5.58 for mortality (median follow up time of 49 months) compared to participants with a negative AlereLAM test at the two-month visit. Participants with a positive AlereLAM at six months had an adjusted HR of 42.1 for mortality during study follow-up. They found no difference (adjusted HR 0.99, P = 0.99) in mortality comparing baseline AlereLAM results. Drain 2017 reported that HIV-positive ART-naïve adult outpatients with a positive AlereLAM test was associated with an adjusted hazard ratio (HR) of 4.26 for mortality (follow up time of 12 months). Hanifa 2016 reported a higher mortality [14% versus 5%, HR 3.6 (95% CI 1.2 to 10.5), P = 0.04] in AlereLAM positive compared to AlereLAM negative (using grade 1 as the test positivity threshold) HIV-positive adults attending HIV clinics (CD4 ≤ 200 cells per µL) irrespective of symptoms or presentation. Lawn 2012b found that among 23 TB participants who were AlereLAM positive, five people died (22%) compared to zero deaths (0/36) among TB participants who were AlereLAM negative (secondary analysis, Lawn 2012a). Peter 2015 reported mortality of 25% and 11% in AlereLAM positive and AlereLAM negative participants, respectively. In another secondary analysis to Lawn 2012a, Lawn 2013 reported that AlereLAM sensitivity was 100% among TB participants who died compared to 25% among TB participants who were alive at 90 days (P = 0.002).

**General observations on AlereLAM positivity, mortality, and CD4 count**

LaCourse 2018a and Drain 2015a adjusted their mortality analysis for baseline CD4 count or percentages but did not report specific mortality data stratified by CD4 count. Lawn 2012a and Lawn 2017 found that those testing AlereLAM positive had lower CD4 counts and a higher prevalence and severity of anaemia (P < 0.001) but mortality analyses evaluated AlereLAM positivity and CD4 count separately.
Association between AlereLAM test positivity with mortality in individuals with CD4 ≤ 200 cells per µL

*Drain 2017* reported that in HIV positive ART-naïve adult outpatients with CD4 count of < 200 cells per µL, a positive AlereLAM test (grade 2 on the old reference scale card with five band intensities) was associated with an adjusted hazard ratio (HR) of 2.71 (0.95 to 7.71, $P = 0.06$) for mortality compared to those who had no evidence of TB and a negative AlereLAM result, which rose to 3.61 (1.69 to 7.71, $P = 0.0009$) when a positive AlereLAM result of grade 3+ or above was analysed (old reference card).

Association between AlereLAM test positivity with mortality in individuals with CD4 ≤ 100 cells per µL

*Thit 2017* reported that of the five deaths among 21 inpatients with TB symptoms and a CD4 T-cell count < 100 cells per µL, three (60%) had a positive AlereLAM result but all of these were on TB treatment prior to death. *Balcha 2014* reported that among 21 outpatients with positive AlereLAM results who had not received TB diagnosis (neither by bacteriological nor clinical criteria), five died within six months of inclusion. All five whom had positive WHO symptom screens and had baseline CD4 counts < 100 cells per µL; three had started ART within three months of inclusion but none had started TB treatment. *Drain 2015a* reported that the overall mean urine AlereLAM test grade decreased from 0.7 (+/- 1.3) at baseline to 0.5 (+/- 1.3) at two months to 0.2 (+/- 0.7) at the six months visit and that these results were similar when stratified by CD4 above/below 100 cells per µL. *Drain 2017* reported that in HIV-positive ART-naïve adult outpatients with CD4 count ≤ 100 cells per µL, a positive AlereLAM test (grade 2 on the old reference scale card with five band intensities) was associated with an adjusted hazard ratio (HR) of 2.96 (1.01 to 8.70, $P = 0.05$) for mortality compared to those who had no evidence of TB and a negative AlereLAM result, which rose to 3.04 (1.34 to 6.91, $P = 0.008$) when a positive AlereLAM result of ≥ 3+ was analysed.

Although these studies did not directly assess the impact of AlereLAM on patient-important outcomes, *Lawn 2012b* found that patients who had a positive AlereLAM result initiated treatment within eight days, compared to 21 days for those with a negative test result. *Peter 2015* demonstrated that LAM positivity was associated with same day treatment initiation, compared to treatment initiation between day 2 to 56 for those with a negative test result.
<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Design and timing for mortality analysis</th>
<th>Population for mortality analysis</th>
<th>Mortality in LAM positive</th>
<th>Mortality in LAM negative</th>
<th>Other outcomes assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaCourse 2018a</td>
<td>181 HIV-positive children, inpatients (unselected)</td>
<td>Nested prospective cohort 6 months</td>
<td>137 HIV-positive inpatient children with valid LAM results</td>
<td>134/100 person years</td>
<td>32/100 person years</td>
<td>Hazard ratio adjusted for CD4 %</td>
</tr>
<tr>
<td>Lawn 2017</td>
<td>427 HIV-positive adults, inpatients (unselected)</td>
<td>Prospective cohort 90 days</td>
<td>136 TB cases</td>
<td>24.5% (13/53) aOR 4.2 95% CI: 1.50-11.75</td>
<td>7.2% (6/83)</td>
<td>LAM+ participants had a lower CD4 count and more severe anaemia (P &lt; 0.001)</td>
</tr>
<tr>
<td>Manabe 2014</td>
<td>506 HIV-positive adults with at least 1 TB symptom, inpatients (symptomatic)</td>
<td>Prospective cohort 6 months</td>
<td>351 enrollees</td>
<td>40% (54/134) 39% (35/90) 22% (2/9) 49% (17/35) Unadjusted HR for LAM positivity 1.67; P = 0.025</td>
<td>28% (60/217) 20% (11/55) 17% (2/12) 31% (47/150)</td>
<td></td>
</tr>
<tr>
<td>Thit 2017</td>
<td>517 HIV-positive adults, inpatients (unselected)</td>
<td>Prospective cohort 6 months</td>
<td>54 TB cases</td>
<td>11.4% (4/35)</td>
<td>10.5% (2/19)</td>
<td></td>
</tr>
<tr>
<td>Balcha 2014*</td>
<td>757 HIV-positive adults eligible for ART (CD4 &lt; 350 or WHO stage 4), outpatients (unselected)</td>
<td>Prospective cohort 6 months</td>
<td>148 TB cases</td>
<td>20% (7/35)</td>
<td>2.7% (3/113)</td>
<td></td>
</tr>
<tr>
<td>Bjerrum 2015</td>
<td>469 HIV-positive adults eligible for ART (WHO stage 3/4, CD4 &lt; 350) (unselected)</td>
<td>Prospective cross-sectional 6 months</td>
<td>469 enrollees</td>
<td>49% 22/45 54% (13/24) 32% (5/16) Kaplan-Meier log-rank test P &lt; 0.001</td>
<td>14% (59/424) 16% (5/31) 4% (1/23)</td>
<td></td>
</tr>
<tr>
<td>Drain, 2015a</td>
<td>90 adults with ≥ 2 TB symptoms for ≥ 2 weeks being initiated on TB</td>
<td>Prospective cohort 2 months and 6 months</td>
<td>90 outpatients</td>
<td>50% (4/8), AHR 5.58 95% CI: 1.24-25.2, P = 0.02 AHR 42.1 95% CI: 1.87-9.52, P = 0.02</td>
<td>18.5% (12/65)</td>
<td>Treatment monitoring: the % of LAM+ participants decreased from 32% (baseline) to 16% (2</td>
</tr>
<tr>
<td>Study</td>
<td>Population Description</td>
<td>Study Design</td>
<td>Follow-up Duration</td>
<td>Therapy, no sputum or smear negative, outpatients (symptomatic)</td>
<td>6 months</td>
<td>31.2% (29/93) MHR 4.26 95% CI: 2.65-6.84 (includes LAM grade 1 positive result)</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------</td>
<td>----------</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>Drain 2017</td>
<td>796 HIV-positive adults, ART-naïve, outpatients (unselected)</td>
<td>Prospective cohort</td>
<td>12 months</td>
<td>726 HIV-positive ART-naïve outpatients</td>
<td>6 months</td>
<td>14% (4/28) HR 3.6 95% CI: 1.2-10.5 P = 0.04</td>
</tr>
<tr>
<td>Hanifa 2016*</td>
<td>586 HIV-positive adults (CD4 &lt; 200) attending HIV clinics, outpatients (unselected)</td>
<td>Nested within</td>
<td>6 months</td>
<td>426 enrollees with evaluable data</td>
<td>4% (4/28) HR 3.6 95% CI: 1.2-10.5 P = 0.04</td>
<td>8.1% 11.1% 11.1% 22.8% 22.8% 15.8% 15.8% 28.1% 8.1% 11.1% 9.9%</td>
</tr>
<tr>
<td>Huerga 2017*</td>
<td>474 HIV-positive adults with cough or cough plus other TB symptom, inpatients and outpatients (symptomatic)</td>
<td>Prospective cohort</td>
<td>2 months</td>
<td>468 enrollees with vital status data</td>
<td>21.7% (5/23) (same at 30 days)</td>
<td>0% (0/36) (same at 30 days)</td>
</tr>
<tr>
<td>Lawn 2012b*</td>
<td>325 HIV-positive adults, ART-naïve, outpatients (unselected)</td>
<td>Prospective cohort</td>
<td>90 days</td>
<td>59 TB cases</td>
<td>25% (9/32) 35% (6/17) ARR 14% P = 0.02</td>
<td>11% (40/361) 14% (15/106)</td>
</tr>
<tr>
<td>Peter 2015</td>
<td>583 HIV-positive adults with suspected TB (symptomatic)</td>
<td>Cross-sectional</td>
<td>6 months</td>
<td>583 enrollees</td>
<td>25% (9/32) 35% (6/17) ARR 14% P = 0.02</td>
<td>11% (40/361) 14% (15/106)</td>
</tr>
</tbody>
</table>

Abbreviations: AHR: adjusted hazard ratio; aOR: adjusted odds ratio; ARR: absolute risk reduction; ART: antiretroviral therapy; MHR: mortality hazard ratio; TB: tuberculosis.

*Denotes study in which grade 1 (using the old reference card with five band intensities) was used.
Discussion

This updated systematic review summarizes the current literature and includes 15 unique studies on the accuracy of the urine-based lateral flow lipoarabinomannan assay, Alere Determine™ TB LAM Ag, 'AlereLAM', for tuberculosis (TB) in adults with human immunodeficiency virus (HIV) and integrates nine new studies identified since the original WHO and Cochrane Review (WHO Lipoarabinomannan Policy Guidance 2015; Shah 2016). Eight studies used AlereLAM for TB diagnosis in symptomatic adult participants with signs and symptoms suggestive of TB. These studies largely focused on inpatient settings and had high TB prevalence. Seven studies used AlereLAM for diagnosing TB in unselected participants that may or may not have had symptoms suggestive of TB 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 TB 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 TB/HIV burden, and only one study outside sub-Saharan Africa. We identified two randomized controlled trials that evaluated the impact of AlereLAM implementation among hospitalized patients on morality and other outcomes.


Summary of main results

For TB diagnosis in HIV-positive adults presenting with signs and symptoms of TB, the diagnostic accuracy of AlereLAM is:
- in inpatient settings, sensitivity 52% and specificity 87% (PICO1a)
- in outpatient settings, sensitivity 29% and specificity 96% (PICO 1b)
- in all settings, sensitivity 42% and specificity 91% (PICO1c)

For TB diagnosis in HIV-positive adults irrespective of signs and symptoms of TB, the diagnostic accuracy of AlereLAM is:
- in inpatient settings, sensitivity 62% and specificity 84% (PICO 2a)
- in outpatient settings, sensitivity 31% and specificity 95% (PICO 2b)
- in all settings, sensitivity 35% and specificity 95% (PICO 2c)

For diagnosis of TB in adults with advanced HIV disease irrespective of signs and symptoms of TB, the diagnostic accuracy of AlereLAM is (limited data available):
- in inpatient setting CD4 ≤ 200, sensitivity of 64% and specificity 82% (one study, PICO 3a)
- in outpatient setting CD4 ≤ 200, sensitivity 21% and specificity 96% (PICO 3b)
- in all settings CD4 ≤ 200, sensitivity 26% and specificity 96% (PICO 3c)
- in inpatient setting CD4 ≤ 100, sensitivity 57% and specificity 90% (PICO 3d)
- in outpatient setting CD4 ≤ 100, sensitivity 40% and specificity 87% (PICO 3e)
- in all settings CD4 ≤ 100, sensitivity 47% and specificity 90% (PICO 3f)

For diagnosis of TB in HIV-positive children, the diagnostic accuracy of AlereLAM is (limited data available)
- in all settings, including all children, for individual studies, sensitivity and specificity were 42% and 94% (outpatient setting); 56% and 95% (inpatient setting); and 43% and 80% (both inpatient and outpatient settings)

For use of AlereLAM to reduce mortality associated with advanced HIV disease (two randomized trials)
- the pooled risk ratio for mortality was 0.85 (0.76 to 0.94) and the absolute effect was 35 fewer deaths per 1,000 (from 14 fewer to 55 fewer) (PICO 4)

AlereLAM for TB diagnosis in symptomatic participants

For TB diagnosis among symptomatic adults, the pooled sensitivity of AlereLAM was 42% and pooled specificity was 91%. In planned investigations of heterogeneity, we found an inverse correlation between AlereLAM sensitivity and CD4 count, with increasing sensitivity as patient CD4 count decreased (increased from 16% in patients with CD4 cell count > 200 cells per µL to 24% in patients with CD4 cell count between 101-199 cells per µL, to 54% 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 (sensitivity increased from 29% among outpatients to 52% among inpatients) while specificity decreased (from 96% among outpatients to 87% among inpatients).

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

AlereLAM for TB diagnosis in unselected participants

For TB diagnosis among unselected HIV-positive adults (with or without signs or symptoms of TB), the pooled sensitivity was low (35%), with a relatively high pooled specificity (95%). In the investigations of heterogeneity, we expected and found a higher sensitivity in patients with low CD4 cell count and among inpatients compared to patients with higher CD4 cell count and outpatients 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 TB (a positive WHO TB screen), and in the studies evaluating inpatients the majority of participants (> 80%) were in fact presenting with signs and symptoms suggestive of TB. These studies may be considered more similar to studies with exclusively symptomatic participants. In
additional analysis of heterogeneity, we examined diagnostic accuracy based on TB 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 TB prevalence within the study population was ≥ 10%, compared to only 16% when TB 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 TB, 80 would be AlereLAM-positive: of these, 45 (56%) would not have TB (false-positives); and 920 would be AlereLAM-negative: of these, 65 (7%) would have TB (false-negatives).

**AlereLAM for TB diagnosis, overall**

The findings of this updated review are consistent with those of the original review (WHO Lipoarabinomannan Policy Guidance 2015; Shah 2016). Inclusion of additional studies in this updated review provided the basis for a more precise estimate of the AlereLAM 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 TB among people living with HIV than the internationally suggested target of minimum 65% overall for rapid non-sputum TB 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.

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 high yield of urine LAM test positivity (Sabur 2017). However, only a few studies included patients who could not provide sputum samples for diagnostic testing. To the extent that inability to produce sputum is correlated with severity of TB 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 AlereLAM 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 specificities were approaching 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.

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 TB. We included studies that evaluated AlereLAM for diagnosis of pulmonary TB, extrapulmonary TB, or both pulmonary and extrapulmonary TB. However, we recognize that a substantial number of TB 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 TB with extrapulmonary and disseminated forms of disease and considered a standardized reference standard using two or more specimen types to be of higher quality than a reference standard using one specimen type. The higher quality reference standard is better at classifying which patients have and do not have TB. A lower quality reference standard may miss some TB cases and classify some TB patients as not having TB. This may make a truly positive AlereLAM result seem like an FP leading to an underestimation of specificity. In this review, we did not assess performance against a composite reference standard that uses microbiological or clinical information to classify TB. This was done in the original WHO and Cochrane Review (WHO Lipoarabinomannan Policy Guidance 2015; 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 (Qvist 2014 and Nel 2017) have postulated that infection with (disseminated) non-tuberculous mycobacteria 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.

We decided a priori to evaluate performance of AlereLAM in HIV-positive individuals with signs and symptoms of TB (symptomatic) separately from HIV-positive individuals irrespective of signs and symptoms of TB (unselected participants). We considered evaluating AlereLAM performance among specifically asymptomatic (i.e. exclusively those without symptoms) participants to assess the role of AlereLAM for TB screening, 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 TB, suggesting relative similarities to studies that enrolled exclusively symptomatic participants. Consequently, the overall performance of AlereLAM 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 study (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. In a secondary analysis combining studies among symptomatic participants and unselected participants, we found a pooled sensitivity of 54% for inpatients compared to 30% among outpatients and a pooled sensitivity of 52% versus 18% among participants with CD4 ≤ 100 cells per µL and CD4 > 100 cells per µL, respectively. In the analysis of all studies combined, the sensitivity remained higher for inpatients than for outpatients across all CD4 strata. This indicate that other characteristics than lower CD4 may explain the higher sensitivity among inpatients like higher TB prevalence, higher mycobacterial burden, renal or genitourinary tract TB with LAM secretion in urine.

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

**Impact on mortality**

We sought to systematically analyse data on patient-important outcomes. Since the publication of the original Cochrane Review (Shah 2016), a second randomized trial that evaluated the impact of AlereLAM implementation on mortality in unselected HIV-positive inpatients (i.e. as a screening test rather than diagnostic test used in patients with TB symptoms) has been published (Gupta-Wright 2018).

Both trials demonstrated mortality reduction in patients with a CD4 count < 100 cells per µL. Both trials also demonstrated an increase in the number of patients started on treatment. Importantly, Gupta-Wright 2018 demonstrated that only 57% of patients could produce sputum for Xpert
MTB/RIF testing, in contrast to 99% of patients who could produce urine for AlereLAM testing. Of note, in both trials, patients who could not give informed consent were ineligible to participate. This accounted for 1074/9728 (12.3%) in Peter 2016 and 654/4788 (13.7%) in Gupta-Wright 2018. Since some of these patients could not consent due to the severity of their illness, this may have biased the effect of the intervention towards the null. Since both trials were conducted in sub-Saharan Africa, it is possible that this may limit applicability to other populations.

In additional analyses we demonstrated that within diagnostic accuracy studies that included follow-up for clinical outcomes, without using AlereLAM results for clinical decision making, there appeared to be an association between AlereLAM positivity among both participants with and without confirmed TB (by microbiological and/or clinical study reference standards) and mortality. These data must be interpreted cautiously as they represent secondary analyses within observational cohorts, are limited in size, and may not control for important biases or other factors. It is likely that these findings may represent the effect of missed diagnoses (that could be averted through earlier diagnosis using rapid AlereLAM testing) and/or that there is a biological association between disease severity resulting in AlereLAM excretion in urine.

**Differences from original Cochrane Review**

In comparison to the original Cochrane Review (Shah 2016), this updated review includes 15 published studies (eight among studies with symptomatic participants, seven among studies with unselected participants). By contrast, the prior review included data from twelve studies, of which three used an older threshold for determining test positivity that is no longer recommended and three were abstracts of which one was included in this review as an updated published manuscript (Lawn 2014a).

In the evaluation of diagnostic accuracy among symptomatic participants, the pooled estimates for sensitivity (42% versus 45%) and specificity (91% versus 92%) remained similar, comparing the current review and prior review, respectively. When stratified by setting, the pooled estimates among inpatients for sensitivity (52% versus 53%) and specificity (87% versus 90%) 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 the prior review (Shah 2016), some studies were classified as ‘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 was 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 TB) 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 relative lack of data on evaluating the incremental yield of AlereLAM 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 AlereLAM in people living with HIV with signs and symptoms of TB and irrespective of signs and symptoms of TB. 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 TB prevalence. For some analyses and subgroup 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 due to exclusion of patients unable to produce sputum. Moreover, only a single study was conducted outside sub-Saharan Africa.

We had overall low concern about the applicability of the included studies to our review question as assessed by QUADAS-2. However, studies of HIV-positive adults irrespective of signs and symptoms had low representation of asymptomatic individuals, as a majority of participants in fact presented with signs and symptoms of TB. To date, no study evaluated AlereLAM in an asymptomatic population.

Using the GRADE approach, we judged the evidence for diagnostic accuracy of AlereLAM to be of low or very low certainty. This means that our confidence in the effect estimate is limited and the true effect may be substantially different from the estimate of the effect.

**Authors' conclusions**

**Implications for practice**
For people living with HIV, this review found overall lower sensitivity of AlereLAM than the internationally suggested target of minimum 65% overall for non-sputum based TB diagnostic tests (WHO TTP 2014). This was consistent whether the test is used for diagnosis of TB among symptomatic (sensitivity of 42%) or unselected participants (sensitivity of 35%). The estimated sensitivity suggests that if AlereLAM were to be used alone, more than half of all TB cases would be missed.

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

The proposed role for the AlereLAM test is to be used in combination with existing TB tests to assist TB diagnosis and possibly improve important outcomes among HIV-positive patients with advanced disease. The test does not require sputum collection and is not site-specific. Other favorable 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, AlereLAM testing may be the only possible way to confirm a diagnosis when a sputum sample cannot be produced.

Findings suggest that sensitivity increases with lower CD4 counts and in inpatient settings compared to outpatient settings. We found differences in AlereLAM performance based on TB prevalence of the target population (when stratified at greater or less than 10% among unselected participants irrespective of symptoms), with higher diagnostic sensitivity when the study population had higher TB prevalence (≥ 10%). The overall association of differing diagnostic accuracy of AlereLAM when examined by study setting and degree of immunosuppression were consistent irrespective of approach to patient selection. However, we had limited data and these findings should be interpreted with caution.

Clinicians must consider the need for additional testing when interpreting negative AlereLAM results. The consequences of false-negative results are increased risk of morbidity and mortality, delayed treatment initiation, and the continued risk of TB transmission. The consequences of false- positive results are delayed alternative diagnosis, likelihood of anxiety and morbidity caused by additional testing, unnecessary treatment, and possible adverse events; possible stigma associated with a diagnosis of TB. As AlereLAM 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 TB may be inappropriately treated with a regimen appropriate only for drug-sensitive disease.

Implications for research
Future studies that evaluate the diagnostic accuracy of non-sputum-based tests for TB, such as AlereLAM, in people living with HIV should use a reference standard that includes at least two specimen types or extrapulmonary specimens in addition to sputum. Moreover, future studies should include patients unable to expectorate sputum in the analysis. While some studies enrolled unselected participants, our review suggests that a large proportion were symptomatic, particularly in the inpatient setting. These features of study design may decrease the risk of bias in the accuracy estimates. Performance of AlereLAM for TB detection among a cohort of exclusively asymptomatic participants is largely unknown. The indication of increased sensitivity with use of fresh urine needs further investigations, and studies in settings outside sub-Saharan Africa are lacking. Further research on effective implementation of AlereLAM within routine clinical practice is needed because the test can only influence clinical practice if the results are believed and acted upon.
Acknowledgements

We are grateful to Vittoria Lutje, the CIDG Information Specialist, for help with the search strategy. The CIDG editorial base is funded by UK aid from the UK government for the benefit of low- and middle-income countries (project number 300342-104). The views expressed do not necessarily reflect the UK government’s official policies.

We thank all authors of the included studies for answering our questions and providing data.

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. ME and KRS drafted the summary in children. RRN reviewed articles with impact data for inclusion, extracted and analysed data, and drafted the sections on impact. MK, SB, MS, and KRS drafted the GRADE tables. ND and IS drafted the statistical analysis section and the statistical appendix. 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 TB 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. 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.
References

Included studies

Bjerrum 2015

Drain 2015

Drain 2016

Floridia 2017

Hanifa 2016

Huerga 2017

Juma 2017

Lawn 2017

Nakiyingi 2014

Pandie 2016

Peter 2012

Peter 2015

Peter 2016

Thit 2017
Additional studies for impact and association with mortality

Drain 2015a

Drain 2017

LaCourse 2018a

Lawn 2012b

Lawn 2013

Manabe 2014

Peter 2013

Additional references

Alere 2017
Andrews 2014

Balcha 2014

Balshem 2011

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

Briken 2004

Broger 2018

Chu 2009

Covidence 2017
Drain 2014a

Drain 2014c

Ford 2016

Gupta 2015

Gupta-Wright 2016

Gupta-Wright 2018

Gupta-Wright 2018a

Harris 2009
Kohli 2018

Lawn 2012

Lawn 2012a

Lawn 2013a

Lawn 2014a

Lawn 2015

Lawn 2016

Lunn 2009

Macaskill 2010

Minion 2011

Nel 2017

Peter 2010

Qvist 2014

R Statistical Computing 2015

Reitsma 2005

Review Manager

Sabur 2017
Sabur NF, Esmail Aliasgar, 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;17:803.
Schünemann 2008

Schünemann 2016

OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark https://www.sdu.dk/ki/open (Accessed 10 Jun 2018).

Shah 2010

Spiegelhalter 2004
Spiegelhalter DJ, Abrams KR., Myles JP. Bayesian approaches to clinical trials and health-care evaluation [Bayesian approaches to clinical trials and health-care evaluation]. Wiley 2004.

StataCorp 2017

TB CARE I 2014

Whiting 2011

WHO End TB 2014

WHO Global Report 2018
WHO Lipoarabinomannan Policy Guidance 2015

WHO TTP 2014

World Bank 2018/2019

Other published versions of this review
Shah 2014

Shah 2016
Appendices

Appendix 1. Reference card grading of Alere Determine™ TB LAM

Current Reference Card (after 2014)

<table>
<thead>
<tr>
<th>Grade</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>[image]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>[image]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prior Reference Card (before 2014)

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>[image]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>[image]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Control
Patient
Specimen
Appendix 2. PICO questions

1. **What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in all HIV-positive adults and children with signs and symptoms of TB?**
   a. in inpatient settings (adults, adolescents and older children)
   b. in outpatient settings (adults, adolescents and older children)
   c. all settings (adults, adolescents and older children)
   d. in inpatient settings (children ≤ 5 years)
   e. in outpatient settings (children ≤ 5 years)
   f. all settings (children ≤ 5 years)

2. **What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in all HIV-positive adults and children irrespective of signs and symptoms of TB?**
   a. in inpatient settings (adults, adolescents and older children)
   b. in outpatient settings (adults, adolescents and older children)
   c. all settings (adults, adolescents and older children)
   d. in inpatient settings (children ≤ 5 years)
   e. in outpatient settings (children ≤ 5 years)
   f. all settings (children ≤ 5 years)

3. **What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in adults with advanced HIV disease irrespective of signs and symptoms of TB?**
   a. in inpatient setting CD4 ≤ 200
   b. in outpatient setting CD4 ≤ 200
   c. in all settings CD4 ≤ 200
   d. in inpatient setting CD4 ≤ 100
   e. in outpatient setting CD4 ≤ 100
   f. in all settings CD4 ≤ 100

4. **Can the use of LF-LAM in HIV-positive adults reduce mortality associated with advanced HIV disease?**
   a. in all settings
   b. in inpatient settings
   c. in outpatient settings
   d. in individuals with CD4 ≤ 200
   e. in inpatient setting CD4 ≤ 200
   f. in outpatient setting CD4 ≤ 200
   g. in individuals with CD4 ≤ 100
   h. in inpatient setting CD4 ≤ 100
   i. in outpatient setting CD4 ≤ 100

**Other questions:** What is the cost and cost-effectiveness of LF-LAM implementation for TB diagnosis, based on review of the published literature?
Appendix 3. Detailed search strategies

**MEDLINE (Pubmed) search history**

<table>
<thead>
<tr>
<th>Search</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>#9</td>
<td>Search (#3) AND (#7) AND #8</td>
</tr>
<tr>
<td>#8</td>
<td>Search test OR assay OR antigen OR Ag OR lateral flow assay* OR urine antigen OR point of care Field: Title/Abstract</td>
</tr>
<tr>
<td>#7</td>
<td>Search (#4) OR #5) OR #6</td>
</tr>
<tr>
<td>#6</td>
<td>Search LAM; Field: Title/Abstract</td>
</tr>
<tr>
<td>#5</td>
<td>Search &quot;lipoarabinomannan&quot; [Supplementary Concept]</td>
</tr>
<tr>
<td>#4</td>
<td>Search lipoarabinomannan ; Field: Title/Abstract</td>
</tr>
<tr>
<td>#3</td>
<td>Search (#1) OR #2</td>
</tr>
<tr>
<td>#2</td>
<td>Search tuberculosis Or TB Field: Title/Abstract</td>
</tr>
<tr>
<td>#1</td>
<td>Search (&quot;Tuberculosis&quot;[Mesh]) OR &quot;Mycobacterium tuberculosis&quot;[Mesh]</td>
</tr>
</tbody>
</table>

**Database: 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
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 4. Data collection form, diagnostic accuracy

**AlereLAM - Lateral flow urine lipoarabinomannan assay for diagnosing active tuberculosis in people living with HIV**

### Data form

<table>
<thead>
<tr>
<th>I. STUDY IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 First author</td>
</tr>
<tr>
<td>2 Corresponding author and email</td>
</tr>
<tr>
<td>3 Title of study</td>
</tr>
<tr>
<td>4 Year of publication</td>
</tr>
<tr>
<td>5 Year of study start</td>
</tr>
<tr>
<td>6 Language if other than English</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. STUDY DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Population</td>
</tr>
<tr>
<td>8 In which country or countries was the study conducted?</td>
</tr>
<tr>
<td>9 Country World Bank Classification (income)</td>
</tr>
<tr>
<td>10 Country WHO classification for high TB burden country (WHO 2015)</td>
</tr>
<tr>
<td>11 Study design</td>
</tr>
<tr>
<td>12 Was a case-control design avoided?</td>
</tr>
</tbody>
</table>

### Table

<table>
<thead>
<tr>
<th>Population</th>
<th>1. Adults (15 years of age)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Children and adolescents</td>
</tr>
<tr>
<td></td>
<td>3. Both adults, children and adolescents</td>
</tr>
<tr>
<td></td>
<td>4. Other</td>
</tr>
<tr>
<td></td>
<td>If other, describe</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country World Bank Classification (income)</th>
<th>1. Low income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Lower-middle income</td>
</tr>
<tr>
<td></td>
<td>3. Upper-middle-income</td>
</tr>
<tr>
<td></td>
<td>4. High income</td>
</tr>
<tr>
<td></td>
<td>7. Other combination</td>
</tr>
<tr>
<td></td>
<td>9. Unknown/Not reported</td>
</tr>
<tr>
<td></td>
<td>If other, describe:</td>
</tr>
</tbody>
</table>

| Country WHO classification for high TB burden country (WHO 2015) | 1. Yes, part of the High TB/HIV burden list |
|                                                                  | 2. No, not part of the High TB/HIV burden list |

<table>
<thead>
<tr>
<th>Study design</th>
<th>1. Randomized controlled trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Cross-sectional</td>
</tr>
<tr>
<td></td>
<td>3. Cohort</td>
</tr>
<tr>
<td></td>
<td>7. Other, specify</td>
</tr>
<tr>
<td></td>
<td>9. Could not tell</td>
</tr>
<tr>
<td></td>
<td>If other, describe:</td>
</tr>
</tbody>
</table>

| Was a case-control design avoided? | 1. Yes |
|                                   | 2. No  |
|                                   | 9. Unclear |
### III. PATIENT SELECTION

<table>
<thead>
<tr>
<th>13</th>
<th>What was the manner of participant selection into the study?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Consecutive</td>
</tr>
<tr>
<td>2.</td>
<td>Random</td>
</tr>
<tr>
<td>3.</td>
<td>Convenience</td>
</tr>
<tr>
<td>7.</td>
<td>Other, specify</td>
</tr>
<tr>
<td>9.</td>
<td>Unknown/Not Reported/Unclear</td>
</tr>
<tr>
<td>If other, describe:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14</th>
<th>Direction of study data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Prospective</td>
</tr>
<tr>
<td>2.</td>
<td>Retrospective</td>
</tr>
<tr>
<td>9.</td>
<td>Unknown/Not reported</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>15</th>
<th>Please select the statement that best describes the selection of participants into the study.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>HIV-positive participants with signs or symptoms suggestive of active TB were tested using AlereLAM. Please provide study definition of ‘signs and symptoms’:</td>
</tr>
<tr>
<td>2.</td>
<td>A predetermined target population of HIV-positive individuals, irrespective of signs and symptoms of TB, were tested using AlereLAM. Please specify target population:</td>
</tr>
<tr>
<td>3.</td>
<td>Both 1 and 2</td>
</tr>
<tr>
<td>4.</td>
<td>Neither 1 nor 2.</td>
</tr>
<tr>
<td>This is what was done:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>16</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>______</td>
</tr>
<tr>
<td>9.</td>
<td>Unknown/Not reported</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>17</th>
<th>Did the study avoid inappropriate exclusions?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Yes</td>
</tr>
<tr>
<td>2.</td>
<td>No</td>
</tr>
<tr>
<td>9.</td>
<td>Unknown/Not reported/Unclear</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>17a</th>
<th>Could the selection of patients have introduced bias?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>High risk</td>
</tr>
<tr>
<td>2.</td>
<td>Low risk</td>
</tr>
<tr>
<td>9.</td>
<td>Unclear risk</td>
</tr>
</tbody>
</table>

### IV. PATIENT CHARACTERISTICS AND SETTING

<table>
<thead>
<tr>
<th>19</th>
<th>Presenting signs and symptoms</th>
<th>List</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>20</th>
<th>Age (years)</th>
<th>If age is reported in median indicate IQR If age is reported in mean indicate SD</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>21</th>
<th>Age of all study participants, Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>Lower</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>HIV infection (%)</td>
</tr>
<tr>
<td>23</td>
<td>Participants included of female sex (%)</td>
</tr>
<tr>
<td>24</td>
<td>CD4</td>
</tr>
<tr>
<td>25</td>
<td>Number (percent) of TB cases in the study (%)</td>
</tr>
<tr>
<td>26</td>
<td>What was the target condition? 1.Pulmonary TB 2.Extra pulmonary TB 3.Mycobacteraemia 4.Both 1 and 2 5.Any of 1,2,3 7.Other, specify</td>
</tr>
<tr>
<td>27</td>
<td>Did the study include patients with prior TB history? 1.Yes 2.No 9.Unknown/Not reported If yes, what is the % _____ Specify the numerator/denominator <em><strong><strong>/</strong></strong></em></td>
</tr>
<tr>
<td>28</td>
<td>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</td>
</tr>
<tr>
<td>29</td>
<td>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</td>
</tr>
<tr>
<td>30</td>
<td>Are there concerns that the included patients and setting do not match the review question? 1.High concern 2.Low concern 9.Unclear concern</td>
</tr>
<tr>
<td>31</td>
<td>NOTES ON CHARACTERISTICS</td>
</tr>
<tr>
<td>V. INDEX TEST</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Was a AlereLAM 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 ¼ 4.Yes, Grade 2/4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
| 33 | What AlereLAM threshold was used to define positivity for data extraction? | 1. Grade 2/5  
2. Grade 1/4  
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 AlereLAM 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 AlereLAM 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 AlereLAM 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 TB, what reference standard was used to identify TB and not TB? | 1. Sputum: solid culture  
2. Sputum: liquid culture  
3. Sputum: both solid and liquid culture  
4. Nucleic acid amplification test, specify  
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 |
<p>| 42 | Were patients without sputum specimens (for example, no expectorated, no induced sputum) included in this study? | 1. Yes |</p>
<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify N/% included without sputum ____</td>
<td>2.No</td>
</tr>
<tr>
<td>Specify N/% excluded due to lack of sputum____</td>
<td></td>
</tr>
<tr>
<td>Were non-pulmonary specimens evaluated to allow diagnosis of extrapulmonary TB?</td>
<td>1. All participants received testing of non-pulmonary specimens, please specify sites/fluids:</td>
</tr>
<tr>
<td></td>
<td>2. Some participants received testing of non-pulmonary specimens, please specify which patients were tested, and sites/fluids:</td>
</tr>
<tr>
<td></td>
<td>3. Extrapulmonary TB was not evaluated</td>
</tr>
<tr>
<td></td>
<td>7. Other, please specify:</td>
</tr>
<tr>
<td>For the diagnosis of extrapulmonary TB, what tests were used to identify TB and not TB (circle all that apply)?</td>
<td>1. Solid culture</td>
</tr>
<tr>
<td></td>
<td>2. Liquid culture</td>
</tr>
<tr>
<td></td>
<td>3. Both solid and liquid culture</td>
</tr>
<tr>
<td></td>
<td>4. Nucleic acid amplification test, specify</td>
</tr>
<tr>
<td></td>
<td>7. Other, specify:</td>
</tr>
<tr>
<td></td>
<td>8. Not applicable, extrapulmonary TB was not evaluated</td>
</tr>
<tr>
<td>Did the study speciate mycobacteria isolated in culture?</td>
<td>1. Yes</td>
</tr>
<tr>
<td></td>
<td>2. No</td>
</tr>
<tr>
<td></td>
<td>9. Unknown/Not reported</td>
</tr>
<tr>
<td>Was the reference standard likely to correctly classify the target condition</td>
<td>1. Yes</td>
</tr>
<tr>
<td></td>
<td>2. No</td>
</tr>
<tr>
<td></td>
<td>9. Unclear</td>
</tr>
<tr>
<td>Was the reference standard result interpreted without knowledge of the result of AlereLAM?</td>
<td>1. Yes</td>
</tr>
<tr>
<td></td>
<td>2. No</td>
</tr>
<tr>
<td></td>
<td>9. Unclear</td>
</tr>
<tr>
<td>How many sputum specimens were obtained in order to detect pulmonary TB?</td>
<td>1. Single</td>
</tr>
<tr>
<td></td>
<td>2. Multiple</td>
</tr>
<tr>
<td></td>
<td>8. Not applicable</td>
</tr>
<tr>
<td>How many specimens from fluid (sites) other than sputum were obtained to detect extrapulmonary TB?</td>
<td>1. Single</td>
</tr>
<tr>
<td></td>
<td>2. Multiple</td>
</tr>
<tr>
<td></td>
<td>8. Not applicable</td>
</tr>
<tr>
<td>Could the reference standard, its conduct, or its interpretation have introduced bias?</td>
<td>1. High risk</td>
</tr>
<tr>
<td></td>
<td>2. Low risk</td>
</tr>
<tr>
<td></td>
<td>9. Unclear risk</td>
</tr>
</tbody>
</table>
| 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: TB detection against a microbiological reference standard

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

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

(Table example to extract TP, FP, FN, TN values)

<table>
<thead>
<tr>
<th>LAM result</th>
<th>TB</th>
<th>Not TB</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Provide additional tables for each of the applicable PICO questions 1-4 (Appendix 2. PICO questions) and the following additional questions:
5. What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in all HIV-positive adults with advanced HIV disease and signs and symptoms of TB?
   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 101-200
   j. in outpatient settings CD4 101-200
   k. in all settings CD4 101-199

6. What is the diagnostic accuracy of LF-LAM for the diagnosis of TB in adults with advanced HIV disease irrespective of signs and symptoms of TB?
   a. in all settings CD4 > 200
   b. in all settings CD4 > 100
   c. in inpatients settings CD4 101-200
   d. in outpatient settings CD4 101-200
   e. in all settings CD4 101-199
# Appendix 5. Data collection form, impact data

**AlereLAM - Lateral flow urine lipoarabinomannan assay for diagnosing active tuberculosis in people living with HIV**

**Data form for impact data extraction**

<table>
<thead>
<tr>
<th>I. STUDY IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 First author</td>
</tr>
<tr>
<td>2 Journal</td>
</tr>
<tr>
<td>3 Year of publication</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. STUDY DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Population</td>
</tr>
<tr>
<td>Adults or children</td>
</tr>
<tr>
<td>HIV status</td>
</tr>
<tr>
<td>Other details re: study inclusion criteria</td>
</tr>
<tr>
<td>In which country or countries was the study conducted?</td>
</tr>
<tr>
<td>List all countries:</td>
</tr>
<tr>
<td>Study design</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>Cohort</td>
</tr>
<tr>
<td>Cross-sectional</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. PATIENT SELECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Direction of study data collection</td>
</tr>
<tr>
<td>Prospective</td>
</tr>
<tr>
<td>Retrospective</td>
</tr>
<tr>
<td>Unknown/Not reported</td>
</tr>
<tr>
<td>8 Please select the statement that best describes the selection of participants into the study.</td>
</tr>
<tr>
<td>HIV-positive participants with signs or symptoms suggestive of active TB were tested using AlereLAM. Please provide study definition of ‘signs and symptoms’:</td>
</tr>
<tr>
<td>A predetermined target population of HIV-positive individuals, irrespective of signs and symptoms of TB, were tested using AlereLAM. Please specify target population:</td>
</tr>
<tr>
<td>Both 1 and 2</td>
</tr>
<tr>
<td>Neither 1 nor 2.</td>
</tr>
<tr>
<td>This is what was done:</td>
</tr>
<tr>
<td>Number enrolled</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
</tbody>
</table>

### IV. PATIENT CHARACTERISTICS AND SETTING

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>If age is reported in median indicate IQR If age is reported in mean indicate SD</th>
</tr>
</thead>
</table>

| What was the clinical setting of the study? | Outpatient  
Inpatient  
Both out-patient and in-patient  
Other, describe:  
Unknown/Not reported |
|--------------------------------------------|---------------------------------------------------------------|

### V. INDEX TEST (LAM)

| What AlereLAM threshold was used to define positivity for data extraction? | Grade 2/5  
Grade 1/4  
Other, specify |
|---------------------------------------------------------------------------|---------------------------------------------------------------|

| Was old or new AlereLAM card used | Old (5 grades)  
New (4 grades) |
|----------------------------------|---------------------------------------------------------------|

### VI. MORTALITY ASSESSMENT

<table>
<thead>
<tr>
<th>How was mortality assessed?</th>
<th>Describe</th>
</tr>
</thead>
</table>

| Type of mortality? | All-cause  
TB-related |
|--------------------|---------------------------------------------------------------|

<table>
<thead>
<tr>
<th>When was mortality assessed?</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality (ARR)</th>
<th>Describe results</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality (HR, aHR, MHR, or Kaplan Meier)</th>
<th>Describe results</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality (OR or aOR)</th>
<th>Describe results</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>What was the comparator?</th>
<th>Describe</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality in intervention</th>
<th>Percentage and number</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality in control</th>
<th>Percentage and number</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality in LAM positive</th>
<th>Percentage and number</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality in LAM negative</th>
<th>Percentage and number</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality in LAM positive confirmed TB cases</th>
<th>Percentage and number</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Mortality in LAM negative confirmed TB cases</th>
<th>Percentage and number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>28</td>
<td>Mortality in LAM positive patients with an inconclusive evaluation for TB</td>
</tr>
<tr>
<td>29</td>
<td>Mortality in LAM negative patients with an inconclusive evaluation for TB</td>
</tr>
<tr>
<td>30</td>
<td>Mortality in LAM positive non-TB cases</td>
</tr>
<tr>
<td>31</td>
<td>Mortality in LAM negative non-TB cases</td>
</tr>
<tr>
<td>32</td>
<td>Was analysis stratified by CD4 count or percentage?</td>
</tr>
<tr>
<td>33</td>
<td>Time to diagnosis</td>
</tr>
<tr>
<td>34</td>
<td>Time to treatment</td>
</tr>
<tr>
<td>35</td>
<td>Other outcomes assessed in the study</td>
</tr>
<tr>
<td>36</td>
<td>Comments</td>
</tr>
</tbody>
</table>
Appendix 6. QUADAS-2

Domain 1: patient selection

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

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

Signalling question 2: Was a case-control design avoided?
We answered ‘yes’ to all included studies given that we are excluding case-control study designs.

Signalling 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 TB. We scored ‘unclear’ if we could not tell.

Applicability: Are there concerns that the included patients and setting do not match the review question?
We were interested in how AlereLAM 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 TB and patients investigated for TB irrespective of signs and symptoms for TB.

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

For AlereLAM used as a TB diagnostic test among patients that were investigated for TB irrespective of signs and symptoms of TB, we judged 'low concern' for studies in which the AlereLAM was performed uniformly within the predetermined study target populations of HIV-infected individuals, 'high concern' if AlereLAM 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/TB coinfection.

Domain 2: index test
Risk of bias: could the conduct or interpretation of the index test have introduced bias?

Signalling question 1: were the index test results interpreted without knowledge of the results of the reference standard?
We answered 'yes' if the study interpreted the result of AlereLAM blinded to the result of the reference standard; we answered 'no' if the study did not interpret the result of AlereLAM blinded to the result of the reference standard. We answered 'yes' for studies in which AlereLAM 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.

Signalling question 2: if an AlereLAM 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?

Signalling question 1: is the reference standard likely to correctly classify the target condition?
HIV-infected TB patients may have pulmonary TB, extrapulmonary TB, or both pulmonary and extrapulmonary TB. A microbiological reference standard, primarily culture, is considered the gold standard for TB. Due to the difficulties in diagnosing HIV-associated TB, it is recommended that multiple cultures from sputum and other specimens be evaluated.

We answered 'yes' when appropriate specimens were obtained for the diagnosis of HIV-associated TB. For presumed pulmonary TB, sputum specimens should be obtained for culture, NAAT, or both culture and NAAT. If the patient cannot produce sputum, induced sputum should be
performed. For presumed extrapulmonary TB, specimens should be consistent with Standard 4 of
the International Standards for TB Care which states: "For all patients, including children,
suspected of having extrapulmonary tuberculosis, appropriate specimens from the suspected sites of
involvement should be obtained for microbiological and histological examination" (TB CARE I
2014). We answered yes if multiple specimens were collected from different sites for
extrapulmonary TB. An Xpert® MTB/RIF test is recommended as the preferred initial
microbiological test for suspected TB meningitis because of the need for a rapid diagnosis". We
also answered 'yes' if studies followed a standardized approach of collecting appropriate specimens
from "suspected sites of involvement", for example, blood or lymph nodes on all patients.
We answered 'no' when the reference standard was restricted to sputum specimens or the reference
standard was restricted to extrapulmonary specimens (for example, urine, blood, etc.). We also
answered 'no' if a consistent approach was not followed for all patients (for example, some but not
all patients with presumed TB lymphadenitis receive lymph node tissue sampling). We answered
'unclear' if we could not tell.

Signalling question 2: were the reference standard results interpreted without knowledge of
the results of the index test?
We answered 'yes' if the study interpreted the result of the reference standard blinded to the result of
AlereLAM, 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 AlereLAM, and
'unclear' if we could not tell.

Applicability: are there concerns that the target condition as defined by the reference
standard does not match the question?
In general, we thought there was low concern for almost included studies based on the current
definitions of the reference standard. We judged 'high concern' if included studies did not speciate
mycobacteria isolated in culture, 'low concern' if speciation was performed, and 'unclear' if we
could not tell. We also judged high concern if there was no protocol to ensure a minimum standard
of testing with a reference standard.

Domain 4: Flow and timing
Risk of bias: could the patient flow have introduced bias?
Signalling question 1: was there an appropriate interval between the index test and reference
standard?
We expected urine specimens for AlereLAM and the reference standards to be obtained at the same
time and answered 'yes' for all studies that meet this criterion, or if index and reference standard
tests were performed on specimens collected no greater than seven days apart. We chose seven days
as a time period during which either treatment of TB or natural progression of TB without treatment
could impact test results. We answered 'no' if specimens were collected for index and reference
standard tests greater than seven days apart, and 'unclear' if we could not tell.
Signalling question 2: did all patients receive the same reference standard?
We answered 'yes' if all participants in the study received the reference standard to confirm TB; 'no' if not all patients received the reference standard to confirm TB, and 'unclear' if we could not tell.

Signalling question 3: were all patients included in the analysis?
We determined the answer to this question by comparing the number of participants enrolled in the study with the number of participants included in the two-by-two tables. We answered 'yes' if all participants enrolled in the study were tested with results presented and accounted for. We answered 'no' if participants meeting enrolment criteria were not tested or results were not presented, and 'unclear' if we could not tell.

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 approach

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 TPi, FPi, TNi, FNi. The sensitivity in i-th study is denoted by sei and the specificity by spi.

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 $\mu$ 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:
TPi $\sim$ Binomial(TPRi, TPi + FNi), and
FPi $\sim$ Binomial(FPRi, TNi + FPi)

Between studies:
The bivariate vector (logit(TRPi), logit(FPRi)) $\sim$ BVN($\mu = (\mu_1, \mu_2)$, TAU) where TAU is a 2 X 2 matrix with entries

TAU[1,1] = variance of logit(TPRi) = tau12,

TAU[2,2] = variance of logit(FPRi) = tau22 and

TAU[1,2] = TAU[2,1] = covariance between logit(TPRi) and logit(FPRi) = rho $\times$ tau1 $\times$ tau2

and rho is the correlation between logit(TPRi) and logit(FPRi) across studies.

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

Prior distributions:

\(\mu_1\) and \(\mu_2\) \(\sim\) \(N(m=0, \tau_2=4)\),

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

\((1/\tau_{12})\) and \((1/\tau_{22})\) \(\sim\) \(\text{Gamma}(\text{shape}=2, \text{rate}=0.5)\)

**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.

```openbugs
define {
L = \text{number of studies in the Meta-analysis}
}

for(i in 1:L) {  ## L is the number of studies in the Meta-analysis

  # Likelihood
  pos[i]<-TP[i]+FN[i]
  neg[i]<-TN[i]+FP[i]

  TP[i] ~ dbin(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] ~ dnorm(mu[1:2], T[1:2, 1:2])
}

# Prior Distributions
```

97
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[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]))

}
## Appendix 8: Characteristics of Included Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants (% symptomatic)</th>
<th>Setting</th>
<th>Median CD4 cell count per µL (IQR)</th>
<th>TB prevalence % (n/N)</th>
<th>Did the study avoid inappropriate exclusion</th>
<th>Specimens collected</th>
<th>High quality reference standard*</th>
<th>Unique Study Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drain 2016</td>
<td><strong>Symptomatic:</strong> Two of four TB related symptoms (cough, fever weight loss, night sweat) for &gt; 2 weeks; smear microscopy negative x 2</td>
<td>Outpatient</td>
<td>168 (89-256)</td>
<td>63% (57/90)</td>
<td>No</td>
<td>Pulmonary samples</td>
<td>No</td>
<td>Adults (&gt;18 years); HIV positive (93.2%); Targeting a relatively well outpatient population; Karnofsky performance score &gt;50</td>
</tr>
<tr>
<td>Huerga 2017</td>
<td><strong>Symptomatic:</strong> Cough &gt; 2 weeks or any cough and one of weight loss, night sweats or fever; severely ill; CD4 &lt; 200 or BMI below 17</td>
<td>Outpatients (67%); Inpatients (33%)</td>
<td>109 (43-214)</td>
<td>57% (156/275)</td>
<td>No</td>
<td>Pulmonary samples; Urine Xpert only for patients without sputum available</td>
<td>No</td>
<td>Adults (&gt;15 years); LAM guided treatment; Excluded many participants from analysis**</td>
</tr>
<tr>
<td>Juma 2017</td>
<td><strong>Symptomatic:</strong> Suggestive of extrapulmonary TB, not specified</td>
<td>Inpatients</td>
<td>not stated</td>
<td>33% (29/87)</td>
<td>No</td>
<td>Extrapulmonary samples only, no sputum samples</td>
<td>No</td>
<td>Adults (&gt;14 years), HIV- positive (68%); Excluded patients with concomitants active pulmonary TB</td>
</tr>
<tr>
<td>Nakiyingi 2014</td>
<td><strong>Symptomatic:</strong> Any of cough, fever weight loss, night sweat</td>
<td>Outpatients (45%); Inpatients (55%)</td>
<td>152 (41-337)</td>
<td>37% (367/997)</td>
<td>No</td>
<td>Pulmonary samples; Blood culture for all</td>
<td>Yes</td>
<td>Adults (&gt;18 years); multisite; large sample size</td>
</tr>
<tr>
<td>Pandie 2016</td>
<td><strong>Symptomatic:</strong> Presence of a pericardial effusion and suspected of pericardial TB</td>
<td>Inpatients</td>
<td>139 (81-249)</td>
<td>95% (36/38)</td>
<td>No</td>
<td>Extrapulmonary samples (pericardial effusion); pulmonary samples for some</td>
<td>No</td>
<td>Adults (&gt;18 years); HIV-positive (74%); Excluded participants from analysis affecting specificity***</td>
</tr>
<tr>
<td>Peter 2012</td>
<td><strong>Symptomatic:</strong> Any of cough, fever weight loss, night sweat</td>
<td>Inpatients</td>
<td>90 (47-197)</td>
<td>48% (116/241)</td>
<td>Yes</td>
<td>Clinically relevant pulmonary samples; clinically relevant extrapulmonary samples. No study defined algorithm.</td>
<td>No</td>
<td>Adults (&gt; 18 years). Multisite; TB diagnostic work-up was not standardised but up to clinical judgements</td>
</tr>
<tr>
<td>Peter 2015</td>
<td><strong>Symptomatic:</strong> Any of cough, fever weight loss, night sweat</td>
<td>Outpatient</td>
<td>210 (103-375)</td>
<td>32% (181/569)</td>
<td>No</td>
<td>Pulmonary samples</td>
<td>No</td>
<td>Adults (&gt; 18 years), Multisite; nested within a randomised, parallel-arm trial,</td>
</tr>
<tr>
<td>Peter 2016</td>
<td><strong>Symptomatic:</strong> Any of cough, fever weight loss, night sweat</td>
<td>Inpatients</td>
<td>81 (26-198)</td>
<td>29% (342/1172)</td>
<td>Yes</td>
<td>Pulmonary samples; Clinically relevant extrapulmonary samples. No study defined algorithm.</td>
<td>No</td>
<td>Adults (&gt;18 years); Multisite; LAM arm of a randomised controlled trial</td>
</tr>
</tbody>
</table>

HIV positive adults irrespective of signs and symptoms of TB
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants (% symptomatic)</th>
<th>Setting</th>
<th>Median CD4 cell count per µL (IQR)</th>
<th>TB prevalence % (n/N)</th>
<th>Did the study avoid inappropriate exclusion</th>
<th>Specimens collected</th>
<th>High quality reference standard*</th>
<th>Unique Study Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bjerrum 2015</td>
<td>Unselected; 91% symptomatic</td>
<td>Outpatients</td>
<td>127 (35-256)</td>
<td>12% (55/469)</td>
<td>No</td>
<td>Pulmonary samples</td>
<td>No</td>
<td>Adults (&gt;18 years); Majority symptomatic.</td>
</tr>
<tr>
<td>Drain 2015</td>
<td>Unselected; proportion symptomatic not stated</td>
<td>Outpatient</td>
<td>248 (107-379)</td>
<td>17% (54/320)</td>
<td>No</td>
<td>Pulmonary samples</td>
<td>No</td>
<td>Adults (&gt;18 years)</td>
</tr>
<tr>
<td>Floridia 2017</td>
<td>Unselected; 34% symptomatic</td>
<td>Outpatient</td>
<td>278 (142-395)</td>
<td>9% (90/972)</td>
<td>No</td>
<td>Pulmonary samples</td>
<td>No</td>
<td>Adults (&gt; 15 years). LAM guided treatment.</td>
</tr>
<tr>
<td>Hanifa 2016</td>
<td>Unselected; 53% symptomatic</td>
<td>Outpatient</td>
<td>111 (56-161)</td>
<td>9% (40/408)</td>
<td>Yes</td>
<td>Pulmonary samples; Blood culture for all</td>
<td>Yes</td>
<td>Adults (&gt;18 years); CD4 &lt; 200; Reference standard included any sample taken within six months from enrolment.</td>
</tr>
<tr>
<td>LaCourse 2016</td>
<td>Unselected; 19% symptomatic</td>
<td>Outpatient</td>
<td>437 (342-565)</td>
<td>1% (3/266)</td>
<td>No</td>
<td>Pulmonary samples</td>
<td>No</td>
<td>Pregnant women (&gt;16 years) attending ANC; Healthy population; one person with CD4 &lt; 400; Few TB cases (n=3).</td>
</tr>
<tr>
<td>Lawn 2017</td>
<td>Unselected; 91% symptomatic</td>
<td>Inpatients</td>
<td>149 (55-312)</td>
<td>33% (139/413)</td>
<td>Yes</td>
<td>Pulmonary samples; Blood culture for all; Clinically relevant extrapulmonary samples</td>
<td>Yes</td>
<td>Adults (&gt;18 years). Included many samples from different sites</td>
</tr>
<tr>
<td>Thit 2017</td>
<td>Unselected; 33% symptomatic</td>
<td>Outpatients</td>
<td>270 (128-443)</td>
<td>10% (54/517)</td>
<td>Yes</td>
<td>Pulmonary samples</td>
<td>No</td>
<td>Adults (median age 34). Reference standard included samples taken within six months from enrolment.</td>
</tr>
</tbody>
</table>

Abbreviations: TB: tuberculosis; AlereLAM: AlereLAM: Alere Determine™ TB lipoarabinomannan assay; Xpert: Xpert MTB/RIF

* 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.

**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).

*** 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).
Appendix 9: Diagnostic accuracy of AlereLAM among HIV-positive children, summary

Background
Among children in 2014, there were an estimated one million incident TB cases and 140,000 deaths attributable to TB. Approximately 40% of TB deaths were among those coinfected with TB and HIV (Carlucci 2017).

Diagnosis of TB in children
Conventional diagnosis by culture or microscopy yields a positive result in less than 50% of children with clinically diagnosed TB and HIV infection (Thomas 2016). Active TB, therefore, remains unrecognised in a large number of children in high burden countries as evident from autopsy studies from five African countries that identified TB in roughly 10% of 811 children (both HIV-positive and HIV-negative) who died from presumed pneumonia (Bates 2013).

To obtain specimens for microscopy or culture, methods used in children include gastric lavage (GL), which requires uncomfortable insertion of a nasogastric tube and induced sputum. Sputum induction requires special facilities (negative pressure) for infection control and nebulization equipment driven by high flow air or oxygen, not available in rural areas of low-income countries. Using sputum induction for children with presumed pulmonary TB, a study conducted in South Africa reported microbiologic confirmation in 11% of cases (Connell 2011). A lower yield (3.8% to 7%) has been reported for nasopharyngeal aspirate. Bronchoalveolar lavage (BAL) is a resource-intensive and invasive procedure that has a lower yield for culture, compared with GL; therefore, BAL is not indicated for microbiologic confirmation of TB in children.

The WHO recommends the use of Xpert MTB/RIF (Xpert) as follows.
1) Xpert should be used rather than conventional microscopy and culture as the initial diagnostic test in children suspected of having MDR-TB or HIV-associated TB (strong recommendation, very low-certainty evidence)
2) Xpert may be used rather than conventional microscopy and culture as the initial test in all children suspected of having TB (Conditional recommendation acknowledging resource implications, very low-certainty evidence)
3) Xpert may be used as a replacement test for usual practice (including conventional microscopy, culture, and/or histopathology) for testing of specific non-respiratory specimens (lymph nodes and other tissues) from children suspected of having extrapulmonary TB (conditional recommendation, very low-certainty evidence) and
4) Xpert should be used in preference to conventional microscopy and culture as the initial diagnostic test in testing cerebrospinal fluid specimens from children suspected of having TB meningitis (strong recommendation given the urgency of rapid diagnosis, very low-certainty evidence) (WHO 2014).
The next-generation assay, Xpert MTB/RIF Ultra (Ultra), has shown improved sensitivity for detection of TB in HIV-positive people. The WHO now recommends Ultra as a replacement for the current Xpert MTB/RIF cartridge (WHO Ultra 2017).

Methods
We performed literature searches up to 11 May 2018 as part of a larger search for studies in adults with the same inclusion criteria except for age. We included studies that evaluated Alere Determine™ TB LAM Ag test (AlereLAM) on urine specimens. The target condition was active TB disease, which includes pulmonary and extrapulmonary TB. Age groups were defined as younger children ≤ 5 years; adolescents, 10 to 19 years; and older children, 6 to 19 years. Two review authors independently extracted data on methodological quality and 2x2 values for AlereLAM for TB against a microbiological reference standard. Given the differences in population and setting, we did not perform meta-analyses and provide sensitivity and specificity estimates for individual studies.

Results
We identified three published studies involving 266 HIV-positive children that evaluated the accuracy of AlereLAM for TB as the result of a broader search for studies in adults and children using the same inclusion criteria (Kroidl 2015; LaCourse 2018; Nicol 2014). All three studies took place in high TB/HIV burden countries in Africa: Kroidl 2015 in Tanzania; LaCourse 2018 in Kenya; and Nicol 2014 in South Africa.

Methodological quality of included studies
In the Patient Selection Domain, we considered two studies (67%) to have low risk of bias because the study used consecutive or random enrolment of participants and avoided inappropriate exclusions (Kroidl 2015; LaCourse 2018). We considered one study to have high risk of bias because children who could not produce sputum despite sputum induction were excluded (Nicol 2014). In the Index Test we considered all three studies to have low risk of bias. In the Reference Standard Domain, we considered two studies (67%) to have low risk of bias and one study to have high risk of bias because we thought the reference standard used was unlikely to correctly classify the target condition (Nicol 2014). In the Flow and Timing Domain we considered one study to have high risk of bias because not all patients received the same reference standard (Kroidl 2015). Applicability in all domains was of low concern in all three studies, (Figure A1 and Figure A2).
Figure A1. Risk of bias and applicability concerns graph.
Review authors’ judgements about each domain presented as percentages across included studies in children.

Figure A2. Risk of bias and applicability concerns summary.
Review authors’ judgements about each domain for each included study in children.

Findings
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 2018 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.

Kroidl 2015 and Nicol 2014 involved HIV-positive children with TB symptoms. LaCourse 2018 involved HIV-positive children hospitalized for acute illness irrespective of TB signs and symptoms. Kroidl 2015 was conducted in an outpatient setting, LaCourse 2018 in an inpatient setting, and Nicol 2014 in both an inpatient and an outpatient setting. The prevalence of microbiologically-confirmed TB in the studies was 40% in Kroidl 2015, 7% in LaCourse 2018, and 22% in Nicol 2014. Regarding immunosuppression, in Kroidl 2015, 65% of children had advanced
or severe immunosuppression; in LaCourse 2018, 70% of children had severe immunosuppression; and in Nicol 2014, 53% of children had advanced or severe immunosuppression. See Table.

Table. Characteristics of Included Studies, Children

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

* For all participants, including HIV-positive and HIV-negative children.

IQR: Interquartile range

We first present a summary and then present findings for the specific PICO questions.

**AlereLAM testing, studies in children with TB symptoms and irrespective of TB signs and symptoms**

**All settings**

We identified three studies involving 266 children (Kroidl 2015; LaCourse 2018; Nicol 2014), Figure A3. In all settings, including all children, sensitivity and specificity (95% CI) were 42% (15% to 72%) and 94% (73% to 100%), (30 participants, outpatient) Kroidl 2015; 56% (21% to 86%) and 95% (90% to 98%), (130 participants, inpatient) LaCourse 2018; and 43% (23% to 66%) and 80% (69% to 88%), (106 participants, both inpatient and outpatient) Nicol 2014.

**Figure A3. Forest plots of AlereLAM sensitivity and specificity for TB in HIV-positive children with TB symptoms and irrespective of TB signs and symptoms, all settings.**
By age group
Kroidl 2015 and LaCourse 2018 provided AlereLAM accuracy data by age group. Figure A4. Stratified by age group, in adolescents, AlereLAM sensitivities were 100% (33% to 100%) (four participants, inpatient) LaCourse 2018, and 60% (15% to 95%) (nine participants, outpatient) Kroidl 2015; in both studies, specificity was 100%. In children ≤ 5 years, sensitivities were 50% (7% to 93%) (95 participants, inpatient) LaCourse 2018, and 25% (1% to 81%) (13 participants, outpatient) Kroidl 2015; corresponding specificities were 93% (86% to 98%) and 89% (52% to 100%).

Figure A4. Forest plots of AlereLAM sensitivity and specificity for TB in HIV-positive children with TB symptoms and irrespective of TB symptoms, by age group.

<table>
<thead>
<tr>
<th>Adolescents</th>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LaCourse 2018</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1.00 [0.03, 1.00]</td>
<td>1.00 [0.23, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kroidl 2015</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td>0.60 [0.15, 0.95]</td>
<td>1.00 [0.40, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Older children

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaCourse 2018</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>0.60 [0.15, 0.95]</td>
<td>1.00 [0.83, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kroidl 2015</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>0.50 [0.16, 0.84]</td>
<td>1.00 [0.66, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Children ≤ 5

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaCourse 2018</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>88</td>
<td>0.50 [0.07, 0.93]</td>
<td>0.95 [0.86, 0.98]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kroidl 2015</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>0.25 [0.01, 0.81]</td>
<td>0.89 [0.52, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PICO 1. What is the diagnostic accuracy of AlereLAM for the diagnosis of TB in all HIV-positive children with signs and symptoms of TB?

I.a.ii. Inpatient settings, adolescents
We did not identify any studies addressing this question.

I.a.iii. Inpatient settings, older children
We did not identify any studies addressing this question.

I.b.ii. Outpatient settings, adolescents
We identified one study involving nine HIV-positive children, five (56%) with TB, Kroidl 2015. Sensitivity and specificity (95% CI) were 60% (15, 95) and 100% (40, 100).

I.b.iii. Outpatient settings, older children
We identified one study involving 17 HIV-positive children, eight (47%) with TB, Kroidl 2015. Sensitivity and specificity (95% CI) were 50% (16, 84) and 100% (66, 100).

1.c.ii. All settings, adolescents
We identified one study involving nine HIV-positive children, five (56%) with TB, Kroidl 2015. The data are from an outpatient setting. Sensitivity and specificity (95% CI) were 60% (15, 95) and 100% (40, 100).

1.c.iii. All settings, older children
We identified one study involving 17 HIV-positive children, eight (47%) with TB, Kroidl 2015. The data are from an outpatient setting. Sensitivity and specificity (95% CI) were 50% (16, 84) and 100% (66, 100).

1.d. Inpatient settings, children ≤ 5 years
We did not identify any studies addressing this question.

1.e. Outpatient settings, children ≤ 5 years
We identified one study involving 13 HIV-positive children, four (31%) with TB, Kroidl 2015. Sensitivity and specificity (95% CI) were 25% (1, 81) and 89% (52, 100).

1.f. All settings, children ≤ 5 years
We identified one study involving 13 HIV-positive children, four (31%) with TB, Kroidl 2015. The data are from an outpatient setting. Sensitivity and specificity (95% CI) were 25% (1, 81) and 89% (52, 100).

**PICO 2. What is the diagnostic accuracy of AlereLAM for the diagnosis of TB in all HIV-positive children irrespective of signs and symptoms of TB?**

2.a.ii. Inpatient settings, adolescents
We identified one study involving four HIV-positive children, one (25%) with TB, LaCourse 2018. Sensitivity and specificity (95% CI) were 100% (3, 100) and 100% (29, 100).

2.a.iii. Inpatient settings, older children
We identified one study involving 34 HIV-positive children, five (15%) with TB, LaCourse 2018. Sensitivity and specificity (95% CI) were 60% (15, 95) and 100% (88, 100).

2.b.ii. Outpatient settings, adolescents
We did not identify any studies addressing this question.

2.b.iii. Outpatient settings, older children
We did not identify any studies addressing this question.

2.c.ii. All settings, adolescents
We identified one study involving four HIV-positive children, one (25%) with TB, LaCourse 2018. The data are from an inpatient setting. Sensitivity and specificity (95% CI) were 100% (3, 100) and 100% (29, 100).

2.c.iii. All settings, older children
We identified one study involving 34 HIV-positive children, five (15%) with TB, LaCourse 2018. The data are from an inpatient setting. Sensitivity and specificity (95% CI) were 60% (15, 95) and 100% (88, 100).

2.d. Inpatient settings, children ≤ 5 years
We identified one study involving 95 HIV-positive children, four (4%) with TB, LaCourse 2018. Sensitivity and specificity (95% CI) were 50% (7, 93) and 93% (86, 98).

2.e. Outpatient settings, children ≤ 5 years
We did not identify any studies addressing this question.

2.f. All settings, children ≤ 5 years
We identified one study involving 95 HIV-positive children, four (4%) with TB, LaCourse 2018. The data are from an inpatient setting. Sensitivity and specificity (95% CI) were 50% (7, 93) and 93% (86, 98).

Discussion
This systematic review on the urine lateral flow lipoarabinomannan assay, AlereLAM, for active TB in children living with HIV summarizes the current literature and includes three studies. As the studies enrolled children aged 15 years and less and younger children (median age, range 24 months to 6.8 years), the results may not be applicable to older children. All studies took place in high TB/HIV burden countries in Africa. We corresponded with two study authors (Kroidl and LaCourse) to ensure that we had accurate data for AlereLAM applied using the current manufacturer's instructions. In individual studies, AlereLAM sensitivities were 56%, 42%, and 43%; corresponding specificities were 94%, 95%, and 80%.

AlereLAM specificity was lower in children ≤ 5 years than in older children, based on limited data. Urine collection was noted to be difficult in younger and sicker children. In addition, urine collection in children may be affected by dehydration or other medical problems.

Authors' conclusions
We found limited evidence on the accuracy of AlereLAM in children living with HIV. There were too few studies and participants to draw conclusions.
References for included studies
Kroidl 2015

LaCourse 2018

Nicol 2014

Additional references
Bates 2013

Carlucci 2017

Connell 2011

Thomas 2016

WHO 2014

WHO Ultra 2017