Article type: Original article

Title: Diagnostic accuracy of WHO screening criteria to guide lateral-flow lipoarabinomannan testing among HIV-positive inpatients: a systematic review and individual participant data meta-analysis

Ashar Dhana, Yohhei Hamada, Andre P Kengne, Andrew D Kerkhoff, Tobias Broger, Claudia M Denkinger, Molebogeng X Rangaka, Ankur Gupta-Wright, Katherine Fielding, Robin Wood, Helena Huerga, Sekai Chenai Mathabire Rucker, Stephanie Bjerrum, Isik S Johansen, Swe Swe Thit, Mar Mar Kyi, Josh Hanson, David A Barr, Graeme Meintjes, Gary Maartens

Corresponding author:

Gary Maartens

Department of Medicine, Faculty of Health Sciences, University of Cape Town,

Cape Town, South Africa, 7935

Tel: 021-406-6008

E-mail: gary.maartens@uct.ac.za

Name	Highest Degree	Affiliation
Ashar Dhana	MBBCh	Department of Medicine, University of Cape Town, Cape Town, South Africa
Yohhei Hamada	MD	Centre for International Cooperation and Global TB Information, The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan / Institute for Global Health, University College London, London, UK
Prof Andre P Kengne	PhD	Non-communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa
Andrew D Kerkhoff	MD	Division of HIV, Infectious Diseases and Global Medicine, Zuckerberg San Francisco General Hospital and Trauma Center, University of California San Francisco, San Francisco, CA, USA
Tobias Broger	MSc	Division of Tropical Medicine, Center for Infectious Diseases, Heidelberg University Hospital, Heidelberg, Germany/ FIND, Geneva, Switzerland
Claudia M Denkinger	MD	Division of Infectious Disease and Tropical Medicine, Center for Infectious Diseases, Heidelberg University Hospital, Heidelberg, Germany/ German Center of Infection Research, Heidelberg, Germany /FIND, Geneva, Switzerland
Molebogeng X Rangaka	PhD	Institute for Global Health, University College London, London, UK / Wellcome Centre for Infectious Diseases Research in Africa (CIDRI-Africa), Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa
Ankur Gupta-Wright	MD	Institute for Global Health, University College London, London, UK / Clinical Research Department, London School of Hygiene and Tropical Medicine, London, UK
Prof Katherine Fielding	PhD	TB Centre, London School of Hygiene & Tropical Medicine, London, UK
Prof Robin Wood	DSc	Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
Helena Huerga	PhD	Field Epidemiology Department, Epicentre, Paris, France
Sekai Chenai Mathabire Rucker	MPH	Field Epidemiology Department, Epicentre, Paris, France
Stephanie Bjerrum	PhD	Department of Clinical Research, Infectious Diseases, University of Southern Denmark, Odense, Denmark
Prof Isik S. Johansen	PhD	Research Unit for Infectious Diseases, Odense University Hospital, University of Southern Denmark, Odense, Denmark

Prof Swe Swe Thit	FRCP	Department of Medicine, University of Medicine 2, Yangon, Yangon Division, Myanmar
Prof Mar Mar Kyi	FRCP	Department of Medicine, University of Medicine 2, Yangon, Yangon Division, Myanmar
Josh Hanson	FRACP	The Kirby Institute, University of New South Wales, Sydney, NSW, Australia
David A Barr	PhD	Institute of Infection and Global Health, University of Liverpool, Liverpool, UK
Prof Graeme Meintjes	PhD	Department of Medicine, University of Cape Town, Cape Town, South Africa / Wellcome Centre for Infectious Diseases Research in Africa (CIDRI-Africa), Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa
Prof Gary Maartens	MMed	Department of Medicine, University of Cape Town, Cape Town, South Africa / Wellcome Centre for Infectious Diseases Research in Africa (CIDRI-Africa), Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

Manuscript word count: 3394

Abstract word count: 351

Figure count: 2

Table count: 4

Supplementary files: 1

Number of references: 46

ABSTRACT

Background: WHO recommends urine lateral-flow lipoarabinomannan (LF-LAM) testing with AlereLAM in HIV-positive inpatients only if screening criteria are met. We assessed the performance of WHO screening criteria and alternative screening tests/strategies to guide LF-LAM testing and compared diagnostic accuracy of the WHO AlereLAM algorithm (WHO screening criteria → AlereLAM) with AlereLAM and FujiLAM (a novel LF-LAM test).

Methods: We searched MEDLINE, Embase, and Cochrane Library from Jan 1, 2011 to March 1, 2020 for studies among adult/adolescent HIV-positive inpatients regardless of tuberculosis signs and symptoms. The reference standards were 1) AlereLAM or FujiLAM for screening tests/strategies and 2) culture or Xpert for AlereLAM/FujiLAM. We determined proportion of inpatients eligible for AlereLAM using WHO screening criteria; assessed accuracy of WHO criteria and alternative screening tests/strategies to guide LF-LAM testing; compared accuracy of WHO AlereLAM algorithm with AlereLAM/FujiLAM in all; and determined diagnostic yield of AlereLAM, FujiLAM, and Xpert MTB/RIF (Xpert). We estimated pooled proportions with a random-effects model, assessed diagnostic accuracy using random-effects bivariate models, and assessed diagnostic yield descriptively.

Findings: We obtained data from all 5 identified studies (n=3,504). The pooled proportion of inpatients eligible for AlereLAM using WHO criteria was 93% (95%CI 91, 95). Among screening tests/strategies to guide LF-LAM testing, WHO criteria, C-reactive protein (≥5 mg/L), and CD4 count (<200 cells/μL) had high sensitivities but low specificities; cough (≥2 weeks), haemoglobin (<8 g/dL), body mass index (<18.5 kg/m^2), lymphadenopathy, and WHO-defined danger signs had higher specificities but suboptimal sensitivities. AlereLAM in all had the same sensitivity (62%) and specificity (88%) as WHO AlereLAM algorithm. Sensitivity of FujiLAM and AlereLAM was 69% and 48%, while specificity was 48% and 96%, respectively. Diagnostic yield of sputum Xpert was 29-41%, AlereLAM was 39-76%, and urine Xpert was 35-62%. In one study, FujiLAM diagnosed 80% of tuberculosis cases (vs 39% for AlereLAM), and sputum Xpert combined with AlereLAM, urine Xpert, or FujiLAM diagnosed 69%, 81%, and 92% of all cases, respectively.

Interpretation: WHO criteria and alternative screening tests/strategies have limited utility in guiding LF-LAM testing, suggesting that AlereLAM testing in all HIV-positive medical inpatients be implemented. Routine FujiLAM may improve tuberculosis diagnosis.

Funding: None.

Panel: Research in context

Evidence before this study

WHO recommends that rapid molecular diagnostic testing (e.g. Xpert MTB/RIF [Xpert]) be done in all HIV-positive inpatients irrespective of tuberculosis signs and symptoms. However, urine lateral-flow lipoarabinomannan (LF-LAM) testing with AlereLAM is recommended only in those with a positive WHO four-symptom screen (W4SS), CD4 count ≤ 200 cells/μL, WHO stage 3 or 4, or positive WHO-defined danger sign. The performance of

WHO screening criteria and alternative screening tests/strategies to guide LF-LAM testing in HIV-positive inpatients is unknown. A comparison of the diagnostic accuracy of the WHO-recommended AlereLAM algorithm (WHO screening criteria followed by AlereLAM) with AlereLAM (no initial screening criteria) or FujiLAM (a novel LF-LAM test) for all HIV-positive inpatients has also not been conducted in this population.

We searched MEDLINE, Embase, and Cochrane Library from Jan 1, 2011 to March 1, 2020 with search terms related to "human immunodeficiency virus", "tuberculosis", "screening", "algorithm", "sensitivity", and "specificity". We performed a systematic review and individual participant data meta-analysis among HIV-positive inpatients (irrespective of tuberculosis signs and symptoms) to assess the performance of WHO screening criteria and alternative tuberculosis screening tests/strategies to guide LF-LAM testing; compare the performance of the WHO AlereLAM algorithm with AlereLAM or FujiLAM testing in all HIV-positive inpatients for diagnosis of tuberculosis; and determine the diagnostic yield of sputum or urine Xpert, AlereLAM, and FujiLAM.

Added value of this study

We analysed data from 3,504 HIV-positive inpatients. We found that 93% of all inpatients were eligible for AlereLAM using WHO screening criteria. The screening tests/strategies we explored to guide LF-LAM testing performed poorly. WHO screening criteria, W4SS, CRP (≥5mg/L), and CD4 count (<200 cells/μL) had high sensitivities, but low specificities; cough (≥2 weeks), haemoglobin (<8 g/dL), body mass index (<18.5 kg/m^2), lymphadenopathy, and WHO-defined danger signs had higher specificities but suboptimal sensitivities. AlereLAM for all had the same sensitivity as the WHO AlereLAM algorithm (62%). The sensitivity of FujiLAM for all was 21 percentage points higher than AlereLAM for all but specificity was 8 percentage points lower, although data were limited. AlereLAM had similar or higher diagnostic yield than sputum Xpert, likely because almost all inpatients could produce a urine sample for AlereLAM testing but a high proportion were unable to produce sputum for Xpert testing. In one study, FujiLAM diagnosed double the number of tuberculosis cases compared with AlereLAM. Sputum Xpert combined with AlereLAM diagnosed only 69% of all cases but sputum Xpert combined with urine Xpert or FujiLAM diagnosed 81% and 92% of all cases, respectively.

Implications of all the available evidence

Our study has policy implications for the diagnosis of tuberculosis in HIV-positive inpatients. Current WHO criteria and alternative screening tests/strategies to guide LF-LAM testing have suboptimal accuracy and complicate the tuberculosis diagnostic algorithm, potentially serving as a barrier to the widespread use of AlereLAM. Thus, AlereLAM testing in all HIV-positive medical inpatients should be implemented alongside routine Xpert in settings with high tuberculosis prevalence. Routine FujiLAM may substantially improve the rapid diagnosis of tuberculosis in this population, if validation studies confirm our findings

INTRODUCTION

Tuberculosis is the leading cause of hospitalization among people living with HIV (PLHIV) and is responsible for nearly 40% of in-hospital deaths.[1,2] Almost 50% of tuberculosis is undiagnosed at autopsy in PLHIV.[2] The diagnosis of tuberculosis in HIV-positive inpatients is challenging: inpatients typically have advanced immunodeficiency with disseminated or extrapulmonary disease, often produce paucibacillary specimens, and are frequently unable to produce sputum specimens.[3-6]

Urine lateral-flow lipoarabinomannan (LF-LAM) tests may address some of these challenges. They are rapid, inexpensive, non-sputum based, and available at point-of-care. Currently, the only LF-LAM test that WHO recommends is the Alere Determine TB-LAM (AlereLAM).[7] Although AlereLAM has only moderate sensitivity,[8] it reduced mortality in inpatients in randomized trials.[4,9] The novel Fujifilm SILVAMP TB-LAM (FujiLAM) test is more sensitive than AlereLAM in inpatients.[10,11] The 2021 WHO tuberculosis screening and diagnostic algorithm among HIV-positive inpatients recommends rapid molecular diagnostic testing (e.g., Xpert MTB/RIF [Xpert]) in all medical inpatients where tuberculosis prevalence is >10%.[12,13] However, AlereLAM is only recommended in those with a positive WHO four-symptom screen (W4SS), CD4 count ≤ 200 cells/μL, WHO stage 3 or 4, or positive WHO-defined danger sign.[8,12]

The WHO screening criteria to guide AlereLAM testing may be challenging to implement in busy inpatient settings[14] and its diagnostic accuracy is unknown. The W4SS, which was developed among ambulatory PLHIV,[15] has low specificity for diagnosis of tuberculosis in inpatients.[16] CD4 cell count may also have low specificity, since inpatients typically have advanced immunodeficiency. It is also often not rapidly available. Furthermore, the diagnostic accuracy of WHO-defined danger signs was not assessed in the review that led to the recommendation.[12] The diagnostic accuracy of alternative screening tests/strategies to guide LF-LAM testing is also unknown. LF-LAM testing in all HIV-positive inpatients may be more appropriate than testing only if screening criteria are met.

We assessed the performance of WHO screening criteria and other screening tests/strategies to guide LF-LAM testing among HIV-positive inpatients (irrespective of tuberculosis signs and symptoms) using an individual participant data (IPD) meta-analysis. Our primary objectives were to 1) determine the proportion of inpatients eligible for AlereLAM using the WHO AlereLAM algorithm (i.e., WHO screening criteria followed by AlereLAM) and 2) assess the diagnostic accuracy of WHO screening criteria and alternative tuberculosis screening tests/strategies to guide LF-LAM testing. Our secondary objectives were to 1) compare the diagnostic accuracy of the WHO AlereLAM algorithm with AlereLAM or FujiLAM testing in all inpatients for tuberculosis; 2) determine the diagnostic yield of rapid tuberculosis diagnostic testing (i.e., proportion of total tuberculosis cases with a positive sputum or urine Xpert, AlereLAM, or FujiLAM); and 3) evaluate the diagnostic accuracy of the WHO-defined danger signs for tuberculosis.

METHODS

Our findings are reported in accordance with PRISMA-IPD and PRISMA-DTA statements.[17,18] Two authors (AD, YH) independently participated in each step of the systematic review: study selection, data extraction, and study

quality assessment. Disagreements between authors were resolved by discussion. We used similar methods to our recent systematic review that contributed to the 2021 WHO tuberculosis screening guidelines;[13,16,19] LF-LAM analyses were not pre-specified in our protocol. Our initial systematic review was registered on PROSPERO (CRD42020155895).

Literature Search

WHO conducted a systematic review of the accuracy of W4SS for tuberculosis screening in PLHIV and searched PubMed (MEDLINE), Embase, Cochrane Library, and conference abstracts from 1 January 2011 to 12 March 2018 (Table S1).[20] The search was limited to studies conducted after 2011, since WHO only developed the W4SS in that year. We retrieved all included studies from this systematic review and reassessed all full texts to identify any further eligible studies. To perform an updated search, we applied the same search strategy from 12 March 2018 to 1 March 2020. Finally, we searched reference lists of related reviews and included articles and contacted experts to inquire about any additional published or unpublished studies.

Study Selection

We reviewed titles and abstracts from the search and, if potentially eligible, full texts of articles. We included primary datasets that 1) were observational studies (cross-sectional or cohort studies) or randomised trials; 2) included adult or adolescent HIV-positive inpatients regardless of tuberculosis signs and symptoms; 3) collected data on W4SS alone (and in combination with CD4 count, WHO stage, or WHO-defined danger signs); and 4) evaluated AlereLAM and/or FujiLAM. We excluded studies that were case-control as they are prone to bias,[21] had only symptomatic HIV-positive inpatients as an inclusion criterion, or enrolled inpatients who were on tuberculosis treatment or were already diagnosed with active tuberculosis.

The target condition was active tuberculosis. To assess diagnostic accuracy of screening tests/strategies to guide LF-LAM testing, the separate reference standards were AlereLAM and FujiLAM because these analyses only concerned the assessment of screening tests in the context of LF-LAM positive tuberculosis (as opposed to any microbiologically confirmed tuberculosis), as recommended by WHO.[22] To compare diagnostic accuracy of the WHO AlereLAM algorithm with AlereLAM or FujiLAM for all, the reference standard was culture or Xpert of sputum and/or other specimens.

The tuberculosis screening tests/strategies we examined were the W4SS; CD4 count ≤200 cells/μL; W4SS or CD4 count ≤200 cells/μL (either positive); WHO-defined danger signs; CRP; chest X-ray; haemoglobin; BMI; lymphadenopathy; and cough ≥2 weeks. We primarily used W4SS or CD4 count ≤ 200 cells/μL (either positive) as WHO eligibility criteria for AlereLAM testing because few studies included WHO stage and WHO-defined danger signs. Finally, the systematically performed tuberculosis LF-LAM diagnostic tests we examined were AlereLAM and FujiLAM.

Data Extraction, Study Quality, and IPD Synthesis

Study-level variables extracted were first author, publication year, study period, country, setting, exclusion criteria, study design, type of participants, and method of tuberculosis diagnosis. To assess study quality for proportion meta-analyses, we modified a tool used in systematic reviews of prevalence.[23] To assess study quality for diagnostic test accuracy, we used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[24]

We emailed authors of eligible datasets inviting them to contribute data. Table S2 shows the IPD collected. After standardizing IPD, we synthesized a single dataset with individual participant and study-level data. Study participants <10 years of age were excluded. Contaminated cultures were considered negative. We ensured IPD integrity by comparing information against study publications and performing recommended checks.[25,26] Discrepancies were resolved by contacting the corresponding author.

Statistical Analyses

We analysed IPD using a two-stage approach. First, we analysed each study separately to obtain aggregate data. The aggregate data for each study were tuberculosis prevalence, proportion of inpatients eligible for AlereLAM using WHO criteria, and measures of diagnostic accuracy (i.e., sensitivity and specificity). For assessment of proportion of inpatients eligible for AlereLAM, we evaluated the W4SS in combination with CD4 count, WHO stage, or WHO-defined danger signs. Second, we combined aggregate data using an appropriate meta-analysis model. We used a generalized linear mixed model with logit transformation to pool tuberculosis prevalence and proportion of inpatients eligible for AlereLAM.[27] We assessed heterogeneity with Cochran's Q test and I^2 statistic.[28] We used a bivariate generalized linear mixed model to pool sensitivities and specificities.[29] If there were <4 studies or the model did not converge, we used simpler models that assumed no correlation between measures of sensitivity and specificity.[30] We computed binomial 95% CIs by summing the numbers with disease (or no disease) across studies if all studies had 100% sensitivity or specificity.[31] We used summary receiver-operating characteristic curves to jointly illustrate absolute pooled sensitivity and specificity.[32] We compared the accuracy of 2 tests by using indirect comparisons (which includes all studies that evaluated ≥1 of the relevant tests). We also performed direct comparisons (which includes all studies that evaluated all relevant tests). For direct comparisons, we used a bivariate meta-regression with test-type as a covariate.

We calculated diagnostic yield of sputum/urine Xpert, AlereLAM, or FujiLAM in studies that included participants unable to produce sputum samples. We used culture, Xpert, or AlereLAM as the denominator, because a positive result on either of these tests is considered sufficient evidence to treat tuberculosis. Finally, we performed mixed-effect logistic regression analysis with random intercept by cohort to determine whether WHO-defined danger signs (individually and combined) were associated with tuberculosis (defined as positive sputum Xpert or AlereLAM because of limited culture data). We calculated both unadjusted and adjusted odds ratios (ORs).

Since analyses were based on few studies, we were unable to investigate heterogeneity with meta-regression or assess for publication bias. We chose a p-value threshold of 0.05 to characterize statistically significant findings. All meta-analyses were performed using *lme4*, *altmeta*, *meta*, *metafor* and *mada* packages in R software version 3.6.1.

Role of the funding source

None.

RESULTS

Characteristics of primary datasets selected and prevalence of tuberculosis

We identified 5 eligible datasets (figure S1), and IPD was obtained for all 5 datasets (n=3,504).[4,33-37] Table S3 shows the characteristics of included studies. Four studies were conducted in sub-Saharan Africa. All studies included inpatients admitted to medical wards (one was an infectious disease ward). Studies systematically collected sputum for culture (n=3), sputum for Xpert (n=5), urine for Xpert (n=3), urine for AlereLAM (n=5), and urine for FujiLAM (n=2). We judged risk of bias for 5 studies that contributed to the meta-analysis of proportion of inpatients eligible for AlereLAM (Table S4). One study had inadequate response rate, while another study used an inappropriate sample frame. We judged risk of bias for 5 studies that contributed to the diagnostic meta-analysis of LF-LAM and screening tests/strategies (Table S5). For LF-LAM analyses, four studies did not collect extrapulmonary samples or samples for culture and were judged to have high risk of bias for reference test domain. Table S6 shows missing data by study. In 3 studies that included participants unable to produce sputum samples,[4,35,36] LF-LAM was missing for ≤3% of inpatients, but sputum Xpert was missing for 35-54% of participants.

Table 1 and Table S7 shows participant characteristics overall and by study, respectively. Most (57%) participants were women; 49% had a CD4 count ≤ 200 cells/μL. The pooled tuberculosis prevalence (using culture or Xpert as a reference standard) was 23% (95%CI 14, 35; n=543) among 3 studies that collected sputum for culture.

Proportion of inpatients eligible for AlereLAM testing according to WHO AlereLAM algorithm

The proportion with a positive W4SS or CD4 count <200 cells/μL was 93% (95%CI 91, 95; n=3,477) (Table 2 and Figure S2). The pooled proportions of other screening combinations to determine eligibility for AlereLAM testing ranged from 89% to 93%. The pooled proportion of inpatients with a WHO-defined danger sign was 26% (95%CI 19, 35; n=2,961). The addition of any WHO-defined danger signs, WHO stage 3 or 4, and CD4 count <200 cells/μL to W4SS (i.e., either positive) increased eligibility for AlereLAM testing by only 1 (n=2,961), 4 (n=54), and 3 (n=3,477) percentage points, respectively (Table S8).

Diagnostic performance of tuberculosis screening tests/strategies

Figure 1 shows plots of sensitivity and specificity of each screening test/strategy for LF-LAM positive tuberculosis, while table 3 shows indirect comparisons. W4SS alone (or combined with CD4 count <200 cells/μL) and CRP had high sensitivities but low specificities. CD4 count <200 cells/μL had sensitivities between 89-90% and specificities between 37-46%. Cough (≥2 weeks), haemoglobin (<8g/dL), lymphadenopathy, and WHO-defined danger signs had low sensitivities (14-58%) but high specificities (70-89%).

Figure S3 shows forest plots and Figure S4 shows summary receiver operating characteristics curves. The point estimates for the specificities of WHO screening criteria were ≤3% in each individual study (Figure S3). Figure S5

shows the trade-off between AlereLAM positive tuberculosis cases missed and number of AlereLAM tests performed for each individual screening test.

The sensitivity of the WHO AlereLAM algorithm (W4SS or CD4<200 cells/μL → AlereLAM) was 62% (95%CI 47, 75) and specificity was 89% (95%CI 67, 97; n=2,036) (Table 4 and Figure 2); the sensitivity and specificity of AlereLAM for all was similar. Two studies compared FujiLAM with AlereLAM. Sensitivity of FujiLAM and AlereLAM was 69% (95%CI 62, 76) and 48% (95%CI 29, 69), respectively; specificity of FujiLAM and AlereLAM was 88% (95%CI 79, 93) and 96% (95%CI 82, 99), respectively.

Diagnostic yield of tuberculosis from different diagnostic tests and sample types

Table S9 shows diagnostic yield using culture, Xpert, or AlereLAM as the denominator among 3 studies that included participants who were unable to produce sputum samples. Sputum Xpert diagnosed only 29-41% of all tuberculosis cases, as 35-54% had missing sputum Xpert results. In all studies, AlereLAM had similar or higher yield than sputum Xpert. In 2 studies that collected sputum and non-sputum samples for Xpert and/or culture, AlereLAM and urine Xpert diagnosed 39-76% and 35-62% of all cases, respectively. In 1 study that collected sputum and non-sputum samples for Xpert and culture and urine for AlereLAM,[36] FujiLAM diagnosed 80% of cases, while urine Xpert and AlereLAM diagnosed 62% and 39% of cases, respectively. In the same study, sputum Xpert combined with AlereLAM diagnosed only 69% of all cases, but sputum Xpert combined with urine Xpert or FujiLAM diagnosed 81% and 92% of all cases, respectively. Across all studies, AlereLAM was positive in 5.1% (8/158) of inpatients who did not meet WHO criteria for AlereLAM testing, and those with a positive AlereLAM test had negative Xpert or culture results. AlereLAM and FujiLAM were positive in 8.5% (70/819) and 28% (61/218) of inpatients with no available sputum Xpert result, respectively.

Association of WHO-defined danger signs with tuberculosis

In univariable mixed-effects logistic regression analysis, any WHO-defined danger sign was associated with increased risk of tuberculosis (OR 2.62 95%CI 2.01, 3.43) (Table S10). In univariable and multivariable mixed-effects logistic regression, individual danger signs other than respiratory rate >30 breaths/min were associated with tuberculosis. Multivariable adjusted odds ratios were smaller compared with univariable estimates, reflecting a positive correlation between individual danger signs.

DISCUSSION

In this IPD meta-analysis, almost all HIV-positive inpatients were eligible for AlereLAM testing using WHO screening criteria, which had very low specificity. We found that potential screening tests/strategies to guide AlereLAM or FujiLAM testing had either suboptimal sensitivities or specificities. The WHO-recommended AlereLAM inpatient algorithm had a sensitivity of 62%; AlereLAM in all inpatients had identical sensitivity. In 2 studies, sensitivity of FujiLAM was 21 percentage points higher than AlereLAM and specificity was 8 percentage points lower, although confidence intervals overlapped. AlereLAM had similar or higher yield than sputum Xpert, as urine samples were obtained from almost all inpatients, but many were unable to produce sputum. In 1 study,

FujiLAM diagnosed twice as many tuberculosis cases than AlereLAM. Sputum Xpert combined with FujiLAM diagnosed 92% of cases versus 69% when combined with AlereLAM. Our findings suggest that implementation of AlereLAM testing in all HIV-positive medical inpatients in high burden settings be considered alongside routine Xpert testing. FujiLAM testing in all HIV-positive inpatients could substantially improve detection of tuberculosis.

We found that potential screening tests/strategies to guide LF-LAM testing had suboptimal sensitivity and/or specificity. The W4SS and CRP had high sensitivities, but much lower specificities compared with outpatient settings.[19] Conversely, several other tests (e.g., WHO-defined danger signs and low haemoglobin) had moderate to high specificities but low sensitivities. These tests might be a proxy for advanced immunodeficiency and a higher bacillary burden. CD4 count appeared to provide the best trade-off between sensitivity and specificity. In 2019, WHO updated its 2015 recommendations on AlereLAM testing, increasing the CD4 count threshold for testing HIV-positive inpatients from <100 cells/μL to <200 cells/μL.[38,39] However, if eligibility for AlereLAM testing is based solely on the new cut-off, 10% of AlereLAM positive tuberculosis cases would be missed. CD4 count meets WHO minimal sensitivity requirements for a screening/triage test (i.e., 90% sensitivity), but it does not meet WHO optimal requirements (i.e., 95% sensitivity), which may be preferred in inpatient settings.[22]

Our diagnostic yield findings highlight the utility of urine-based tuberculosis diagnostics in HIV-positive inpatients. Urine Xpert or AlereLAM often had higher yield than sputum Xpert, as urine was readily available for testing. However, based on limited data, it is unclear whether urine Xpert or AlereLAM provides higher yield; urine Xpert had higher yield compared with AlereLAM in one included study,[40] but the opposite was true in another included study.[4] In a recent study, sensitivity of urine Xpert Ultra was double that of AlereLAM (33% vs 16%).[40]AlereLAM is less costly than urine Xpert and provides a more rapid diagnosis, since an Xpert result may take several days in the real world.[41] However, urine Xpert provides rifampicin susceptibility. There is a need for implementation science and health economics research to make appropriate recommendations for different settings.

AlereLAM is a rapid, inexpensive point of care test, which would have a number of benefits if testing was implemented in all HIV-positive inpatients in real world settings. First, since most inpatients already meet WHO criteria for testing, routine AlereLAM testing would reduce complexity and accelerate clinical decision making in busy inpatient settings. For example, CD4 cell count is one of the WHO criteria for AlereLAM testing but may not be immediately available to treating clinicians. Second, routine AlereLAM testing (in addition to routine sputum Xpert) was cost-effective in the STAMP trial.[42] Third, AlereLAM was positive in 5% of HIV-positive inpatients who did not meet WHO criteria for AlereLAM testing. Fourth, two randomised trials have demonstrated a reduction in all-cause mortality among HIV-positive inpatients with the use of AlereLAM in addition to routine diagnostics.[4,9] One trial included HIV-positive inpatients irrespective of tuberculosis signs and symptoms,[4] while the other included inpatients with a positive W4SS (which we found was present in >90% of HIV-positive inpatients).[9] Despite these findings, a recent survey of 24 high tuberculosis/HIV burden countries revealed that only 4 (17%) were using AlereLAM in all hospitals.[43] Combined use of sputum Xpert and AlereLAM has also been shown to improve diagnostic yield over either test alone in tuberculosis blood stream infection, which predicts mortality.[44]

AlereLAM and Xpert in all HIV-positive inpatients would increase diagnostic yield. But a negative result on both tests does not rule out tuberculosis. FujiLAM may substantially bridge the diagnostic gap. We found that sputum Xpert when combined with FujiLAM diagnosed 92% of tuberculosis cases versus 69% when combined with AlereLAM. A strategy that incorporates FujiLAM takes advantage of FujiLAM's higher sensitivity and the immediate availability of urine. WHO-defined minimum thresholds for a rapid biomarker-based non-sputum-based test are 65% sensitivity and 98% specificity.[22] FujiLAM met the sensitivity threshold, but not the specificity threshold. However, the reduced specificity could be a result of an imperfect microbiological reference standard, since FujiLAM detects lower concentrations of LAM.[45,46] Nontuberculous mycobacteria could also reduce specificity but were found in only 4% of participants with a false-positive FujiLAM test.[10] Differences in FujiLAM accuracy may also be because studies we included used biobanked samples for testing. However, biobanked samples produce similar results to fresh samples.[11]

We found that any WHO-defined danger sign was associated with increased risk of tuberculosis. Our finding that all danger signs other than respiratory rate were associated with tuberculosis risk is consistent with that of a study that enrolled HIV-positive inpatients with ≥1 WHO-defined danger sign.[47] WHO-defined danger signs likely have limited utility in determining hospital admission, as we found that 74% of inpatients had no danger signs.

Our study has limitations. First, studies had high tuberculosis prevalence and only one study was conducted outside sub-Saharan Africa, limiting generalizability. Hoverer, sub-Saharan Africa has a disproportionate burden of HIV-associated tuberculosis. Second, some tests had wide 95% confidence intervals because of heterogenous or limited data. Third, only 2 studies evaluated FujiLAM and no study evaluated Xpert Ultra. Fourth, some studies excluded participants unable to produce sputum and/or did not collect extrapulmonary samples for microbiological testing. Thus, the reference standard in these studies may be biased because inpatients often produce paucibacillary sputum samples or present with extrapulmonary/disseminated tuberculosis. However, for screening tests, we used a reference standard of LF-LAM, which correctly classifies LF-LAM positive tuberculosis. WHO recommends that screening/triage tests be assessed against confirmatory tests that follow.[22] Furthermore, diagnostic yield analyses and estimates of proportion of inpatients eligible for AlereLAM did not require a reference standard. Therefore, it is unlikely that this limitation would alter our conclusions. Fifth, the small number of included studies precluded exploration of heterogeneity. Finally, we used W4SS or CD4 count ≤ 200 cells/μL as WHO eligibility criteria for AlereLAM given limited data on WHO-defined danger signs and WHO stage.

In conclusion, our findings suggest that AlereLAM testing in all HIV-positive medical inpatients in high burden settings be implemented alongside routine molecular diagnostic testing (e.g., Xpert). WHO criteria and other potential screening tests/strategies to guide AlereLAM testing have suboptimal diagnostic accuracy and complicate the tuberculosis diagnostic algorithm, potentially serving as a barrier to LF-LAM's widespread use. Xpert and AlereLAM testing in all HIV-positive inpatients would improve diagnostic yield, although a negative result on both tests does not rule out tuberculosis. Routine FujiLAM may substantially improve the rapid diagnosis of tuberculosis in this population if validation studies confirm our findings.

CONTRIBUTORS

AD, YH, APK, ADK, MXR, DAB, GrM & GaM designed the study and protocol and interpreted the results. GaM supervised the study. AD & YH did the systematic review. ADK, TB, CMD, AG-W, KF, RW, HH, SCMR, CH, DW, SB, ISJ, SST, MMK, & JH contributed data to the meta-analysis. AD analysed the data with assistance from APK, DAB & YH. AD and GaM wrote the first draft of the manuscript, which was revised based on comments from co-authors. AD, DAB, and YH accessed and verified the data. All authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

TB reports a patent-pending (WO/2019/186486) in the field of lipoarabinomannan detection, as an inventor, but without ownership/commercial rights.

ACKNOWLEDGEMENTS

AD received training in research that was supported by the Fogarty International Center of the National Institutes of Health under Award Number D43 TW010559 and National Research Foundation. GrM and GaM were supported by core funding from the Wellcome Centre for Infectious Diseases Research in Africa (203135/Z/16/Z). GrM was supported by the Wellcome Trust (214321/Z/18/Z), and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa (Grant No 64787). We thank Mrs Joy Oliver (Cochrane South Africa and South African Medical Research Council) for assisting to conduct searches for the review. We would also like to thank FIND, the global alliance for diagnostics, for agreeing to share FujiLAM data. This research was funded, in part, by the Wellcome Trust. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

DISCLAIMER:

None.

REFERENCES

1. Ford N, Matteelli A, Shubber Z, et al. TB as a cause of hospitalization and in-hospital mortality among people living with HIV worldwide: a systematic review and meta-analysis. *J Int AIDS Soc* 2016; **19**(1): 20714.
2. Gupta RK, Lucas SB, Fielding KL, Lawn SD. Prevalence of tuberculosis in post-mortem studies of HIV-infected adults and children in resource-limited settings: a systematic review and meta-analysis. *AIDS (London, England)* 2015; **29**(15): 1987-2002.
3. Swaminathan S, Padmapriyadarsini C, Narendran G. HIV-associated tuberculosis: clinical update. *Clin Infect Dis* 2010; **50**(10): 1377-86.
4. Gupta-Wright A, Corbett EL, van Oosterhout JJ, et al. Rapid urine-based screening for tuberculosis in HIV-positive patients admitted to hospital in Africa (STAMP): a pragmatic, multicentre, parallel-group, double-blind, randomised controlled trial. *Lancet (London, England)* 2018; **392**(10144): 292-301.
5. Huerga H, Ferlazzo G, Bevilacqua P, et al. Incremental Yield of Including Determine-TB LAM Assay in Diagnostic Algorithms for Hospitalized and Ambulatory HIV-Positive Patients in Kenya. *PLoS One* 2017; **12**(1): e0170976.
6. Lawn SD, Kerkhoff AD, Burton R, et al. Diagnostic accuracy, incremental yield and prognostic value of Determine TB-LAM for routine diagnostic testing for tuberculosis in HIV-infected patients requiring acute hospital admission in South Africa: a prospective cohort. *BMC Med* 2017; **15**(1): 67.
7. World Health Organization. WHO operational handbook on tuberculosis. Module 3: diagnosis - rapid diagnostics for tuberculosis detection, 2021 update. Geneva: World Health Organization; 2021.
8. Bjerrum S, Schiller I, Dendukuri N, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV. *Cochrane Database Syst Rev* 2019; **10**: CD011420.
9. Peter JG, Zijenah LS, Chanda D, et al. Effect on mortality of point-of-care, urine-based lipoarabinomannan testing to guide tuberculosis treatment initiation in HIV-positive hospital inpatients: a pragmatic, parallel-group, multicountry, open-label, randomised controlled trial. *Lancet (London, England)* 2016; **387**(10024): 1187-97.
10. Broger T, Nicol MP, Szekely R, et al. Diagnostic accuracy of a novel tuberculosis point-of-care urine lipoarabinomannan assay for people living with HIV: A meta-analysis of individual in- and outpatient data. *PLoS Med* 2020; **17**(5): e1003113.
11. Broger T, Muyoyeta M, Kerkhoff AD, Denkinger CM, Moreau E. Tuberculosis test results using fresh versus biobanked urine samples with FujiLAM. *Lancet Infect Dis* 2020; **20**(1): 22-3.
12. World Health Organization. Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV: policy update 2019; 2019.
13. World Health Organization. WHO operational handbook on tuberculosis: module 2: screening: systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021.
14. Mwaura M, Engel N. Constructing confidence: User perspectives on AlereLAM testing for tuberculosis. *Int J Infect Dis* 2021; **112**: 237-42.

15. Getahun H, Kittikraisak W, Heilig CM, et al. Development of a standardized screening rule for tuberculosis in people living with HIV in resource-constrained settings: individual participant data meta-analysis of observational studies. *PLoS Med* 2011; **8**(1): e1000391.
16. Dhana A, Hamada Y, Kengne AP, Kerkhoff AD, Rangaka MX. Tuberculosis screening among HIV-positive inpatients: a systematic review and individual participant data meta-analysis. *Lancet HIV* 2021(In press).
17. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018; **319**(4): 388-96.
18. Stewart LA, Clarke M, Rovers M, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015; **313**(16): 1657-65.
19. Dhana A, Hamada Y, Kengne AP, et al. Tuberculosis screening among ambulatory people living with HIV: a systematic review and individual participant data meta-analysis. *Lancet Infect Dis* 2021.
20. Hamada Y, Lujan J, Schenkel K, Ford N, Getahun H. Sensitivity and specificity of WHO's recommended four-symptom screening rule for tuberculosis in people living with HIV: a systematic review and meta-analysis. *The lancet HIV* 2018; **5**(9): e515-e23.
21. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. *Acad Emerg Med* 2013; **20**(11): 1194-206.
22. World Health Organization. High priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting, 28-29 April 2014, Geneva, Switzerland. Geneva: World Health Organization, 2014.
23. Munn Z, Moola S, Riitano D, Lisy K. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. *Int J Health Policy Manag* 2014; **3**(3): 123-8.
24. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**(8): 529-36.
25. Higgins JP. Cochrane handbook for systematic reviews of interventions version 5.0. 1. The Cochrane Collaboration. *http://www cochrane-handbook org* 2008.
26. Tierney JF, Vale C, Riley R, et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. *PLoS Med* 2015; **12**(7): e1001855.
27. Lin L, Chu H. Meta-analysis of Proportions Using Generalized Linear Mixed Models. *Epidemiology* 2020; **31**(5): 713-7.
28. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**(7414): 557-60.
29. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006; **59**(12): 1331-2; author reply 2-3.
30. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2017; **26**(4): 1896-911.
31. Dinnes J, Deeks JJ, Chuchu N, et al. Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. *Cochrane Database Syst Rev* 2018; **12**: CD011902.

32. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20**(19): 2865-84.

33. Bjerrum S, Kenu E, Lartey M, et al. Diagnostic accuracy of the rapid urine lipoarabinomannan test for pulmonary tuberculosis among HIV-infected adults in Ghana-findings from the DETECT HIV-TB study. *BMC Infect Dis* 2015; **15**: 407.

34. Heidebrecht CL, Podewils LJ, Pym AS, Cohen T, Mthiyane T, Wilson D. Assessing the utility of Xpert((R)) MTB/RIF as a screening tool for patients admitted to medical wards in South Africa. *Scientific reports* 2016; **6**: 19391.

35. Huerga H, Mathabire Rucker SC, Bastard M, et al. Urine Lipoarabinomannan Testing for All HIV Patients Hospitalized in Medical Wards Identifies a Large Proportion of Patients With Tuberculosis at Risk of Death. *Open Forum Infect Dis* 2021; **8**(2): ofaa639.

36. Lawn SD, Kerkhoff AD, Burton R, et al. Rapid microbiological screening for tuberculosis in HIV-positive patients on the first day of acute hospital admission by systematic testing of urine samples using Xpert MTB/RIF: a prospective cohort in South Africa. *BMC Med* 2015; **13**: 192.

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

38. World Health O. The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV: policy guidance. Geneva: World Health Organization; 2015.

39. World Health Organization. Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV: policy update 2019. Geneva: World Health Organization; 2019.

40. Andama A, Jaganath D, Crowder R, et al. Accuracy and incremental yield of urine Xpert MTB/RIF Ultra versus Determine TB-LAM for diagnosis of pulmonary tuberculosis. *Diagn Microbiol Infect Dis* 2020; **96**(1): 114892.

41. Cohen GM, Drain PK, Noubary F, Cloete C, Bassett IV. Diagnostic delays and clinical decision making with centralized Xpert MTB/RIF testing in Durban, South Africa. *J Acquir Immune Defic Syndr* 2014; **67**(3): e88-93.

42. Reddy KP, Gupta-Wright A, Fielding KL, et al. Cost-effectiveness of urine-based tuberculosis screening in hospitalised patients with HIV in Africa: a microsimulation modelling study. *Lancet Glob Health* 2019; **7**(2): e200-e8.

43. Singhroy DN, MacLean E, Kohli M, et al. Adoption and uptake of the lateral flow urine LAM test in countries with high tuberculosis and HIV/AIDS burden: current landscape and barriers. *Gates Open Res* 2020; **4**: 24.

44. Barr DA, Lewis JM, Feasey N, et al. Mycobacterium tuberculosis bloodstream infection prevalence, diagnosis, and mortality risk in seriously ill adults with HIV: a systematic review and meta-analysis of individual patient data. *Lancet Infect Dis* 2020; **20**(6): 742-52.

45. Sigal GB, Pinter A, Lowary TL, et al. A Novel Sensitive Immunoassay Targeting the 5-Methylthio-d-Xylofuranose-Lipoarabinomannan Epitope Meets the WHO's Performance Target for Tuberculosis Diagnosis. *J Clin Microbiol* 2018; **56**(12).

46. Lawn SD, Kerkhoff AD, Nicol MP, Meintjes G. Underestimation of the True Specificity of the Urine Lipoarabinomannan Point-of-Care Diagnostic Assay for HIV-Associated Tuberculosis. *J Acquir Immune Defic Syndr* 2015; **69**(4): e144-6.

47. Griesel R, Stewart A, van der Plas H, et al. Optimizing Tuberculosis Diagnosis in Human Immunodeficiency Virus-Infected Inpatients Meeting the Criteria of Seriously Ill in the World Health Organization Algorithm. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018; **66**(9): 1419-26.

Tables and figures: Diagnostic accuracy of WHO screening criteria to guide lateral-flow lipoarabinomannan testing among HIV-positive inpatients

Table 1 - Summary of main characteristics for participants

Variable	Overall†	N‡
Participants	3504 (100)	
Demographics		
Age (years)	38 (31-46)	3504
Female	1992 (57)	3504
HIV history		
On ART	2363 (68)	3489
CD4 count (cells/μL)	205 (66-408)	3479
CD4 <=100 cells/μL	1118 (32)	3479
CD4 101 to 200 cells/μL	591 (17)	3479
CD4 >200 cells/μL	1770 (51)	3479
Clinical characteristics		
History of tuberculosis	856 (27)	3115
Positive W4SS*	3162 (90)	3502
Cough	1834 (52)	3500
Fever	1871 (54)	3496
Weight loss	2521 (72)	3495
Night sweats	1414 (40)	3499
Cough >= 2 weeks	731 (24)	3025
Lymphadenopathy	58 (11)	508
WHO-defined danger sign**	678 (23)	2961
WHO stage 3 or 4	96 (80)	120
Tuberculosis diagnostic tests		
AlereLAM positive	368 (17)	2191
FujiLAM positive	141 (30)	477
Total Xpert positive***	369 (13)	2827
Sputum Xpert +	270 (13)	2145
Non-sputum Xpert +	168 (10)	1736
Total culture positive***	126 (23)	543
Sputum culture +	75 (23)	332
Non-sputum culture +	70 (17)	420
Imaging and laboratory tests		
CXR (any abnormality)¶	130 (59)	220
BMI (kg/m2)	20 (18-24)	2966
CRP (mg/L)	75 (18-157)	400
CRP (>=10 mg/L)	334 (84)	400
Hb, Median (g/dL)	10 (8-12)	3481
Hb (<10 g/dL)	1574 (45)	3481

†Data are count (%) or median (25th-75th percentiles)

‡Participants with data available for variable

*W4SS defined as one or more of the following: current cough, fever, night sweats, or weight loss

**WHO-defined danger sign defined as one or more of the following: respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided

Variable	Overall†	N‡

***Sputum and/or non-sputum result

¶A positive chest x-ray was defined by the authors of the included studies

Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Table 2 - Random-effects meta-analysis of proportion of inpatients eligible for AlereLAM testing according to WHO AlereLAM algorithm*

					Heterogeneity	
Screening combination§¶	No studies	N	No screen positive	Proportion % (95% CI)†	I^2 (95% CI)	P-value
Positive W4SS or CD4 <=200 cells/µL	5	3,477	3,225	93 (91-95)	0 (0-71)	0.59
Positive W4SS or WHO-defined danger sign	2	2,961	2,691	91 (90-92)	47 (-)	0.17
Positive W4SS or WHO stage 3 or 4**	1	54	48	89 (77-95)	-	-
Positive W4SS or CD4 <=200 cells/µL or WHO-defined danger sign	2	2,945	2,735	93 (92-94)	66 (0-92)	0.09
Positive W4SS or CD4 <=200 cells/µL or WHO stage 3 or 4**	1	54	50	93 (82-97)	-	-

*According to WHO screening & diagnostic algorithm, AlereLAM testing for tuberculosis is advised if an inpatient has a positive W4SS (defined as one or more of the following: current cough, fever, night sweats, or weight loss), a CD4 count <= 200 cells/µL, is WHO stage 3 or 4, or has a WHO-defined danger sign (defined as one or more of the following: respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided)

§Combinations dependent on available variables. Proportion of inpatients with a positive W4SS was 90 (89-91) (5 studies;3502 participants), a CD4 count <= 200 cells/µL was 62 (49-74) (5 studies; 3479 participants), a WHO-defined danger sign was 26 (19-35) (2 studies; 2961 participants), and WHO stage 3 or 4** was 57 (44-70) (1 study; 54 participants).

¶Screening combination is either variable positive

†Calculated using meta-analysis of proportions

**One study by Bjerrum et al (2015) excluded from analysis as WHO stage 3 or 4 was part of inclusion criteria

Definition of abbreviations: W4SS = WHO four-symptom screen

Table 3 - Pooled sensitivity and specificity along with 95% CIs for each screening test/strategy for the detection of LF-LAM positive tuberculosis using reference standards of AlereLAM or FujiLAM

	AlereLAM†				FujiLAM			
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)
W4SS	5	2,189	94 (88-97)	10 (8-13)	2	475	99 (95-100)	3 (0-51)
CRP (>=10 mg/L)	1	392	99 (87-100)	19 (15-24)	1	391	94 (88-97)	21 (17-26)
CRP (>=8 mg/L)	1	392	99 (87-100)	17 (13-21)	1	391	95 (89-98)	18 (14-23)
CRP (>=5 mg/L)	1	392	99 (87-100)	11 (8-15)	1	391	97 (92-99)	12 (9-17)
CXR (abnormal)	2	220	60 (49-70)	41 (34-50)	-	-	-	-
Cough (any)	5	2,187	62 (57-67)	48 (42-54)	2	473	74 (48-90)	56 (50-61)
Cough (>=2 weeks)	3	1,736	36 (15-65)	79 (55-92)	2	472	33 (10-68)	84 (43-97)
Hb (<10 g/dL)	5	2,170	76 (66-84)	48 (38-59)	2	467	83 (75-88)	47 (30-63)
Hb (<8 g/dL)	5	2,170	48 (35-60)	71 (60-80)	2	467	58 (50-66)	70 (48-86)
BMI (<18.5 kg/m²)	4	1,664	54 (48-60)	62 (52-72)	1	58	55 (34-74)	61 (45-75)
Lymphadenopathy	3	503	14 (9-20)	89 (86-92)	1	67	8 (2-26)	85 (71-93)
WHO-defined danger sign*	2	1,657	44 (38-50)	77 (68-84)	-	-	-	-
CD4 count <=200 cells/µL	5	2,174	90 (55-99)	46 (34-58)	2	468	89 (83-93)	37 (19-59)
W4SS or CD4 count <=200 cells/µL¶§	5	1,990	100 (99-100)	0 (0-4)	2	464	100 (97-100)	1 (0-2)
W4SS or CRP (>=10 mg/L)¶	1	391	99 (87-100)	4 (2-7)	1	390	100 (93-100)	5 (3-8)
W4SS or CXR (abnormal)¶	2	220	93 (85-97)	3 (1-8)	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	1	391	98 (88-100)	19 (15-23)	1	390	96 (91-99)	22 (17-27)

†In one study by Gupta-Wright (2018), only the intervention arm was included since AlereLAM was unavailable for the standard of care arm.

*WHO-defined danger sign defined as one or more of the following: respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided

¶For parallel strategies, two screening tests are offered at the same time. For sequential strategies, a second screening test is offered only if the first screening test is positive

§Bivariate models did not converge; sensitivity estimates computed with binomial 95% CIs and specificity estimates from a univariate random-effects model

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

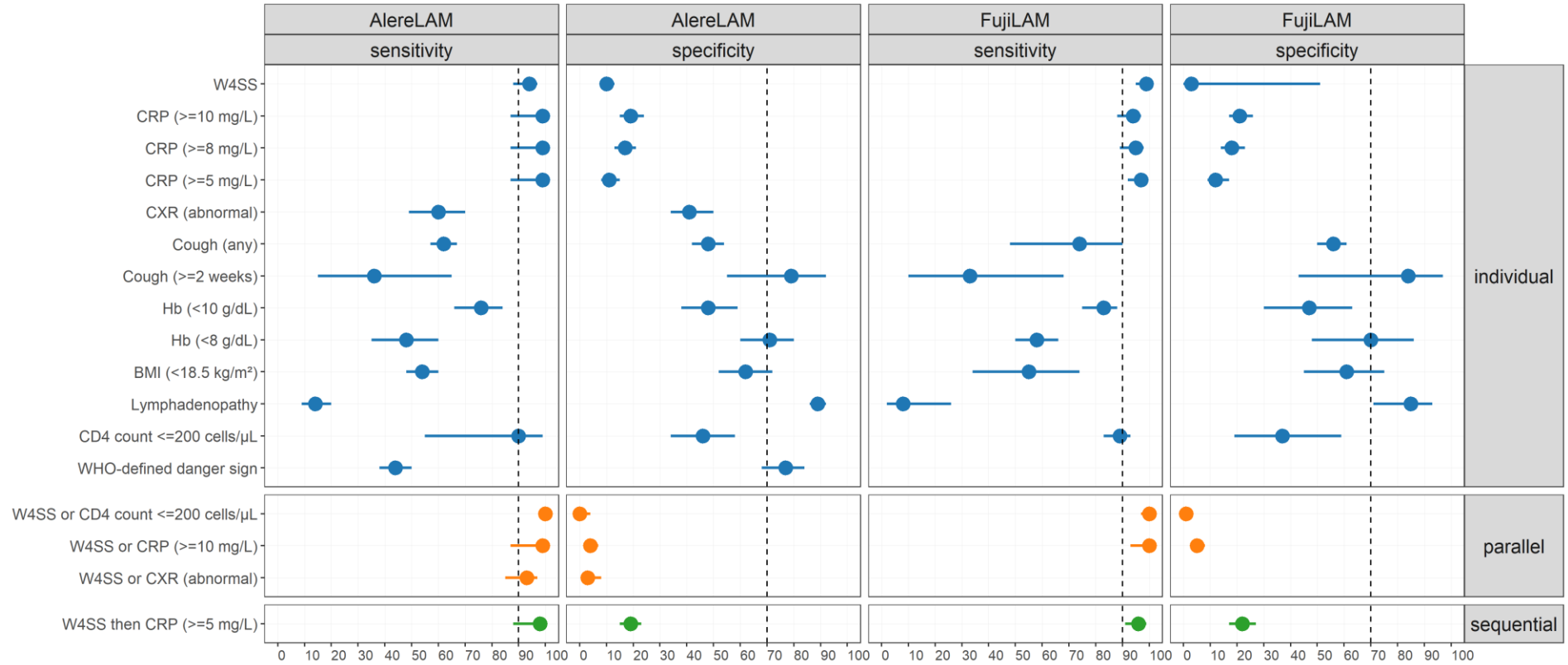
Table 4 - Pooled sensitivity and specificity along with 95% CIs of WHO AlereLAM algorithm, AlereLAM, and FujiLAM for the detection of tuberculosis§

Test	No studies	N	Sensitivity (95% CI)	Specificity (95% CI)
Indirect comparisons†				
WHO AlereLAM algorithm	5	2,036	62 (47-75)	89 (67-97)
AlereLAM alone	5	2,038	62 (47-75)	88 (64-97)
FujiLAM alone	2	477	73 (43-91)	88 (79-93)
Direct comparisons†				
WHO AlereLAM algorithm	2	475	48 (29-68)	96 (82-99)
AlereLAM alone	2	475	48 (29-68)	96 (82-99)
FujiLAM alone	2	475	69 (62-76)	88 (79-93)

§According to WHO screening & diagnostic algorithm, AlereLAM testing is advised if an inpatient has a positive WHO four-symptom screen (defined as one or more of the following: current cough, fever, night sweats, or weight loss), a CD4 count <= 200 cells/µL, is WHO stage 3 or 4, or has a WHO-defined danger sign (defined as one or more of the following: respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided)

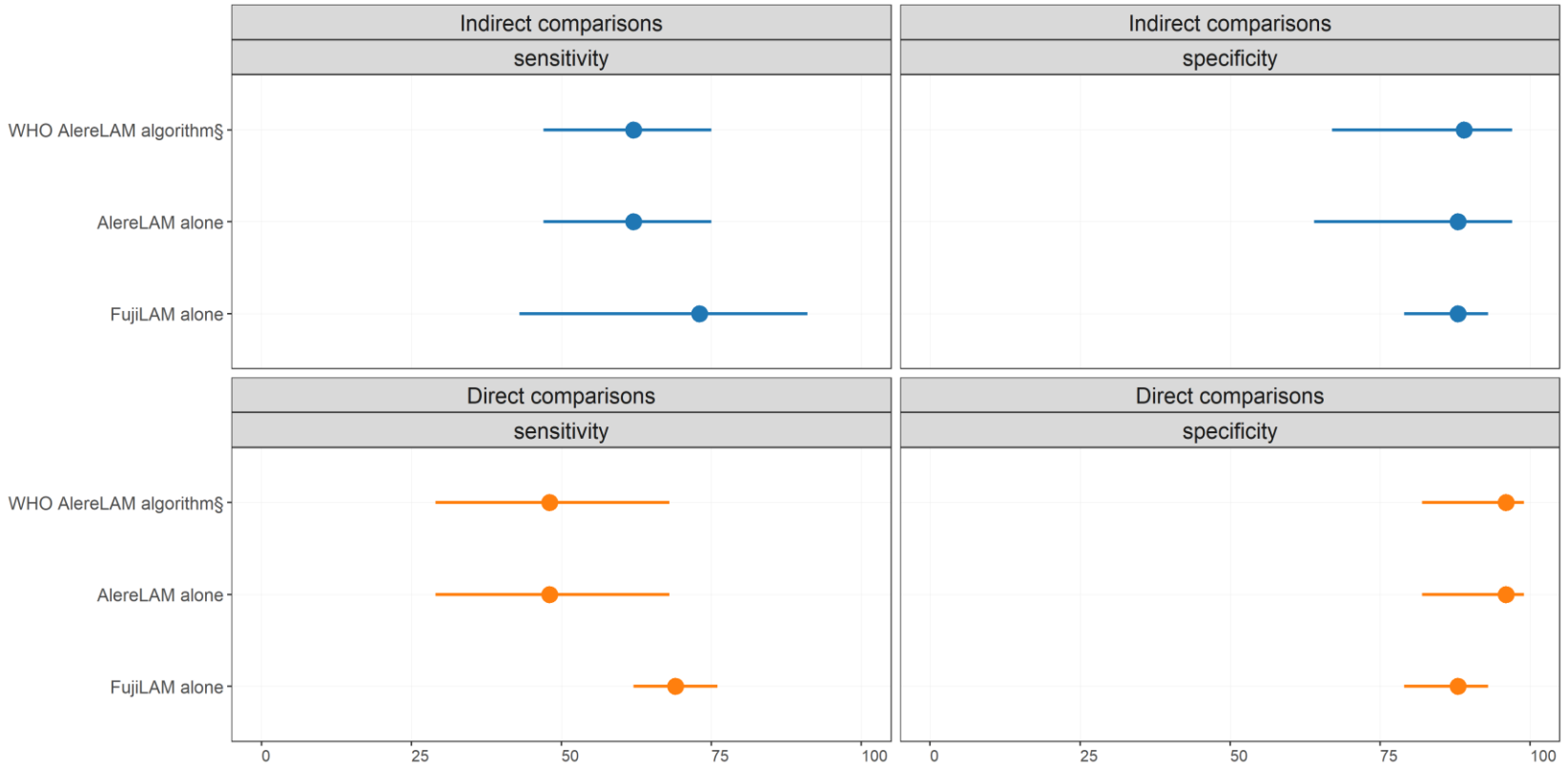
†Indirect comparisons include all studies that evaluated at least one of the relevant tests. Direct comparisons include all studies that evaluated all relevant tests

Figure 1 - Pooled sensitivity and specificity along with 95% CIs for each screening test/strategy for the detection of LF-LAM positive tuberculosis using reference standards of AlereLAM or FujiLAM†*



†Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity)
*For parallel strategies, two screening tests are offered at the same time. For sequential strategies, a second screening test is offered only if the first screening test is positive
Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Figure 2 - Pooled sensitivity and specificity along with 95% CIs for each LF-LAM strategy for the detection of tuberculosis



§AlereLAM testing is done if an inpatient has a positive WHO four-symptom screen (defined as one or more of the following: current cough, fever, night sweats, or weight loss) or a CD4 count <= 200 cells/µL

Supplementary tables and figures: Diagnostic accuracy of WHO screening criteria to guide lateral-flow lipoarabinomannan testing among HIV-positive inpatients

Table S1 - Search terms

Database	Search terms
Pubmed	
#1.	“HIV Infections” [MeSH] OR “HIV”[MeSH] OR “hiv”[tw] OR hiv infect*[tw] OR “human immunodeficiency virus”[tw] OR “human immunedeficiency virus”[tw] OR “human immuno-deficiency virus”[tw] OR “human immune-deficiency virus”[tw] OR ((human immun*) AND (“deficiency virus”[tw])) OR “acquired immunodeficiency syndrome”[tw] OR “acquired immunedeficiency syndrome”[tw] OR “acquired immuno-deficiency syndrome”[tw] OR “acquired immune-deficiency syndrome”[tw] OR ((acquired immun*) AND (“deficiency syndrome”[tw]))
#2.	"Tuberculosis"[Mesh] OR tuberculosis [TW] OR "Mycobacterium tuberculosis"[Mesh] OR TB [Ti]
#3	Screening* OR algorithm* OR “case finding” [TIAB] OR “case findings” [TIAB] OR sensitivit* OR specificit* OR predictor* OR "Sensitivity and Specificity"[MeSH Terms] OR "Tuberculosis/diagnosis"[Mesh] OR "Mass Screening"[Mesh:NoExp]
#4.	("animals"[MeSH Terms] NOT ("humans"[MeSH Terms] AND "animals"[MeSH Terms])) OR case reports[Publication Type]
#5	#1 AND #2 AND #3 NOT #4
	Limit: publication date from 2011/01/01
Embase	
#1	'human immunodeficiency virus infection'/exp OR 'human immunodeficiency virus'/exp OR 'hiv':ti,ab OR 'human immunodeficiency virus':ti,ab OR 'human immuno-deficiency virus':ti,ab OR 'human immunedeficiency virus':ti,ab OR 'human immune-deficiency virus':ti,ab OR 'acquired immune-deficiency syndrome':ti,ab OR 'acquired immunedeficiency syndrome':ti,ab OR 'acquired immunodeficiency syndrome':ti,ab OR 'acquired immuno-deficiency syndrome':ti,ab
#2	’tuberculosis’/exp OR 'tuberculosis':ab,ti OR 'TB':ti OR 'Mycobacterium tuberculosis'/exp
#3	‘Screen’:ti,ab OR ‘Screening’:ti,ab OR ‘algorithm’:ti,ab OR ‘case finding’:ti,ab OR ‘case findings’:ti,ab OR sensitivit*:ti,ab OR specificit*:ti,ab OR predictor*:ti,ab OR ‘sensitivity and specificity’/exp OR ‘case finding’/exp OR ‘Mass Screening’/exp OR ‘screening’/exp
#4	([animals]/lim NOT [humans]/lim)
#5	#1 AND #2 AND #3 NOT #4 AND [2011-]/py
Cochrane	
#1.	“HIV Infections” [MeSH] OR “HIV”[MeSH] OR hiv OR hiv infect* OR “human immunodeficiency virus” OR “human immunedeficiency virus”OR “human immuno-deficiency virus” OR “human immune-deficiency virus” OR ((human immun*) AND (“deficiency virus”)) OR “acquired immunodeficiency syndrome” OR “acquired immunedeficiency syndrome” OR “acquired immuno-deficiency syndrome” OR “acquired immune-deficiency syndrome” OR ((acquired immun*) AND (“deficiency syndrome”))
#2.	"Tuberculosis"[Mesh] OR tuberculosis OR "Mycobacterium tuberculosis"[Mesh]
#3	Screening* OR algorithm* OR “case finding” OR “case findings” OR sensitivit* OR specificit* OR predictor* OR "Sensitivity and Specificity"[MeSH Terms] OR "Tuberculosis/diagnosis"[Mesh] OR "Mass Screening"[Mesh:NoExp]
#4.	("animals"[MeSH Terms] NOT ("humans"[MeSH Terms] AND "animals"[MeSH Terms])) OR case reports[Mesh]
#5	#1 AND #2 AND #3 NOT #4
	Limit: publication year from 2011-

Table S2 - Variables sought

Variable	Description
country	country where the study took place, or if multisite, country individual patient was recruited from
clinical setting	from {inpatient, outpatient, other, NA}
age	patient's age in years
sex	patient's sex {female, male, NA}
hiv status	from {positive, negative, NA}
art status	from {on art, not on art, NA}
history of tuberculosis	from {history of tuberculosis, no history of tuberculosis, NA}
current smoking status	from {currently smoking, not currently smoking, NA}
pregnancy status	from {pregnant, not pregnant, NA}
tuberculosis treatment status	from {currently on tuberculosis treatment, not currently on tuberculosis treatment, NA}
current ipt status	from {yes, no, NA}
current cough	from {yes, no, NA}
cough (more than 2 weeks)	from {yes, no, NA}
fever	from {yes, no, NA}
weight loss	from {yes, no, NA}
night sweats	from {yes, no, NA}
w4ss	number of w4ss symptoms {0, 1, 2, 3, 4, NA}
respiratory rate >30 bpm	from {yes, no, NA}
body temperature >39°C	from {yes, no, NA}
heart rate >120 bpm	from {yes, no, NA}
unable to walk unaided	from {yes, no, NA}
who danger signs	number of who danger signs {0, 1, 2, 3, 4, NA}
who stage 3 or 4	from {yes, no, NA}
body mass index	numerical value (weight/height^2)
lymphadenopathy	from {yes, no, NA}
cd4 count	numerical value (in cells/μL)
c-reactive protein level	numerical value (in mg/L)
haemoglobin	numerical value (in g/dl)
chest x-ray suggestive of tuberculosis	from {yes, no, NA}
chest x-ray abnormal	from {yes, no, NA}
urine AlereLAM result	{positive, negative, NA}
urine FujiLAM result	{positive, negative, NA}
sputum xpert result	{positive, negative, NA}, indeterminate = negative
sputum culture result	{positive, negative, NA}, contaminated culture = negative

Variable	Description
non-sputum xpert result	{positive, negative, NA}, indeterminate = negative
non-sputum culture result	{positive, negative, NA}, contaminated culture = negative

Definition of abbreviations: ART = antiretroviral therapy, IPT = Isoniazid preventive therapy, W4SS = WHO four-symptom screen

Table S3 - Study-level characteristics

Author, year	Country	Study period	Study population	Study setting	Exclusion criteria	Sputum culture	Sputum Xpert	Liquid or solid culture	Non-sputum culture/Xpert	AlereLAM reference card (threshold)	FujiLAM
Bjerrum, 2015[1]	Ghana	2013-2014	ART-naïve inpatient PLHIV aged ≥18 years with WHO stage 3/4 or CD4 cell count ≤350 per µL or pregnant admitted to the Fevers Unit (infectious diseases ward)	1 hospital	On ATT for >2 days in 3 months before admission or unable to produce sputum or urine samples	1 spot and 1 early morning samples	1 or 2 samples	Both	Urine Xpert (biobanked)	Old (2)	Yes (biobanked)*
Gupta-Wright, 2018[2]	South Africa and Malawi	2015-2017	Inpatient PLHIV admitted to medical wards	2 hospitals	On ATT, treated for TB in previous 12 months, IPT in previous 6 months, admitted to hospital for >48 hours at time of screening	-	1 spot sample, induced if physician requested at 1 site	-	Urine Xpert in intervention group	New (1)	-
Huerga, 2021[3]	Malawi	2015-2017	Inpatient PLHIV aged ≥15 years admitted to medical wards	1 hospital	On ATT	-	1 spot sample	-	-	New (1)	-
Lawn, 2015[4]	South Africa	2012-2013	Inpatient PLHIV aged ≥18 years admitted to medical wards	1 district hospital	Current TB diagnosis and/or were receiving ATT at the time of admission	1 spot and 1 induced samples, 2 induced if necessary, if too unwell for induction then 2 spot samples, additional samples according to medical team	1 spot and 1 induced sample, 2 induced if necessary, if too unwell for induction then 2 spot samples, additional samples according to medical team	Liquid	Blood culture, urine Xpert (fresh and frozen), other samples if clinically indicated	Old (2)	Yes (biobanked)¶
Thit, 2017[5]	Myanmar	2015-2015	Inpatient PLHIV admitted to medical wards	1 tertiary hospital	-	1 spot sample, induced if unable to expectorate	1 spot sample, induced if unable to expectorate	Solid	No	New (1)	-

*Follow up study by Bjerrum et al (2020)

¶Follow up study by Broger et al (2019)

Definition of abbreviations: ART = antiretroviral therapy, ATT = anti-tuberculosis treatment, PLHIV = people living with HIV, TB = tuberculosis

Table S4 - Risk of bias results of studies that assessed proportion of HIV-positive inpatients eligible for AlereLAM testing

Domain	Gupta-Wright, 2018	Huerga, 2021	Lawn, 2015	Bjerrum, 2015	Thit, 2017
1. Was the sample frame appropriate to address the target population?¶	Yes	Yes	Yes	No¶¶	Yes
2. Were study participants recruited in an appropriate way?§	Yes	Yes	Yes	Yes	Yes
3. Were the study subjects and setting described in detail?*	Yes	Yes	Yes	Yes	Yes
4. Were valid methods used for the identification of eligibility criteria?#	Yes	Yes	Yes	Yes	Yes
5. Was the response rate adequate (>80%)?	No†	Yes	Yes	Yes	Yes

¶The sample frame was considered inappropriate if a certain group was used and the results were then inferred to the target population?

¶¶For Bjerrum et al (2015), study inclusion criteria were ART naïve and WHO stage 3/4, CD4 count <=350 per µL, or pregnant

§Was recruitment conducted using a consecutive or random sample?

*Was the study sample described in sufficient detail so that other researchers can determine if it is comparable to the population of interest to them? For example, did the study report age, gender, ART status, and CD4 count?

#Were eligibility items (e.g WHO four-symptom screen and CD4 count) assessed based on existing definitions or diagnostic criteria?

Definition of abbreviations: ART = antiretroviral therapy

Table S5 - Risk of bias and applicability results on the QUADAS-2 criteria tool*

	LF-LAM analyses					Screening test analyses#				
	Bjerrum, 2015	**Gupta-Wright, 2018**	**Huerga, 2021**	**Lawn, 2015**	**Thit, 2017**	**Bjerrum, 2015**	**Gupta-Wright, 2018**	**Huerga, 2021**	**Lawn, 2015**	**Thit, 2017**
Patient selection (Risk of Bias)¶	High	High	Low	Low	Low	High	High	Low	Low	Low
Index test (Risk of Bias)¶¶	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Reference test (Risk of Bias)§	High	High	High	Low	High	Low	Low	Low	Low	Low
Flow and timing (Risk of Bias)§§	Low	High	High	Low	Low	Low	Low	Low	Low	Low
Patient selection (Applicability)†	High	Low	Low	Low	Low	High	Low	Low	Low	Low
Index test (Applicability)†	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Reference test (Applicability)†	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low

*Assessment done for all index tests.

#For the domain index test (risk of bias), the risk of bias was judged high for BMI as the index test (>20% missing data).

¶Was a consecutive or random sample of patients enrolled? Did the study avoid inappropriate exclusions?

¶¶Were the index test results interpreted without knowledge of the results of the reference standard?

§Is the reference standard likely to correctly classify the target condition? For example, were both pulmonary and extrapulmonary samples obtained for LF-LAM analyses? Were the reference standard results interpreted without knowledge of the results of the index test?

§§Was there an appropriate interval between index test(s) and reference standard? Did all patients receive a reference standard? Did all patients receive the same reference standard? Were all patients included in the analysis?

†Are there concerns that the included patients (patient selection), index test, or target condition (reference standard) do not match the review question?

Table S6 - Percentage of missing data for each variable by study†§

Variable†	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Huerga	Lawn	Thit
Clinical setting	0	0	0	0	0	0
Age	0	0	0	0	0	0
Sex	0	0	0	0	0	0
ART status	0	0	0	4	0	0
History of tuberculosis	0	0	0	100	0	0
WHO symptoms*	0	0	0	0	0	0
Cough	1	0	0	0	1	0
Fever	4	0	0	1	1	0
Weight loss	1	0	0	1	0	0
Night sweats	0	0	0	0	1	0
Cough >=2 weeks	1	1	1	100	1	100
BMI	14	0	0	28	100	2
Lymphadenopathy	0	100	100	1	100	0
WHO-defined danger sign**	100	0	0	0	100	100
WHO stage	4	100	100	100	100	0
CD4 count	10	0	1	2	0	0
CRP	100	100	100	100	5	100
Haemoglobin	6	0	0	1	1	11
CXR (any abnormality)#	100	100	100	57	100	4
CXR (suggests tuberculosis)	100	100	100	100	100	100
AlereLAM	0	1	100	1	2	0
FujiLAM	3	100	100	100	2	100
Sputum Xpert##	28	35	39	39	54	0
Non-sputum Xpert	20	1	100	100	2	100
Total Xpert##	4	1	39	39	1	0
Sputum culture	0	100	100	100	50	0
Non-sputum culture	100	100	100	100	0	100
Total culture	0	100	100	100	0	0
Total (culture or Xpert)	0	1	39	39	0	0

†<5% missing (green), 5-95% missing (yellow), and >95% missing (red)

§Some datasets received in which some participants with missing data were already excluded

††Study by Gupta-Wright involved an intervention arm (systematically performed AlereLAM, urine Xpert and sputum Xpert) and control arm (systematically performed sputum Xpert only)

*Regarded as missing only if a subject had all four symptoms missing

**Regarded as missing only if a subject had all four WHO-defined danger signs missing. WHO-defined danger sign defined as one or more of the following: respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided

#Study by Huerga et al has a high missing value for CXR (abnormal) because the study site at times had a lack of water and technicians to perform chest x-ray

##Study by Bjerrum et al has a high missing value for Xpert because Xpert only became available after study enrollment began

Variable†	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Huerga	Lawn	Thit

Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Table S7 - Summary of main characteristics for participants overall and by each study

Variable†	All	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Huerga	Lawn	Thit
Frequency	3504 (100)	69 (2)	1287 (37)	1287 (37)	387 (11)	420 (12)	54 (2)
Age (years)	38 (31-46)	37 (32-43)	38 (31-46)	38 (31-46)	38 (32-45)	36 (29-42)	33 (30-44)
N	3504	69	1287	1287	387	420	54
CD4 count (cells/μL)	205 (66-408)	41 (12-115)	231 (78-438)	222 (80-436)	173 (51-370)	150 (56-312)	97 (42-264)
N	3479	62	1286	1279	380	418	54
CD4 <=200 cells/μL	1709 (49)	53 (85)	572 (44)	592 (46)	205 (54)	252 (60)	35 (65)
N	3479	62	1286	1279	380	418	54
Female	1992 (57)	33 (48)	727 (56)	734 (57)	216 (56)	255 (61)	27 (50)
N	3504	69	1287	1287	387	420	54
On ART	2363 (68)	0 (0)	926 (72)	935 (73)	305 (82)	175 (42)	22 (41)
N	3489	69	1287	1287	372	420	54
History of TB	856 (27)	5 (7)	335 (26)	309 (24)	-	190 (45)	17 (31)
N	3115	69	1287	1287	-	418	54
Current Smoker	293 (11)	1 (2)	151 (12)	128 (10)	-	-	13 (24)
N	2693	65	1287	1287	-	-	54
Positive W4SS*	3162 (90)	69 (100)	1152 (90)	1164 (90)	349 (90)	382 (91)	46 (85)
N	3502	69	1287	1287	387	418	54
Cough	1834 (52)	48 (71)	651 (51)	681 (53)	230 (59)	199 (48)	25 (46)
N	3500	68	1287	1287	387	417	54
Fever	1871 (54)	46 (70)	753 (59)	747 (58)	228 (59)	62 (15)	35 (65)
N	3496	66	1287	1287	385	417	54
Weight loss	2521 (72)	65 (96)	906 (70)	875 (68)	277 (73)	356 (85)	42 (78)
N	3495	68	1287	1286	382	418	54
Night sweats	1414 (40)	29 (42)	497 (39)	540 (42)	154 (40)	171 (41)	23 (43)
N	3499	69	1287	1286	386	417	54
Cough >= 2 weeks	731 (24)	35 (51)	342 (27)	321 (25)	-	33 (8)	-
N	3025	68	1271	1270	-	416	-
Lymphadenopathy	58 (11)	8 (12)	-	-	42 (11)	-	8 (15)
N	508	69	-	-	385	-	54
WHO-defined danger sign**	678 (23)	-	277 (22)	275 (21)	126 (33)	-	-
N	2961	-	1287	1287	387	-	-
WHO stage 3 or 4	96 (80)	65 (98)	-	-	-	-	31 (57)
N	120	66	-	-	-	-	54
CXR (any abnormality)	130 (59)	-	-	-	100 (60)	-	30 (58)
N	220	-	-	-	168	-	52
AlereLAM +	368 (17)	18 (26)	158 (12)	-	101 (26)	56 (14)	35 (65)
N	2191	69	1275	-	382	411	54
FujiLAM +	141 (30)	26 (39)	-	-	-	115 (28)	-
N	477	67	-	-	-	410	-
Sputum Xpert +	270 (13)	9 (18)	85 (10)	82 (11)	33 (14)	57 (29)	4 (7)
N	2145	50	832	779	235	195	54
Non-sputum Xpert +	168 (10)	5 (9)	74 (6)	-	-	89 (22)	-

Variable†	All	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Huerga	Lawn	Thit
N	1736	55	1270	-	-	411	-
Total Xpert +§	369 (13)	12 (18)	122 (10)	82 (11)	33 (14)	116 (28)	4 (7)
N	2827	66	1279	779	235	414	54
Sputum culture +	75 (23)	13 (19)	-	-	-	58 (28)	4 (7)
N	332	69	-	-	-	209	54
Non-sputum culture +	70 (17)	-	-	-	-	70 (17)	-
N	420	-	-	-	-	420	-
Total culture +¶	126 (23)	13 (19)	-	-	-	109 (26)	4 (7)
N	543	69	-	-	-	420	54
Total Xpert & culture +	401 (14)	18 (26)	122 (10)	82 (11)	33 (14)	139 (33)	7 (13)
N	2836	69	1279	779	235	420	54
BMI (kg/m2)	20 (18-24)	19 (17-21)	21 (18-24)	21 (18-24)	18 (17-21)	-	20 (17-21)
N	2966	59	1287	1287	280	-	53
CRP (mg/L)	75 (18-157)	-	-	-	-	75 (18-157)	-
N	400	-	-	-	-	400	-
CRP (>=10 mg/L)	334 (84)	-	-	-	-	334 (84)	-
N	400	-	-	-	-	400	-
Hb, Median (g/dL)	10 (8-12)	7 (5-10)	11 (8-13)	11 (8-13)	9 (7-11)	10 (8-12)	9 (7-11)
N	3481	65	1284	1285	385	414	48
Hb (<10 g/dL)	1574 (45)	50 (77)	544 (42)	505 (39)	219 (57)	227 (55)	29 (60)
N	3481	65	1284	1285	385	414	48

†Data are count (%) or median (25th-75th percentiles)

††Study by Gupta-Wright involved an intervention arm (systematically collected AlereLAM, urine Xpert and sputum Xpert) and control arm (systematically collected sputum Xpert only)

*W4SS defined as one or more of the following: current cough, fever, night sweats, or weight loss

**WHO-defined danger sign defined as one or more of the following: respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided

§Sputum and/or non-sputum Xpert result

¶Sputum and/or non-sputum culture result

Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Table S8 - Direct comparisons of proportion of W4SS alone with proportion of W4SS in combination with different components of the WHO AlereLAM inpatient algorithm to determine eligibility AlereLAM testing*

	Combination§¶				Positive W4SS				Difference from W4SS
	No studies	N	No screen positive	Prevalence % (95% CI)†	No studies	N	No screen positive	Prevalence % (95% CI)†	Difference (95% CI)†
Positive W4SS					5	3,502	3,162	90 (89-91)	
Positive W4SS or CD4 <=200 cells/µL	5	3,477	3,225	93 (91-95)	5	3,477	3,137	90 (89-91)	3 (2-4)
Positive W4SS or WHO-defined danger sign	2	2,961	2,691	91 (90-92)	2	2,961	2,665	90 (89-91)	1 (0-3)
Positive W4SS or WHO stage 3 or 4**	1	54	48	89 (77-95)	1	54	46	85 (73-92)	4 (1-14)
Positive W4SS or CD4 <=200 cells/µL or WHO-defined danger sign	2	2,945	2,735	93 (92-94)	2	2,945	2,649	90 (89-91)	3 (2-5)
Positive W4SS or CD4 <=200 cells/µL or WHO stage 3 or 4**	1	54	50	93 (82-97)	1	54	46	85 (73-92)	7 (3-18)

*According to WHO screening & diagnostic algorithm, AlereLAM testing for tuberculosis is advised if an inpatient has a positive W4SS (defined as one or more of the following: current cough, fever, night sweats, or weight loss), a CD4 count <= 200 cells/µL, is WHO stage 3 or 4, or has a WHO-defined danger sign (defined as one or more of the following: respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided)

§Combinations are dependent on available variables

¶Screening combination is either variable positive

†Calculated using meta-analysis of proportions

**One study by Bjerrum et al (2015) excluded from analysis as WHO stage 3 or 4 was part of inclusion criteria

Definition of abbreviations: W4SS = WHO four-symptom screen

Table S9 - Diagnostic yield of different diagnostic tests and sample types as a proportion of total microbiologically confirmed tuberculosis cases†

Study	Gupta-Wright intervention*	Gupta-Wright control*	Huerga	Lawn**
Total sample size	1287	1287	387	420
Microbiological sample available	1282	779	385	420
Microbiologically confirmed tuberculosis¶	209	82	115	143
Sputum culture + (%)	-	-	-	58 (41%)
N	-	-	-	209
Non-sputum culture + (%)	-	-	-	70 (49%)
N	-	-	-	420
Total culture + (%)	-	-	-	109 (76%)
N	-	-	-	420
Sputum Xpert + (%)	85 (41%)	82 (100%)	33 (29%)	57 (40%)
N	832	779	235	195
Urine Xpert + (%)	74 (35%)	-	-	89 (62%)
N	1270	-	-	411
Total Xpert + (%)§	122 (58%)	-	-	116 (81%)
N	1279	-	-	414
Culture or Xpert + (%)	122 (58%)	82 (100%)	33 (29%)	139 (97%)
N	1279	779	235	420
Urine AlereLAM + (%)	158 (76%)	-	101 (88%)	56 (39%)
N	1275	-	382	411
Urine FujiLAM + (%)	-	-	-	115 (80%)
N	-	-	-	410
Urine AlereLAM or urine Xpert + (%)§	179 (86%)	-	-	99 (69%)
N	1275	-	-	411
Urine AlereLAM or sputum Xpert + (%)§	196 (94%)	-	115 (100%)	87 (61%)
N	1282	-	385	414
Urine FujiLAM or sputum Xpert + (%)§	-	-	-	131 (92%)
N	-	-	-	413

†Denominator for % is microbiologically confirmed

*Study by Gupta-Wright et al (2018) involved an intervention arm (systematically collected AlereLAM, urine Xpert, and sputum Xpert) and control arm (systematically collected sputum Xpert only)

**The number (%) of all microbiologically cases diagnosed with concentrated urine Xpert was 82 (57%; 402 participants) and with unconcentrated urine Xpert was 59 (41%; 405 participants).

¶Defined as any AlereLAM, Xpert, or culture positive.

§Yield calculated only if study collected all combination tests of interest

Table S10 - The association between WHO-defined danger signs and tuberculosis†*

Variable**	N	Unadjusted OR (95% CI)	N	Adjusted OR (95% CI)
Any WHO-defined danger sign	1667	2.62 (2.01-3.43)		
Individual WHO-defined danger signs				
Inability to walk unaided	1282	3.33 (2.25-4.92)	1277	3.15 (2.1-4.72)
Respiratory rate >30 breaths/min	1279	1.86 (1-3.47)	1277	0.93 (0.44-1.96)
Heart rate >120 beats/min	1280	4.04 (2.68-6.09)	1277	3.31 (2.13-5.13)
Body temperature >39°C	1282	19.41 (5.29-71.18)	1277	10.49 (2.61-42.14)

†Definition of tuberculosis is a positive AlereLAM or sputum Xpert

*In the trial by Gupta-Wright et al (2018), we only included the intervention arm, which collected sputum and urine for Xpert and urine for AlereLAM, and the 2 study sites were considered as separate cohorts

**For the analysis of any WHO-defined danger sign, both studies by Gupta-Wright et al (2018) and Huerga et al (2021) contributed to the analysis. For each individual danger sign, only the study by Gupta-Wright et al contributed to the analysis because it was the only study with available data on each individual danger sign

Definition of abbreviations: OR = Odds ratio

Figure S1 - Study flow diagram

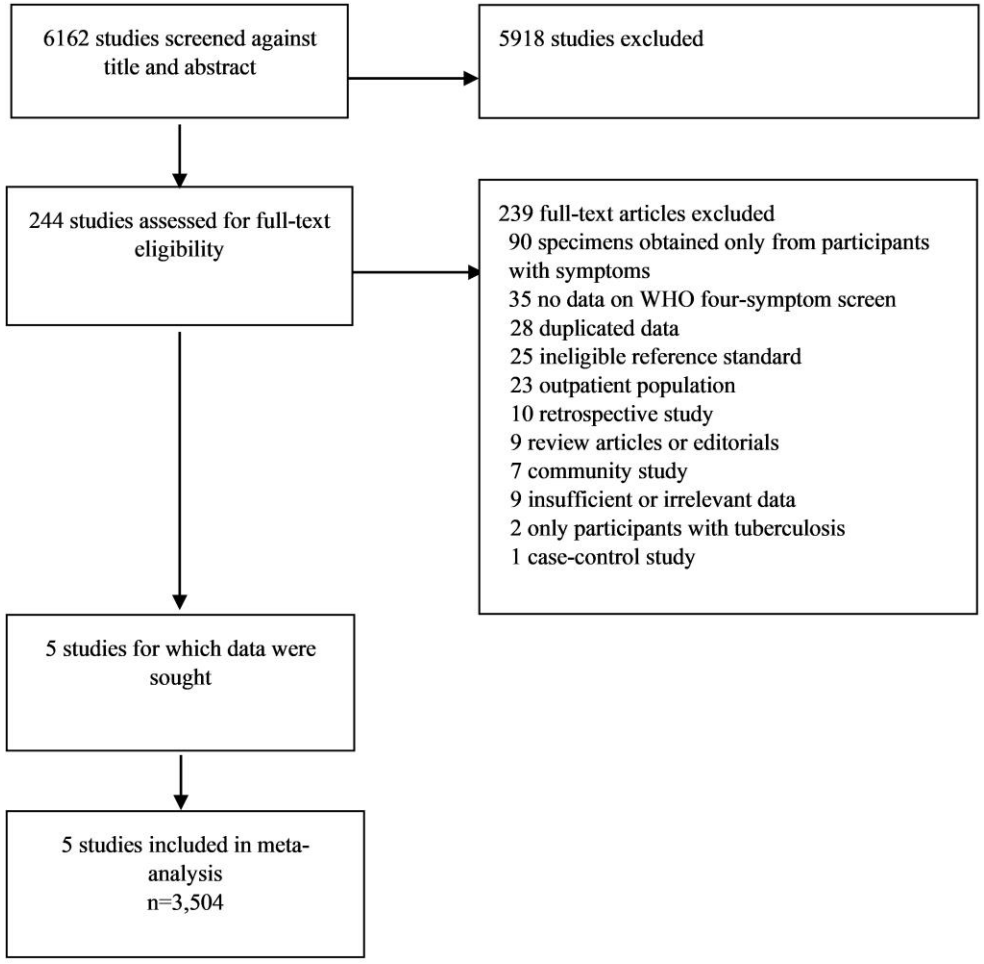
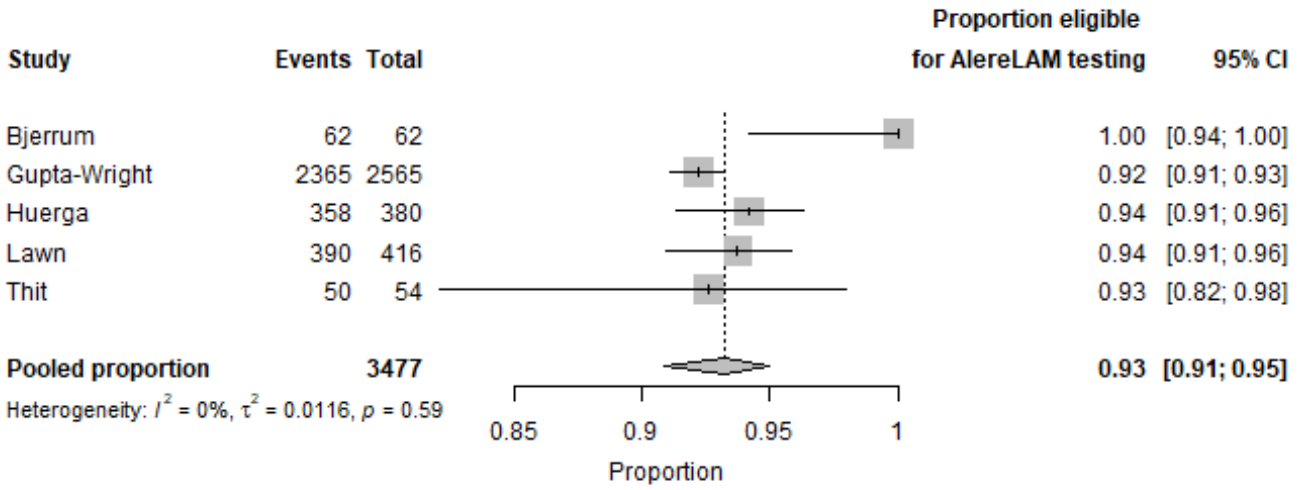
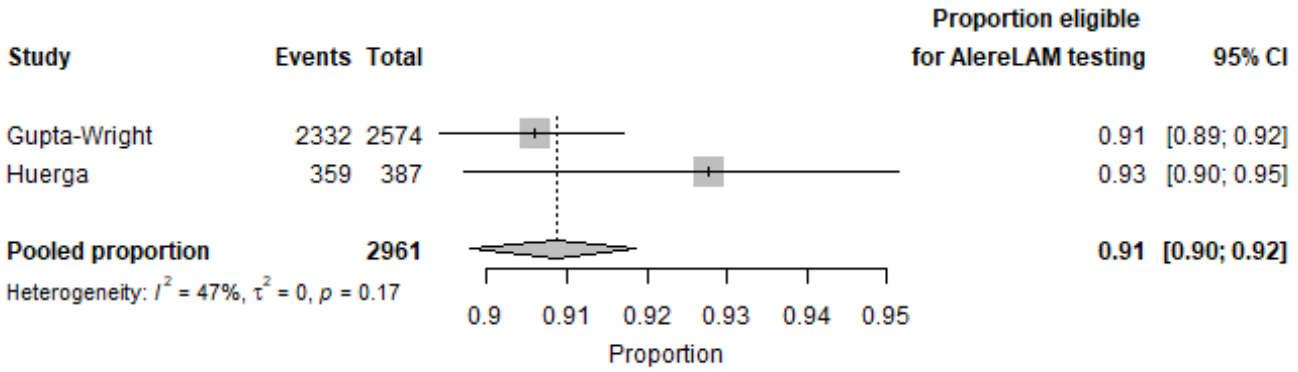


Figure S2 - Forest plots of studies reporting proportion of inpatients eligible for AlereLAM testing according to WHO AlereLAM inpatient algorithm

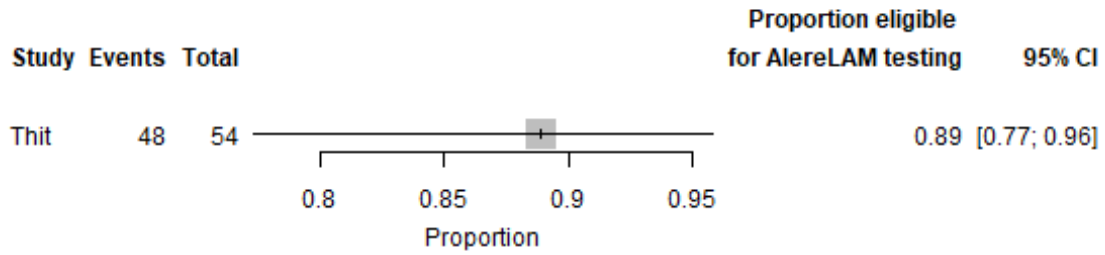
Random-effects meta-analysis of proportion studies reporting screen positivity by W4SS or CD4 <=200 cells/µL in HIV-infected inpatients



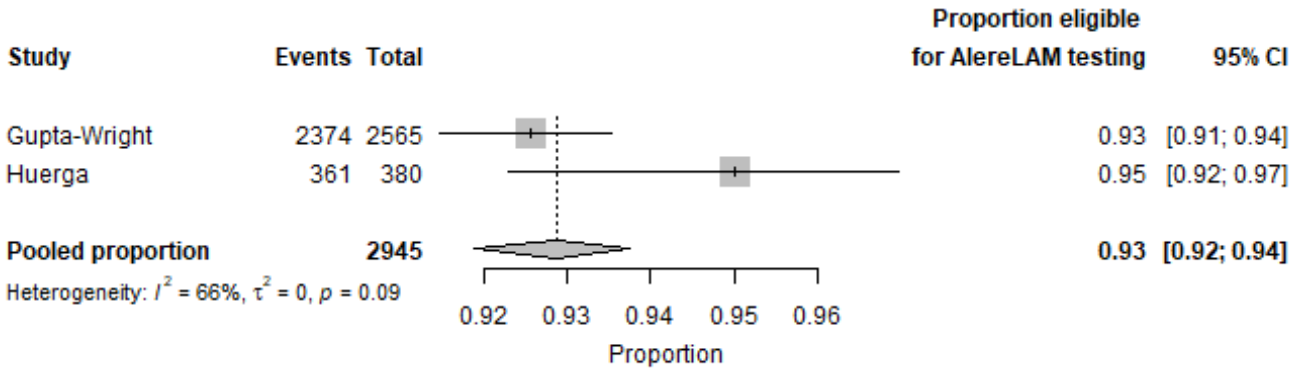
Random-effects meta-analysis of proportion studies reporting screen positivity by W4SS or WHO-defined danger signs in HIV-infected inpatients



Random-effects meta-analysis of proportion studies reporting screen positivity by W4SS or WHO stage 3 or 4 in HIV-infected inpatients



Random-effects meta-analysis of proportion studies reporting screen positivity by W4SS or CD4 <=200 cells/µL or WHO-defined danger signs in HIV-infected inpatients



Random-effects meta-analysis of proportion studies reporting screen positivity by W4SS + CD4 <=200 cells/µL + WHO stage 3 or 4 in HIV-infected inpatients

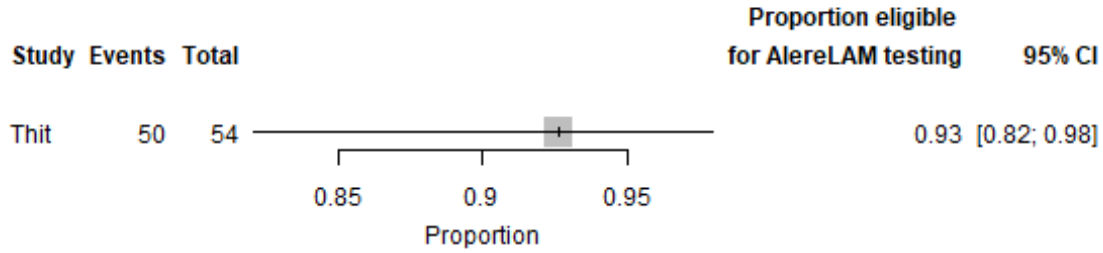
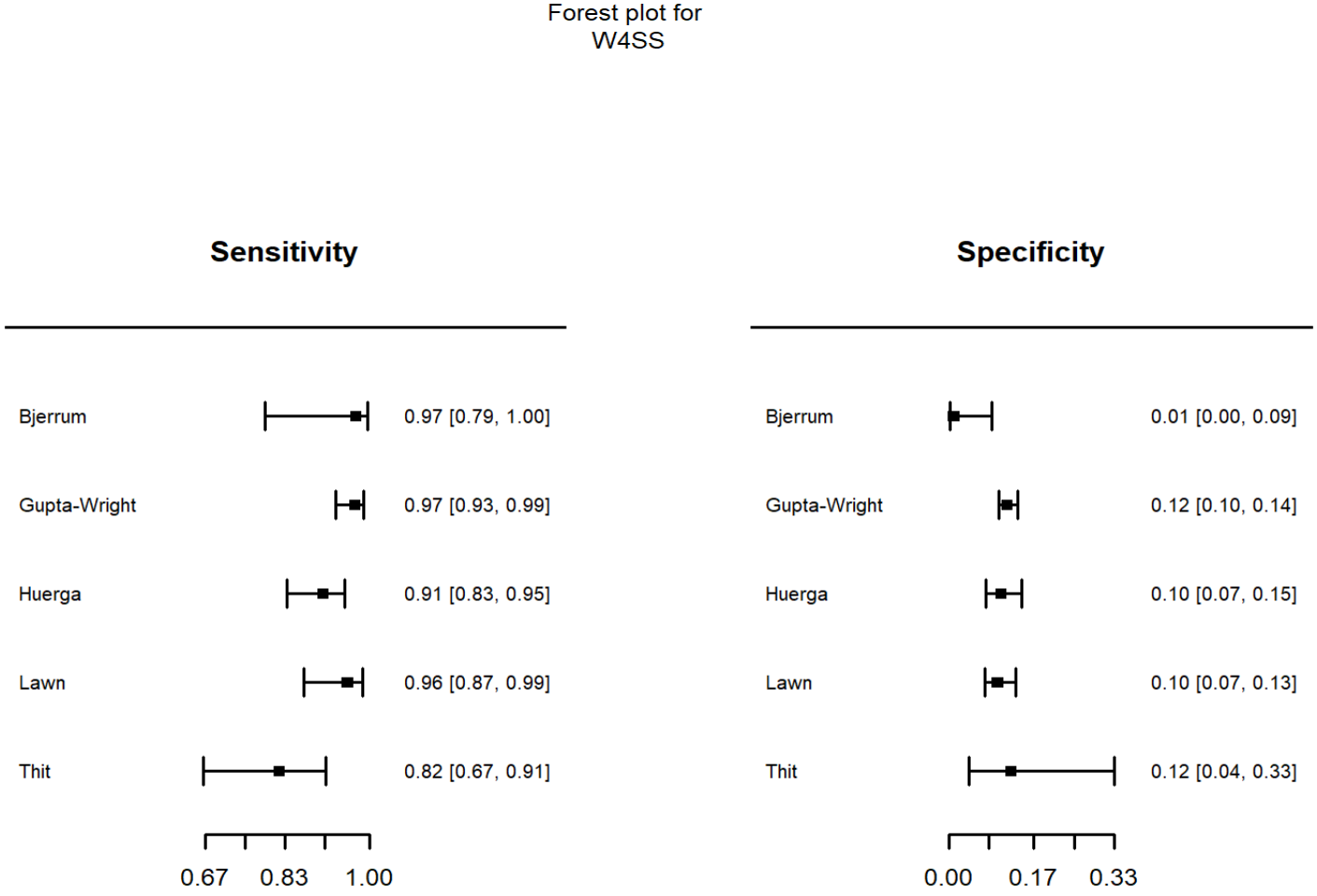
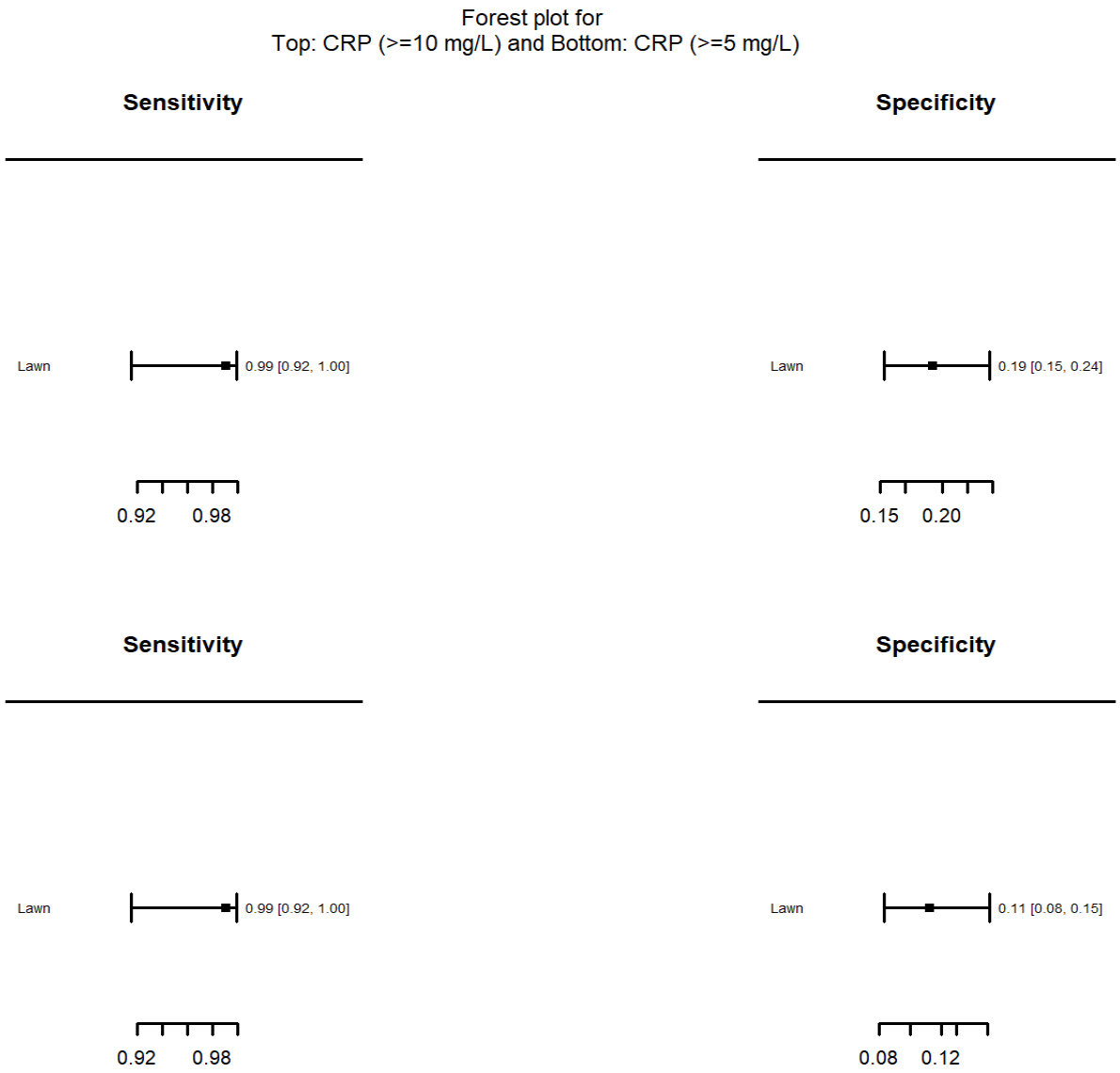
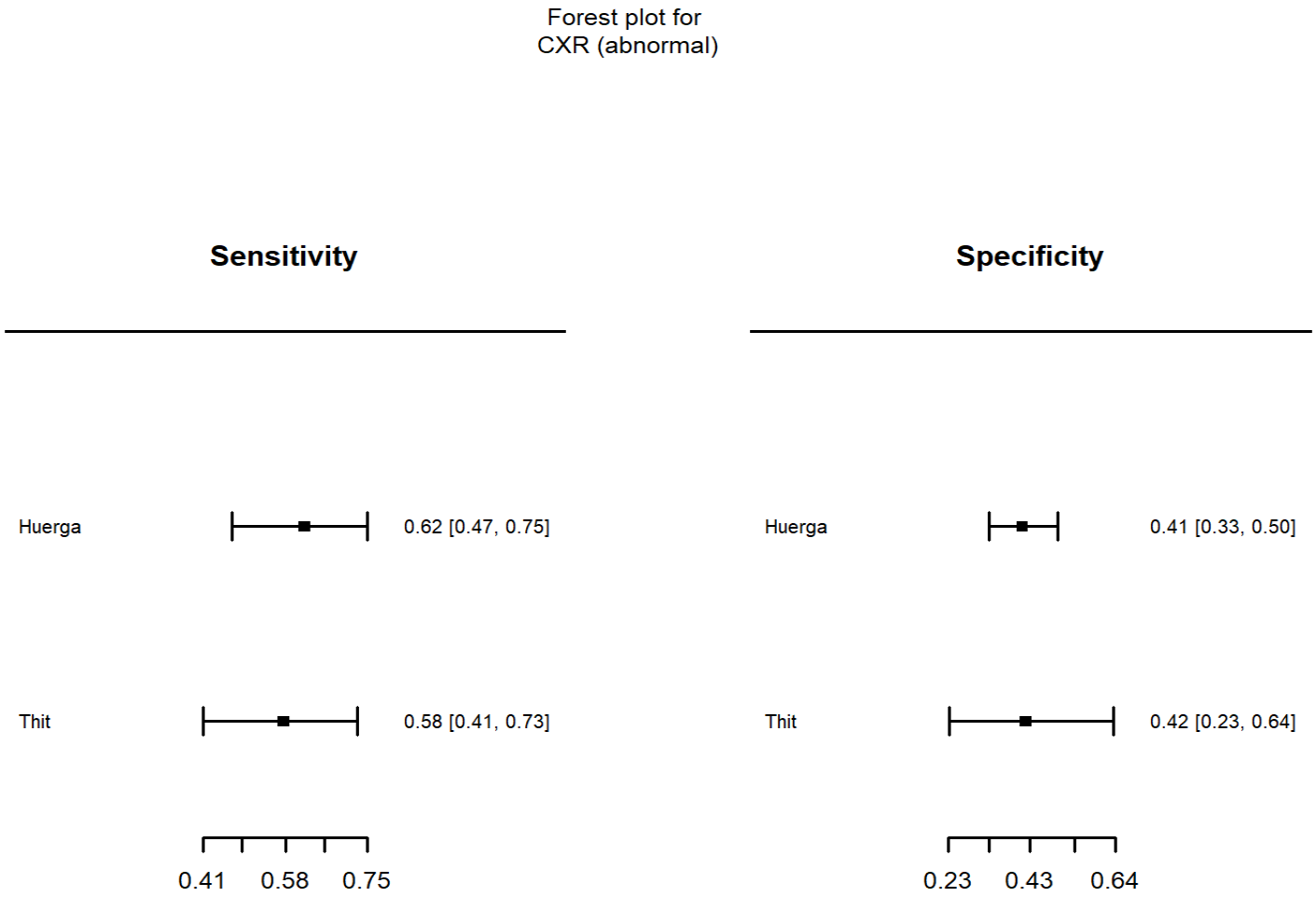


Figure S3 - Forest plots of sensitivity and specificity estimates

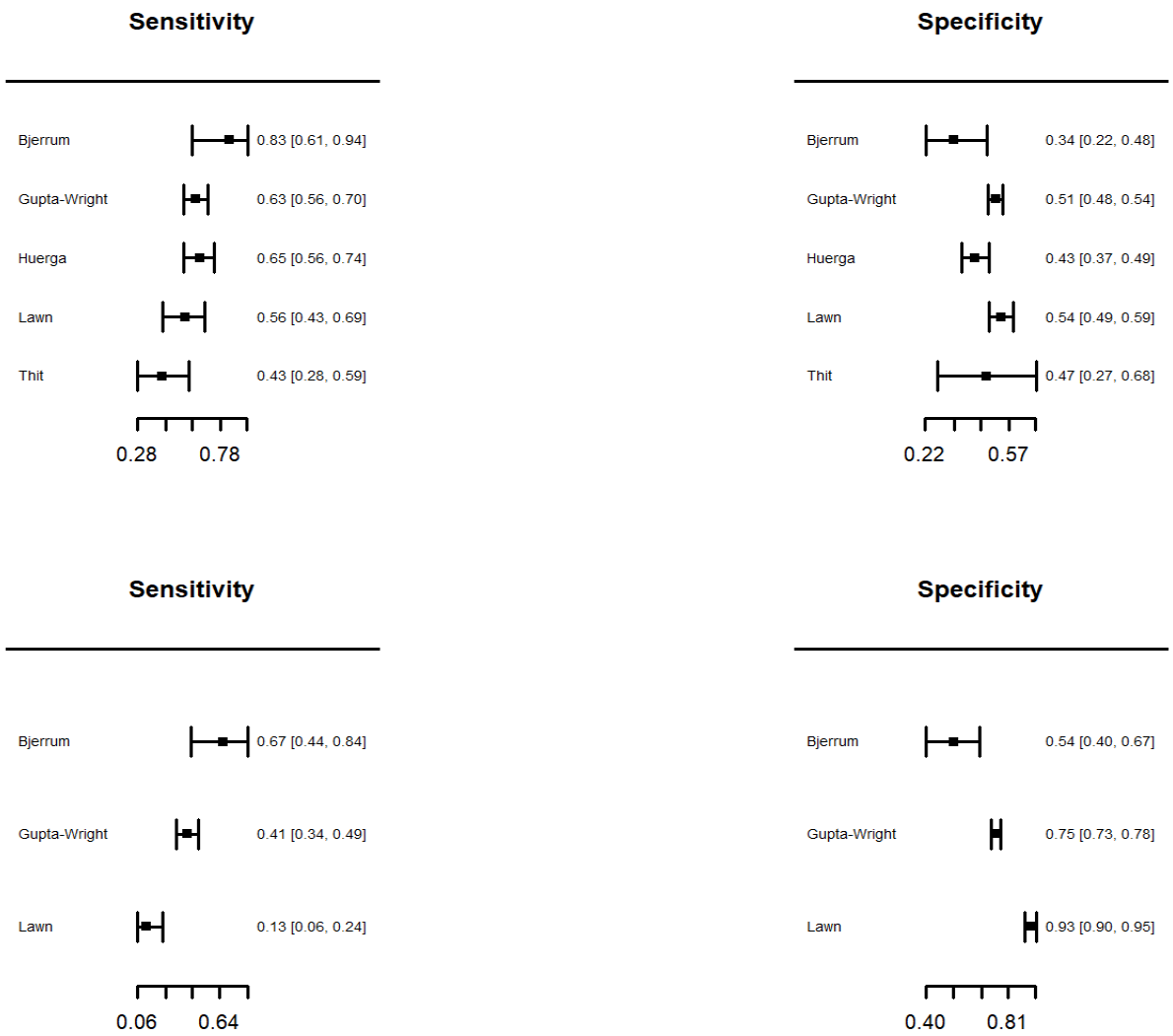
Figure S3A - Forest plots of sensitivity and specificity estimates for each screening test/strategy using AlereLAM as a reference standard



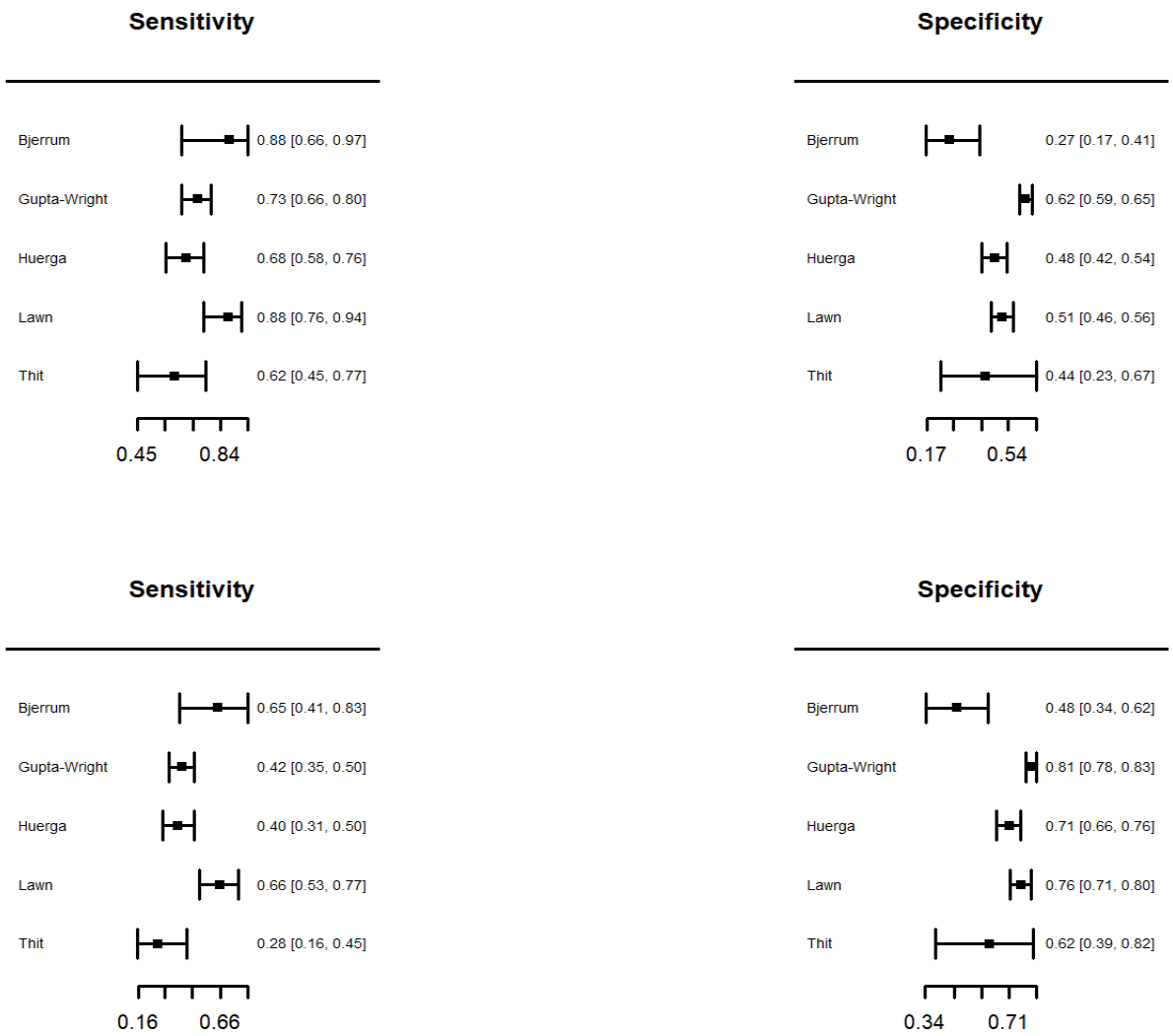

Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
0.99 [0.92, 1.00]
0.19 [0.15, 0.24]
0.92
0.98
0.15
0.20
Sensitivity
Specificity
Lawn
0.99 [0.92, 1.00]
0.11 [0.08, 0.15]
0.92
0.98
0.08
0.12


Forest plot for
CXR (abnormal)
Sensitivity
Specificity
Huerga
0.62 [0.47, 0.75]
Thit
0.58 [0.41, 0.73]
0.41
0.58
0.75
Huerga
0.41 [0.33, 0.50]
Thit
0.42 [0.23, 0.64]
0.23
0.43
0.64

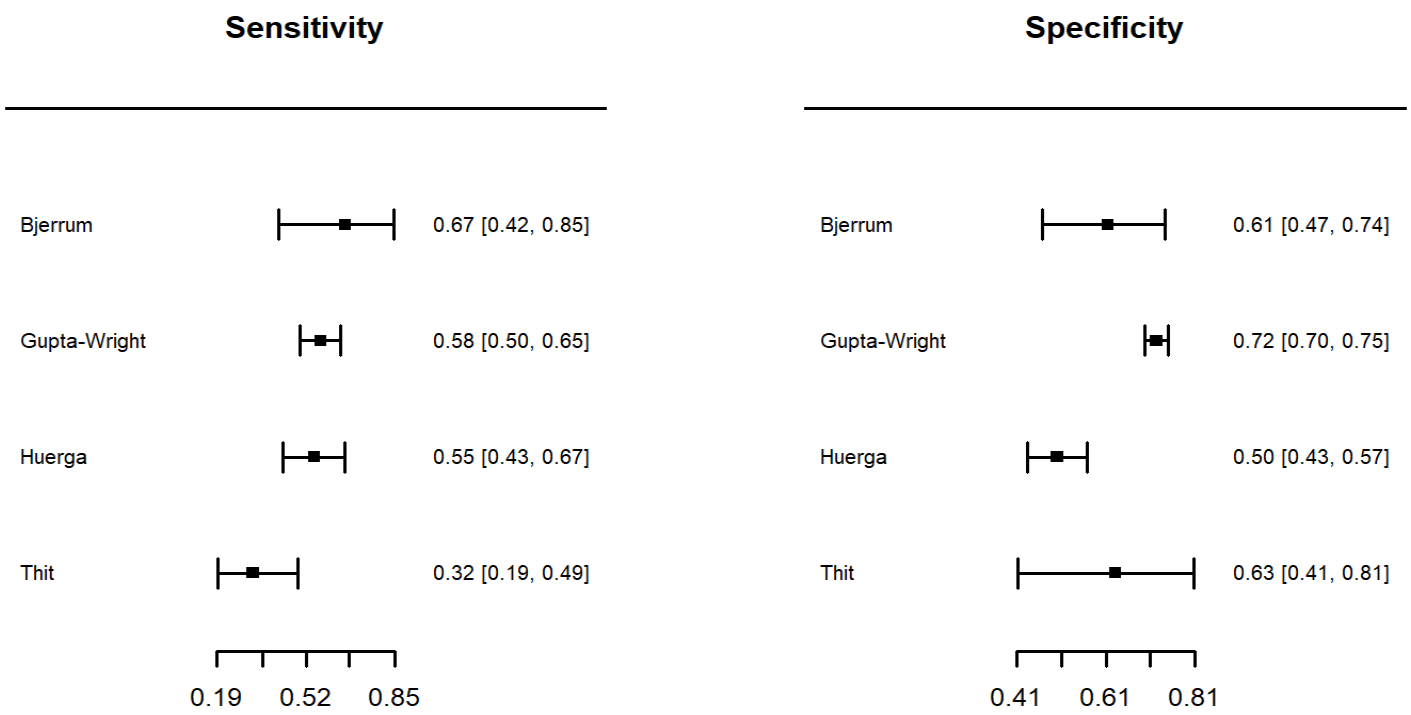
Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

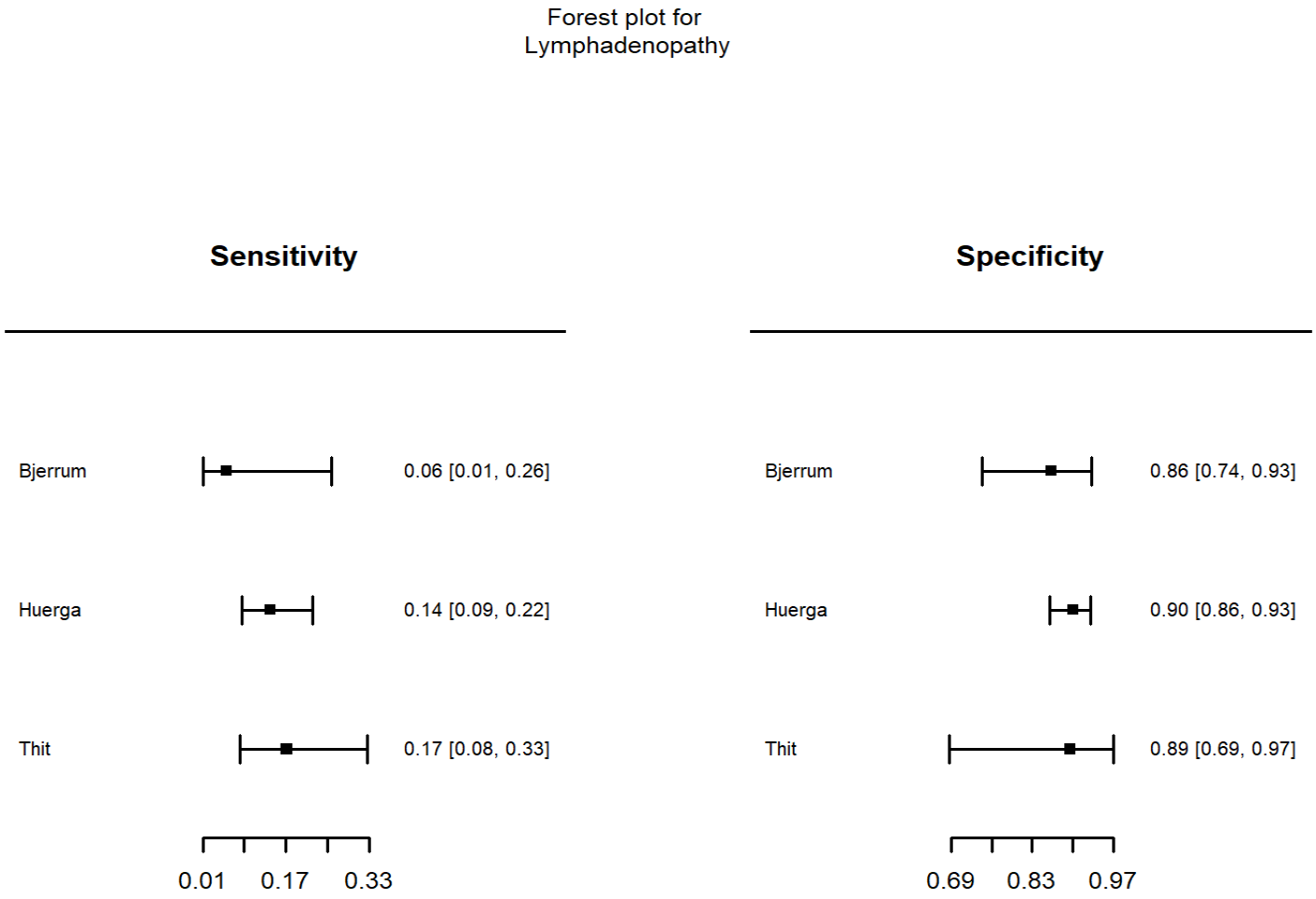


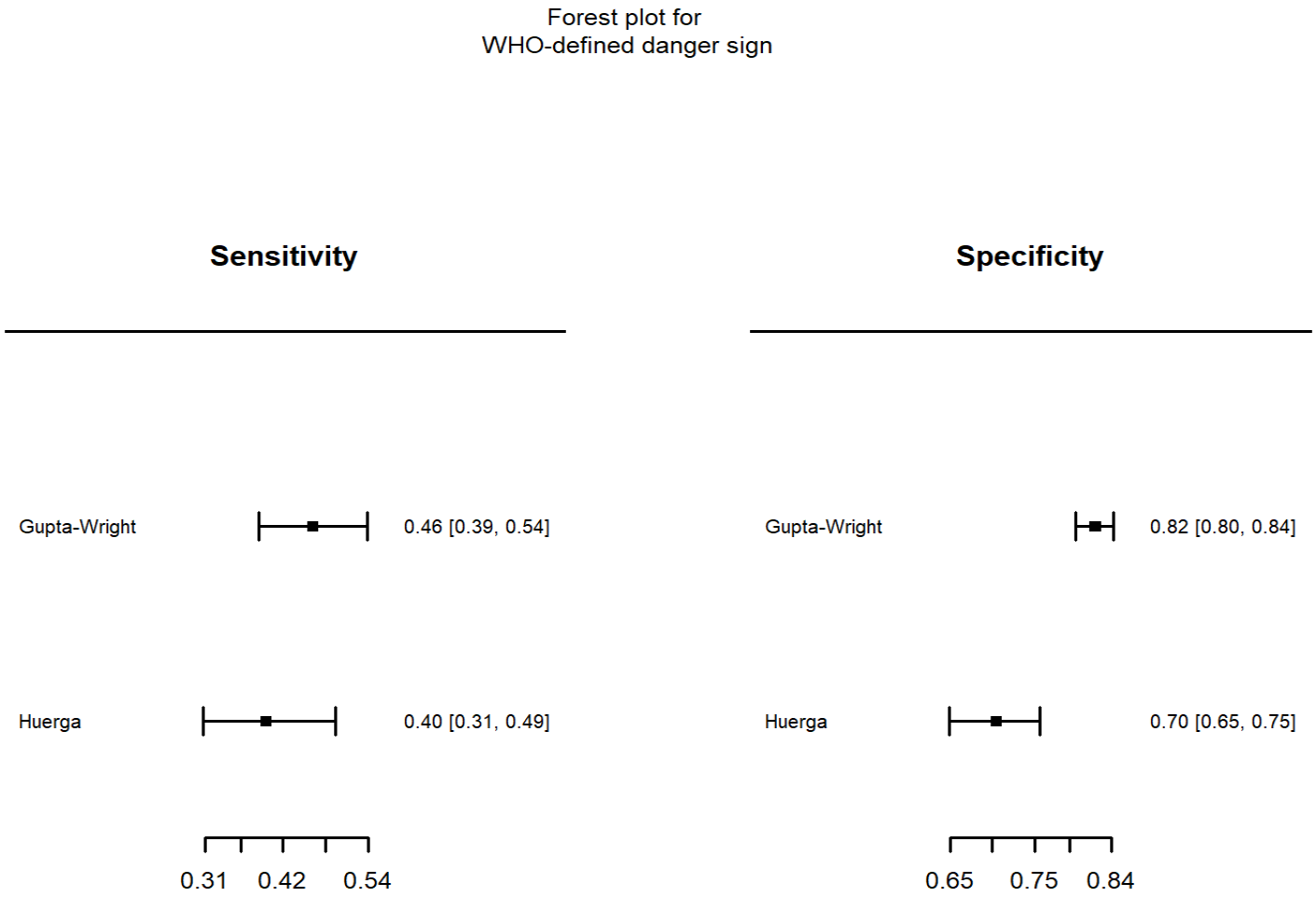
Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)



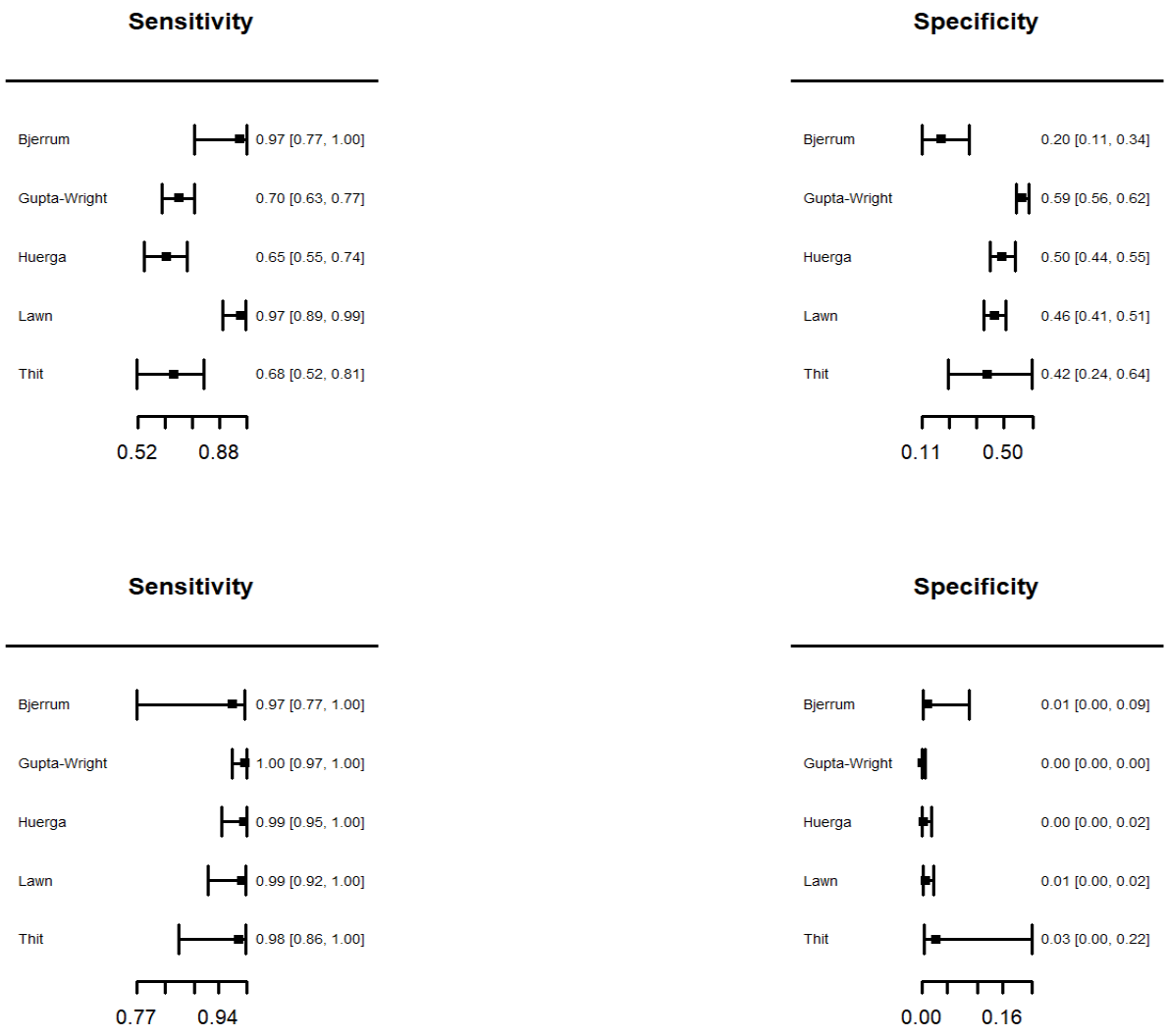
Forest plot for
BMI (<18.5 kg/m²)

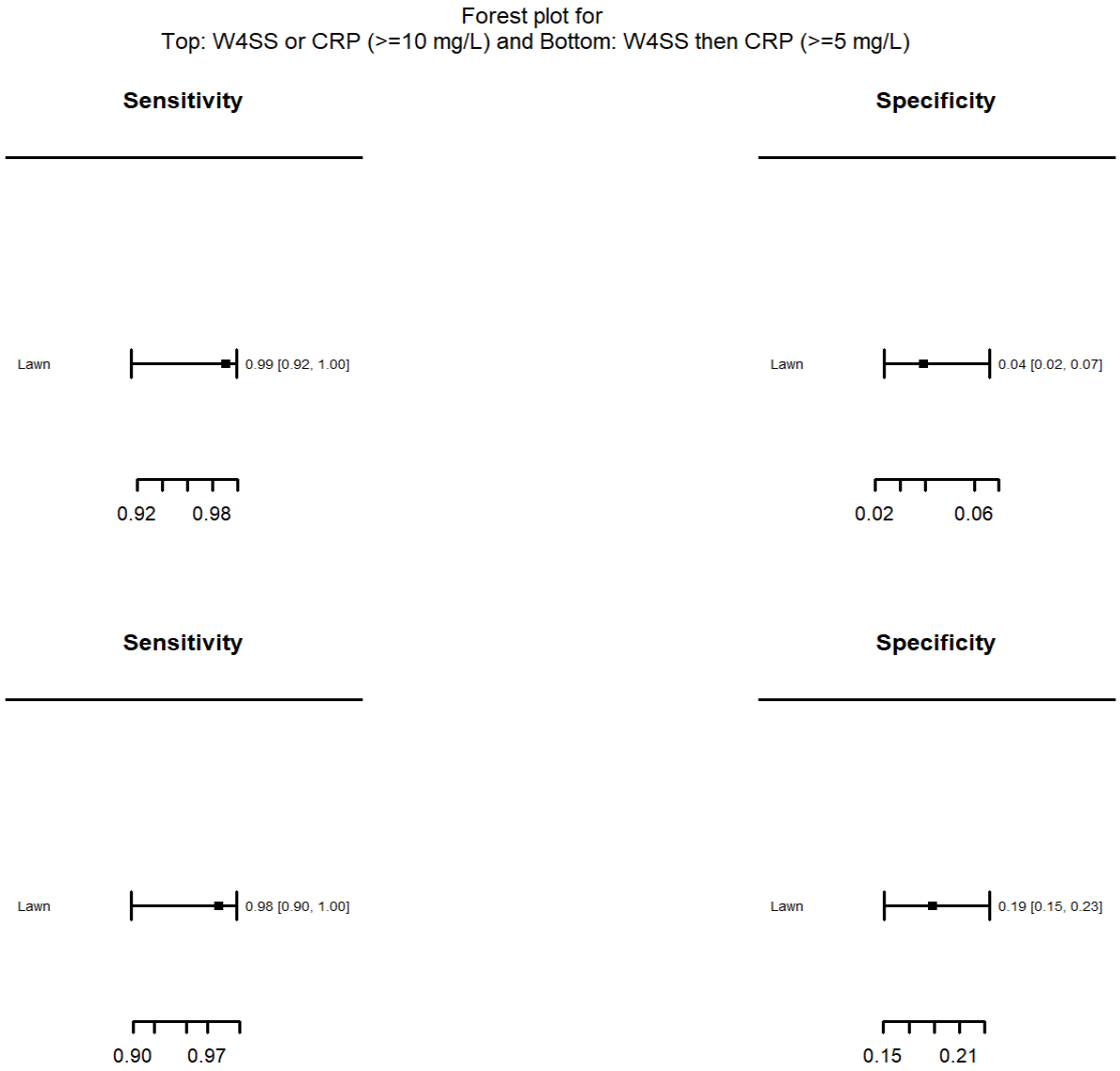



Forest plot for
Lymphadenopathy
Sensitivity
Specificity
Bjerrum
0.06 [0.01, 0.26]
Huerga
0.14 [0.09, 0.22]
Thit
0.17 [0.08, 0.33]
0.01
0.17
0.33
Bjerrum
0.86 [0.74, 0.93]
Huerga
0.90 [0.86, 0.93]
Thit
0.89 [0.69, 0.97]
0.69
0.83
0.97


Forest plot for
WHO-defined danger sign
Sensitivity
Specificity
Gupta-Wright
0.46 [0.39, 0.54]
Huerga
0.40 [0.31, 0.49]
0.31
0.42
0.54
Gupta-Wright
0.82 [0.80, 0.84]
Huerga
0.70 [0.65, 0.75]
0.65
0.75
0.84

Forest plot for
Top: CD4 count <=200 cells/μL and Bottom: W4SS or CD4 count <=200 cells/μL




Forest plot for
Top: W4SS or CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
0.99 [0.92, 1.00]
0.04 [0.02, 0.07]
0.92
0.98
0.02
0.06
Sensitivity
Specificity
Lawn
0.98 [0.90, 1.00]
0.19 [0.15, 0.23]
0.90
0.97
0.15
0.21

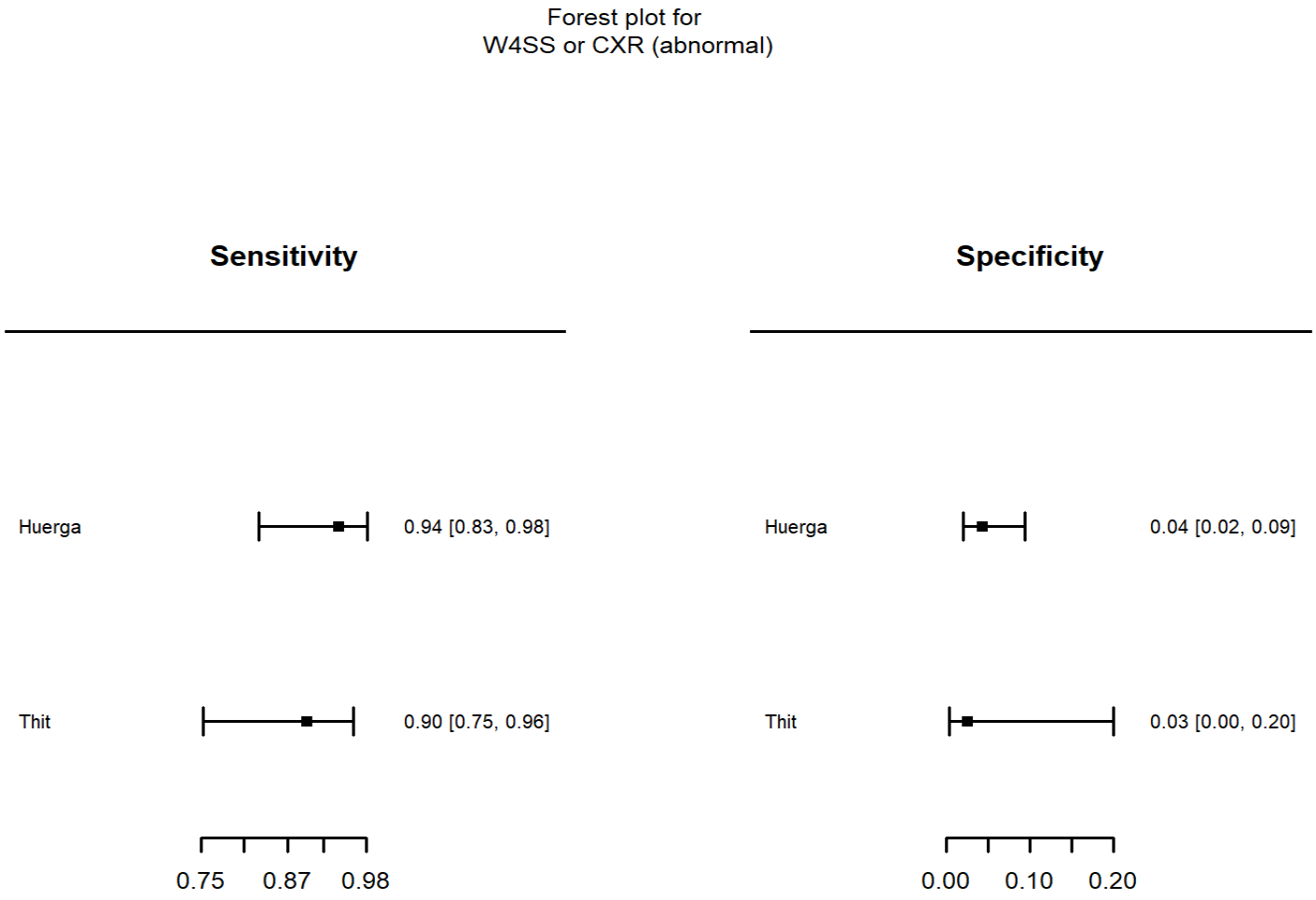
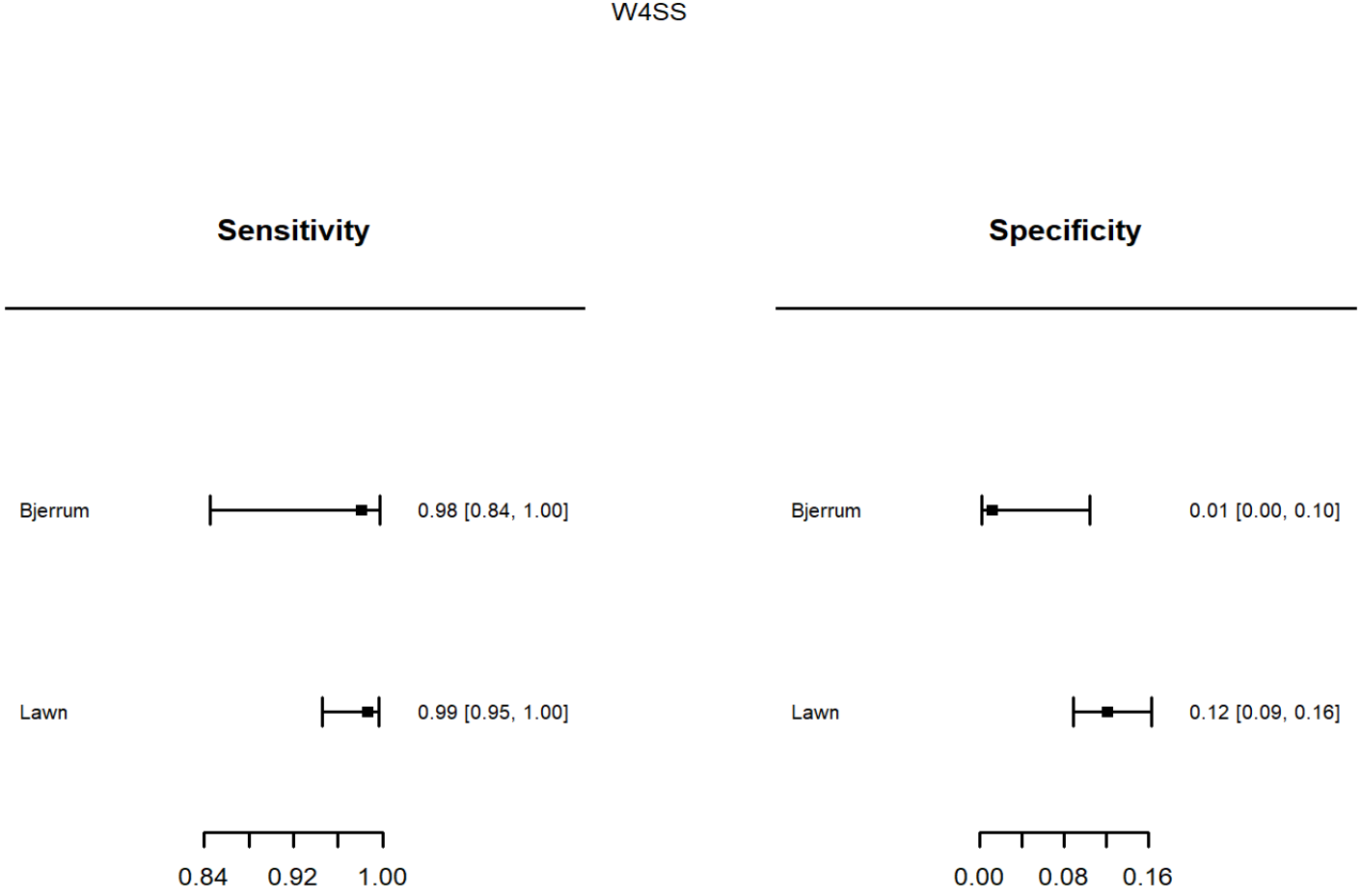
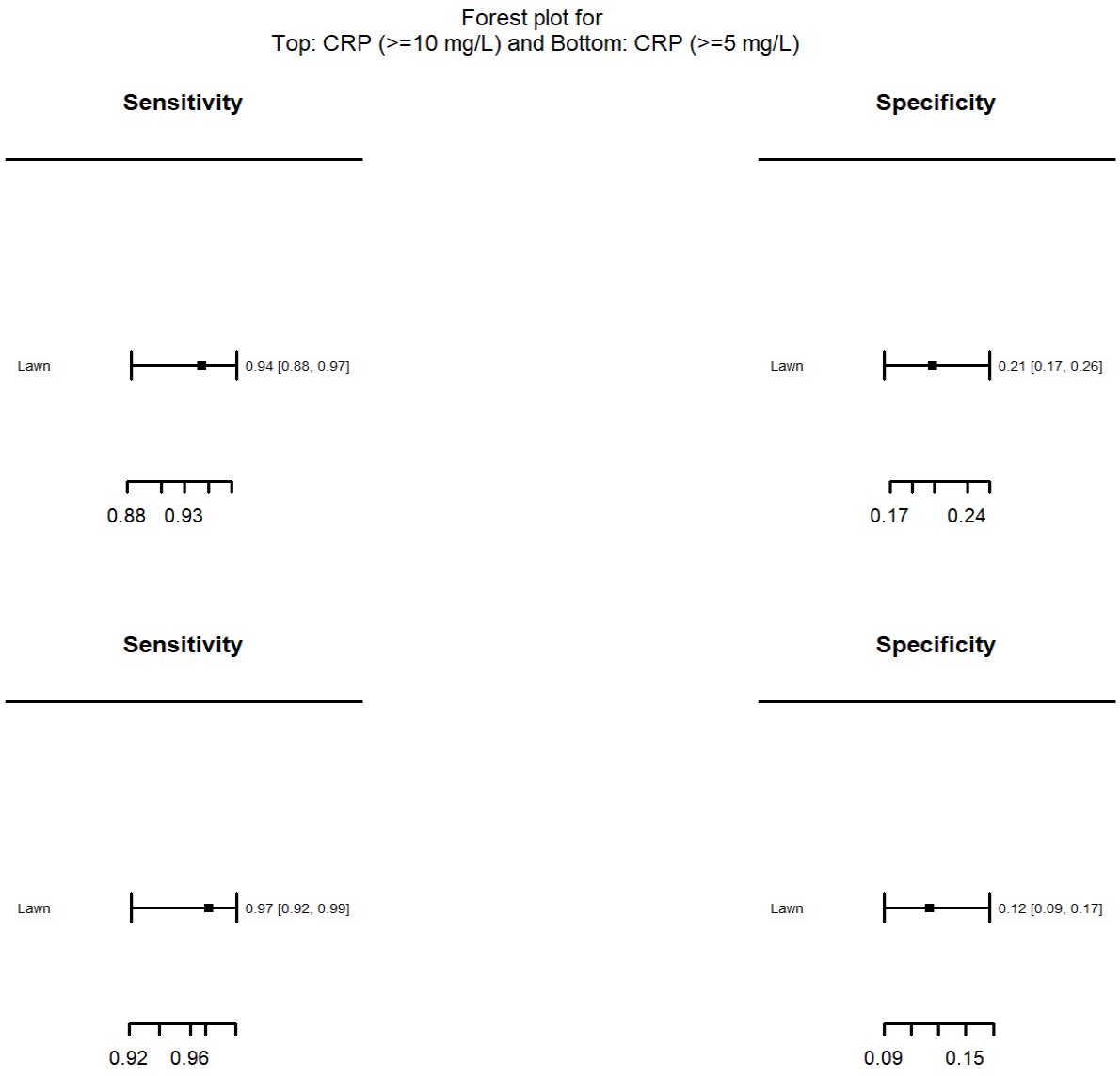
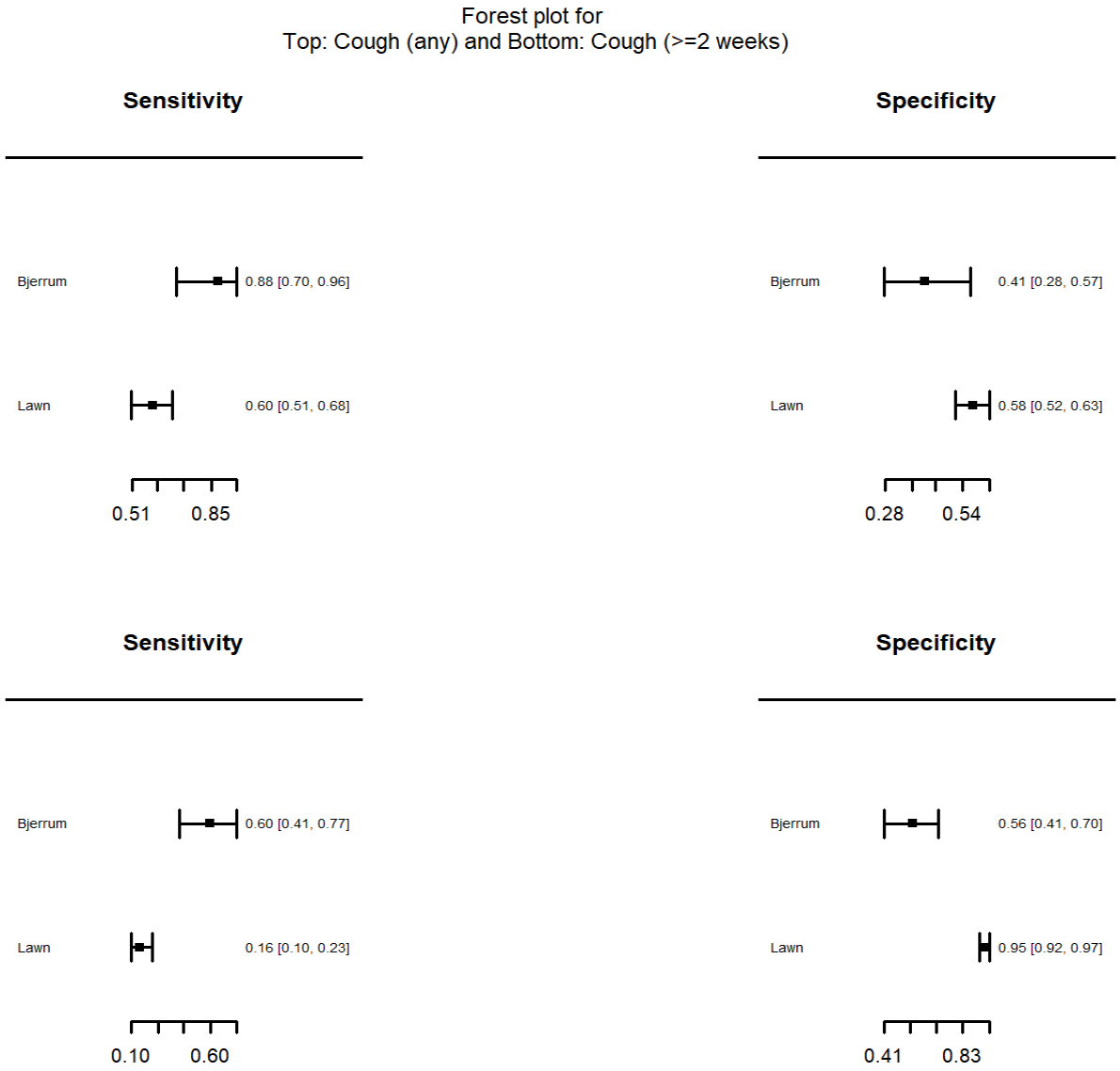
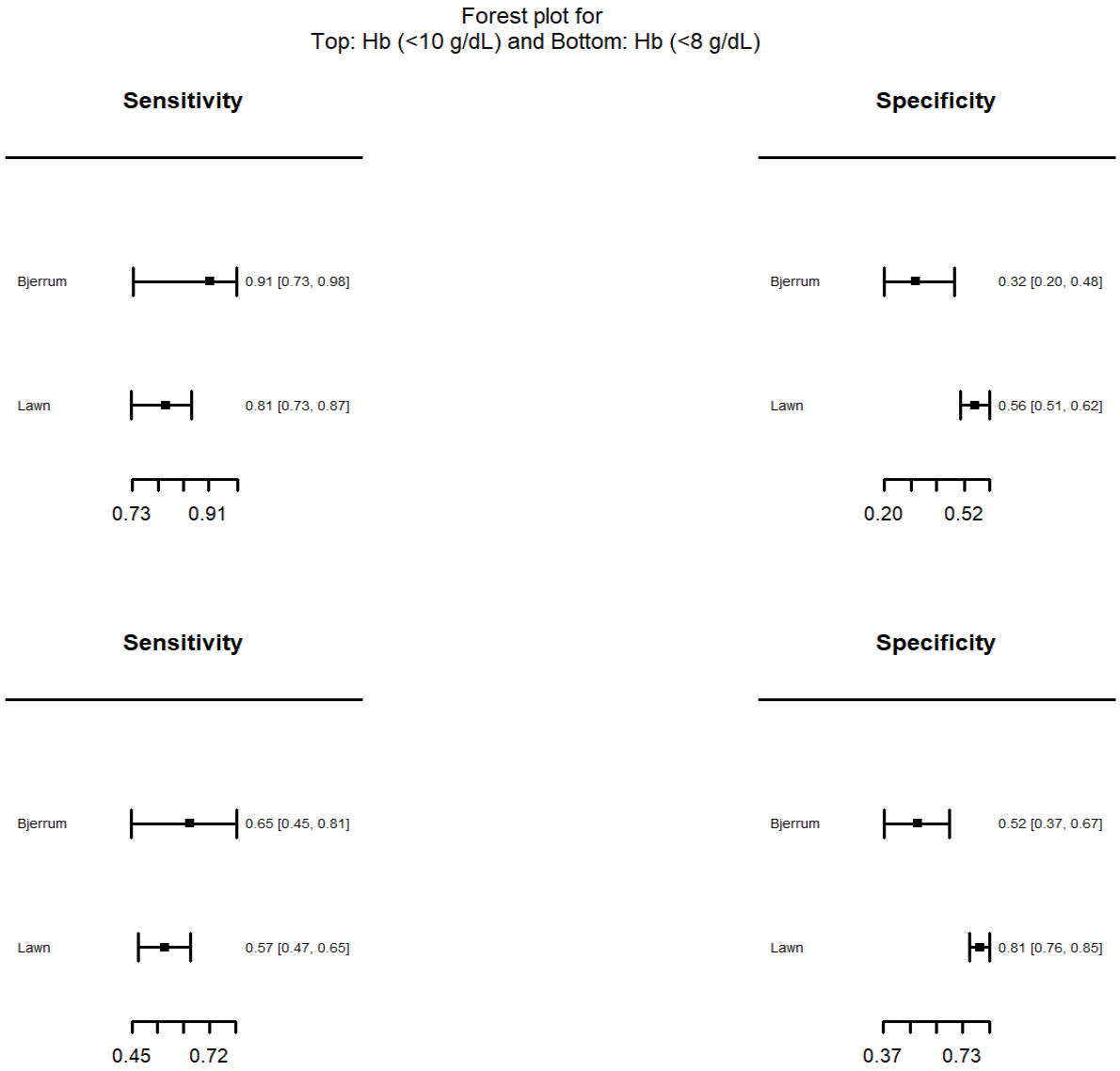

Forest plot for
W4SS or CXR (abnormal)
Sensitivity
Specificity
Huerga
0.94 [0.83, 0.98]
Thit
0.90 [0.75, 0.96]
0.75
0.87
0.98
Huerga
0.04 [0.02, 0.09]
Thit
0.03 [0.00, 0.20]
0.00
0.10
0.20

Figure S3B - Forest plots of sensitivity and specificity estimates for each screening test/strategy using FujiLAM as a reference standard

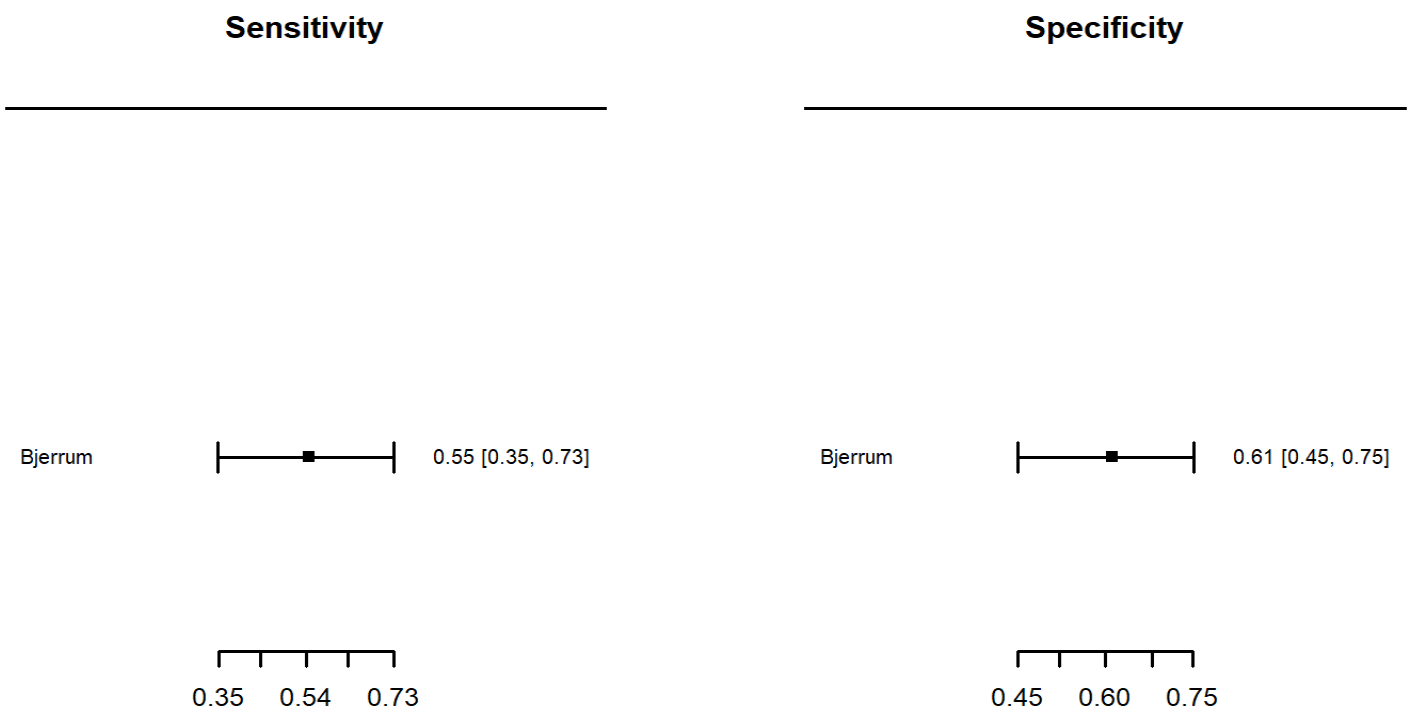


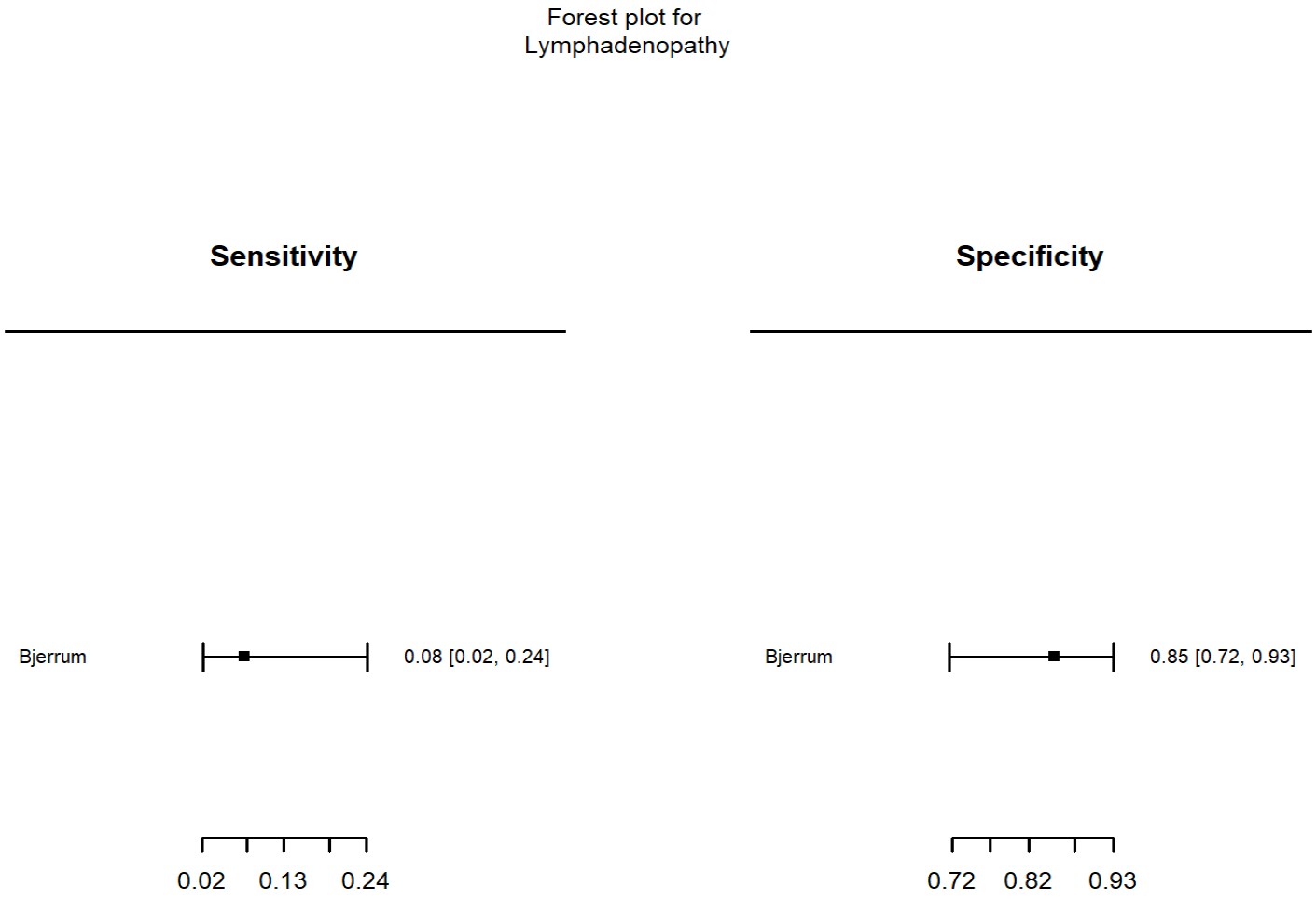

Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
0.94 [0.88, 0.97]
0.21 [0.17, 0.26]
0.88
0.93
0.17
0.24
Sensitivity
Specificity
Lawn
0.97 [0.92, 0.99]
0.12 [0.09, 0.17]
0.92
0.96
0.09
0.15

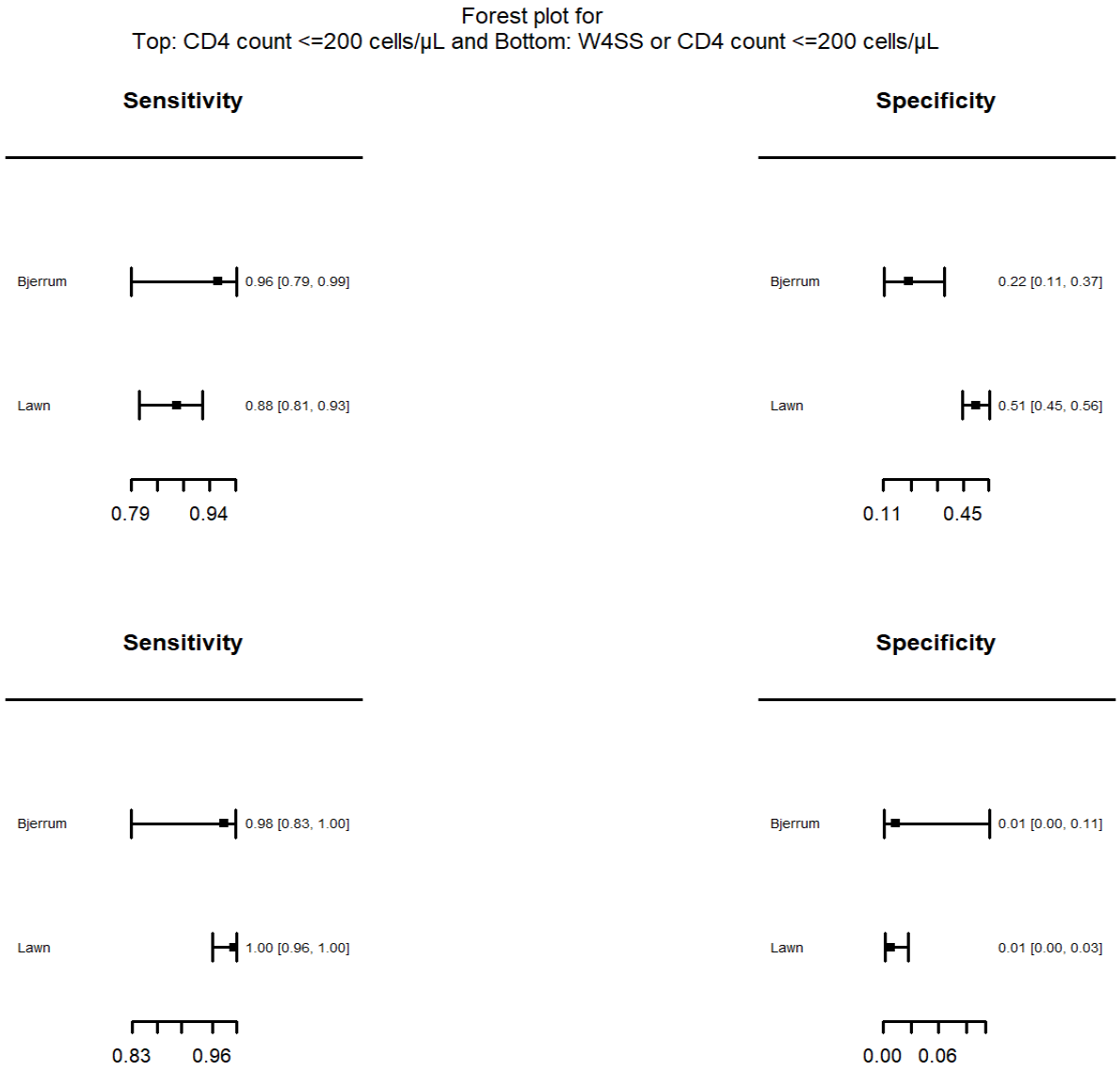

Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)
Sensitivity
Bjerrum
0.88 [0.70, 0.96]
Lawn
0.60 [0.51, 0.68]
0.51
0.85
Specificity
Bjerrum
0.41 [0.28, 0.57]
Lawn
0.58 [0.52, 0.63]
0.28
0.54
Sensitivity
Bjerrum
0.60 [0.41, 0.77]
Lawn
0.16 [0.10, 0.23]
0.10
0.60
Specificity
Bjerrum
0.56 [0.41, 0.70]
Lawn
0.95 [0.92, 0.97]
0.41
0.83


Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)
Sensitivity
Specificity
Bjerrum
0.91 [0.73, 0.98]
Lawn
0.81 [0.73, 0.87]
0.73
0.91
Bjerrum
0.32 [0.20, 0.48]
Lawn
0.56 [0.51, 0.62]
0.20
0.52
Sensitivity
Specificity
Bjerrum
0.65 [0.45, 0.81]
Lawn
0.57 [0.47, 0.65]
0.45
0.72
Bjerrum
0.52 [0.37, 0.67]
Lawn
0.81 [0.76, 0.85]
0.37
0.73

Forest plot for
BMI (<18.5 kg/m²)




Forest plot for
Lymphadenopathy
Sensitivity
Specificity
Bjerrum
0.08 [0.02, 0.24]
0.02
0.13
0.24
Bjerrum
0.85 [0.72, 0.93]
0.72
0.82
0.93


Forest plot for
Top: CD4 count <=200 cells/µL and Bottom: W4SS or CD4 count <=200 cells/µL
Sensitivity
Specificity
Bjerrum
0.96 [0.79, 0.99]
Lawn
0.88 [0.81, 0.93]
0.79
0.94
Bjerrum
0.22 [0.11, 0.37]
Lawn
0.51 [0.45, 0.56]
0.11
0.45
Sensitivity
Specificity
Bjerrum
0.98 [0.83, 1.00]
Lawn
1.00 [0.96, 1.00]
0.83
0.96
Bjerrum
0.01 [0.00, 0.11]
Lawn
0.01 [0.00, 0.03]
0.00
0.06

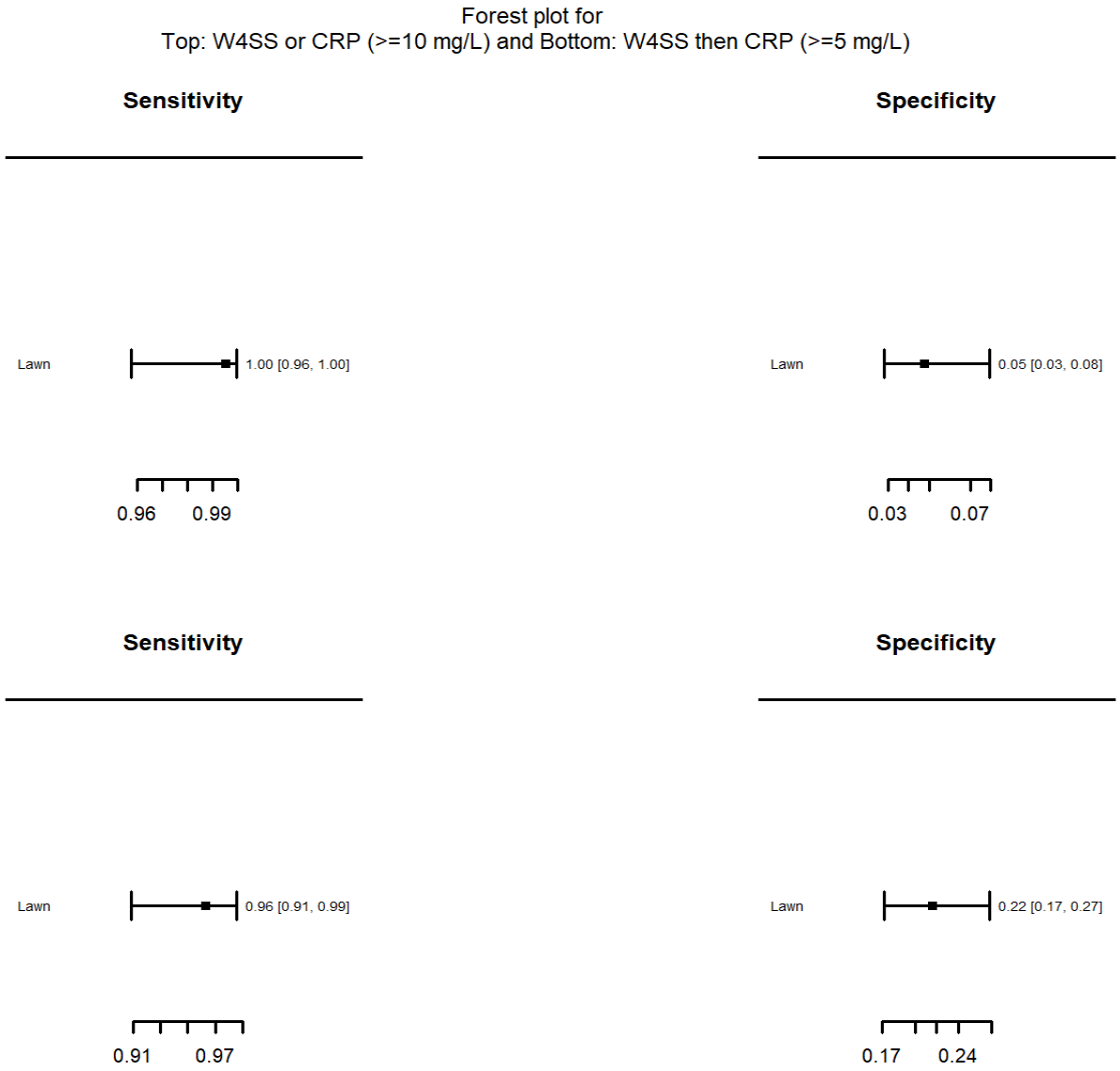
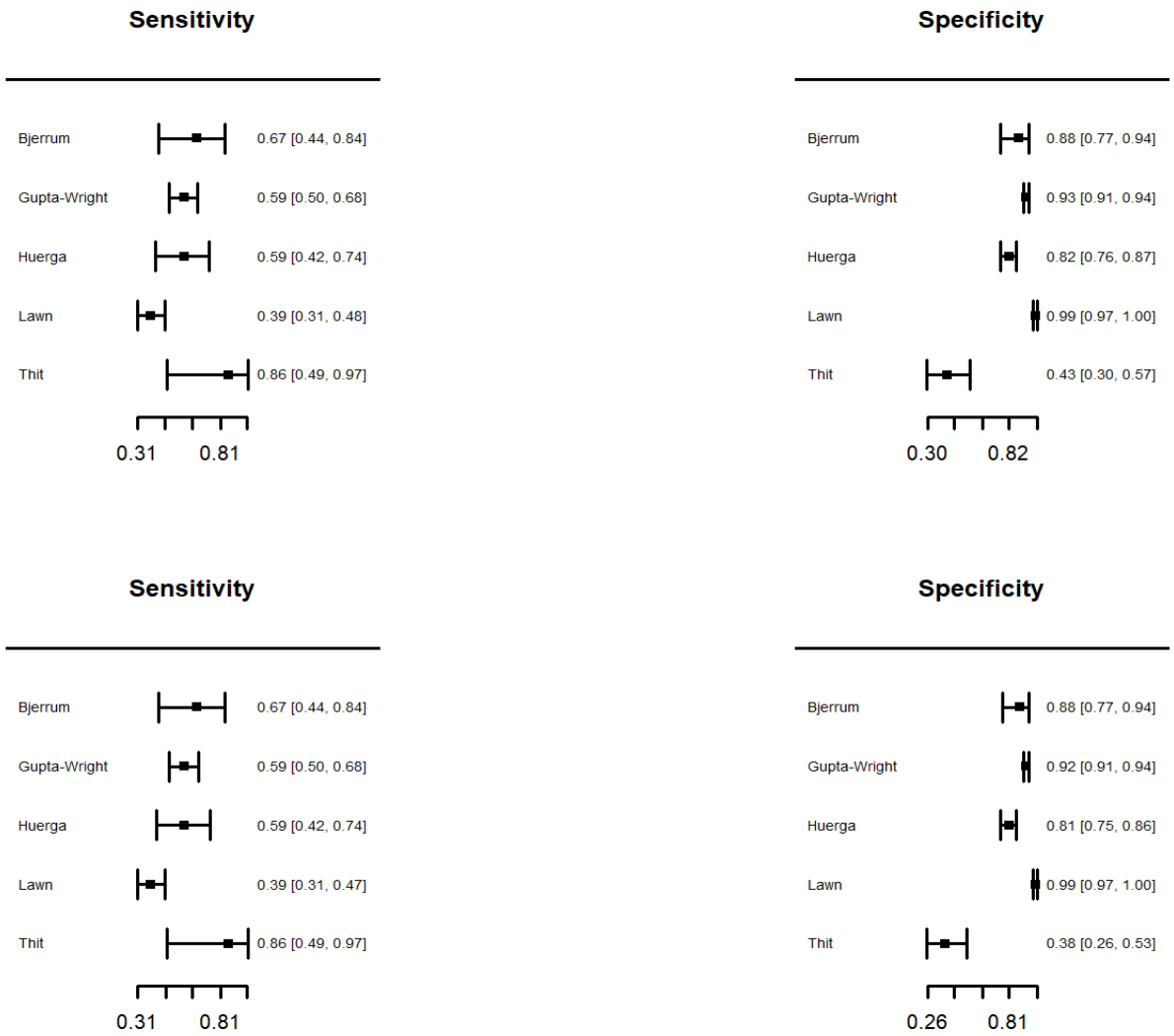

Forest plot for
Top: W4SS or CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
1.00 [0.96, 1.00]
0.05 [0.03, 0.08]
0.96
0.99
0.03
0.07
Sensitivity
Specificity
Lawn
0.96 [0.91, 0.99]
0.22 [0.17, 0.27]
0.91
0.97
0.17
0.24

Figure S3C - Forest plots of sensitivity and specificity estimates for each for each LF-LAM strategy using culture or Xpert as a reference standard
Forest plot for
Top: WHO AlereLAM algorithm and Bottom: AlereLAM alone



Forest plot for
FujiLAM alone

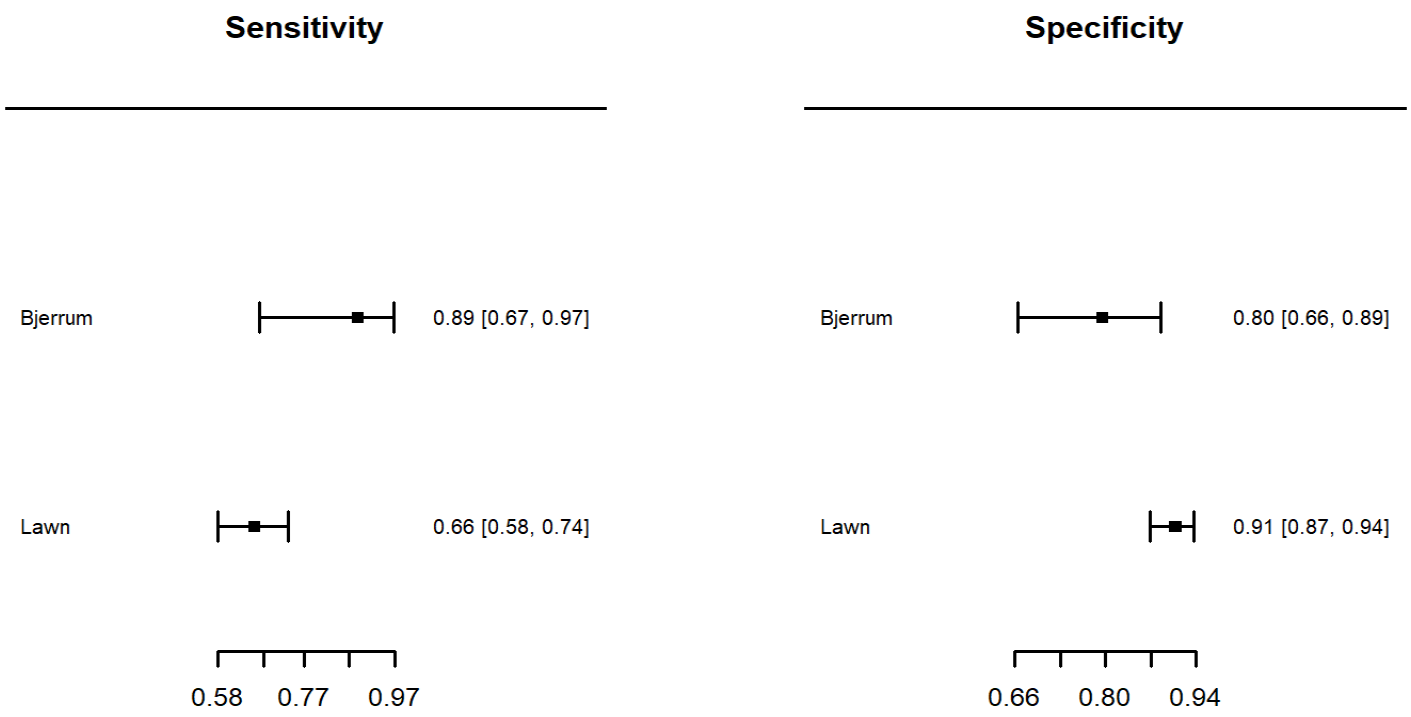
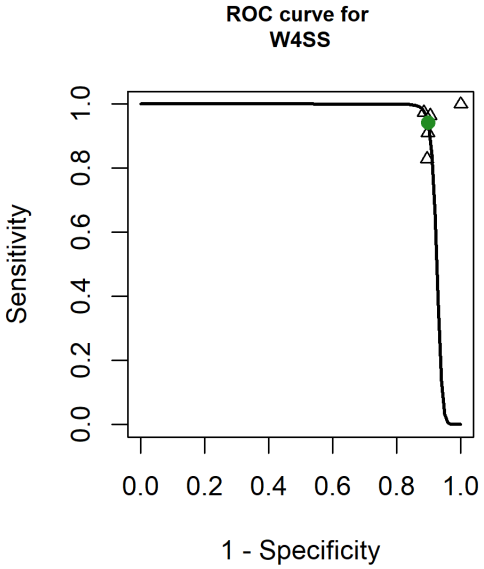
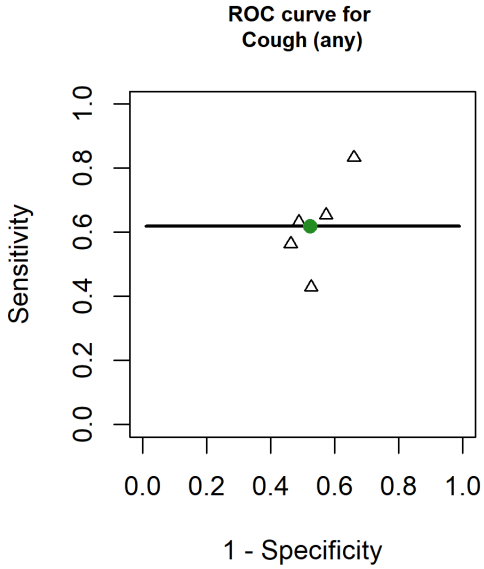
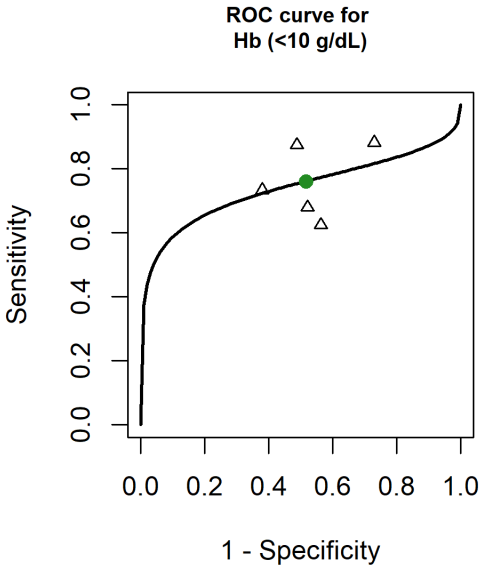


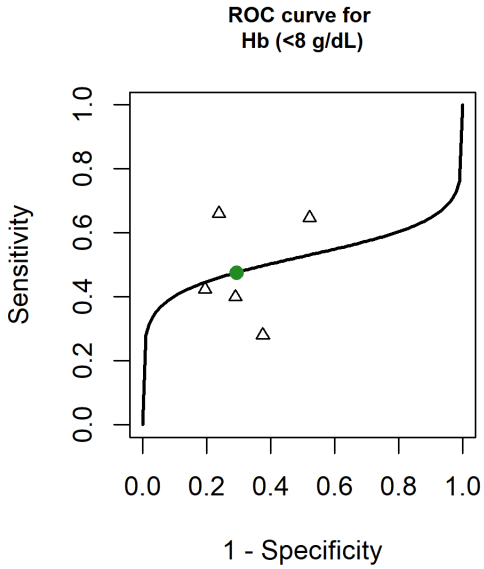
Figure S4 - Summary receiver operating characteristics curves (for tests/strategies with >=4 studies available)

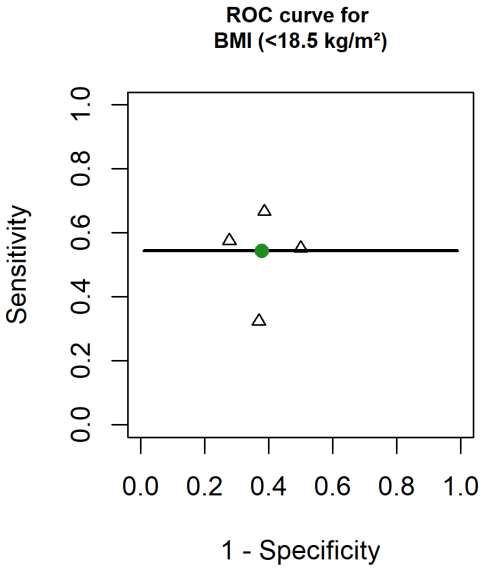
Figure S4A - Summary receiver operating characteristics curves for each screening test/strategy using AlereLAM as a reference standard

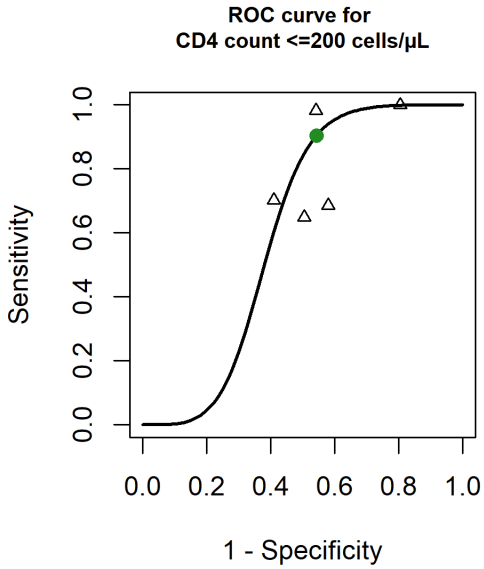





ROC curve for
Hb (<10 g/dL)
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity


ROC curve for
Hb (<8 g/dL)
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity


ROC curve for
BMI (<18.5 kg/m²)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


ROC curve for
CD4 count <=200 cells/µL
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

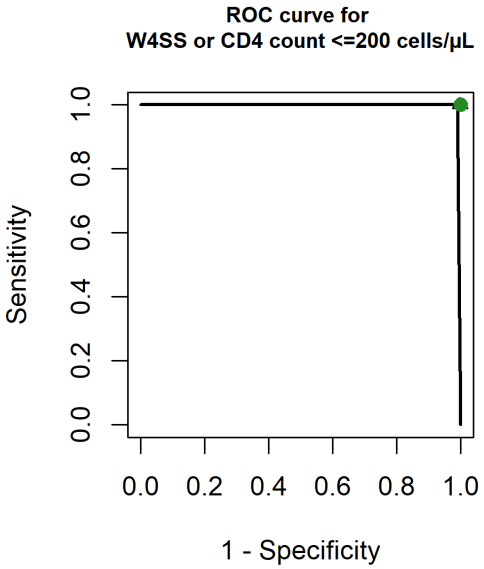
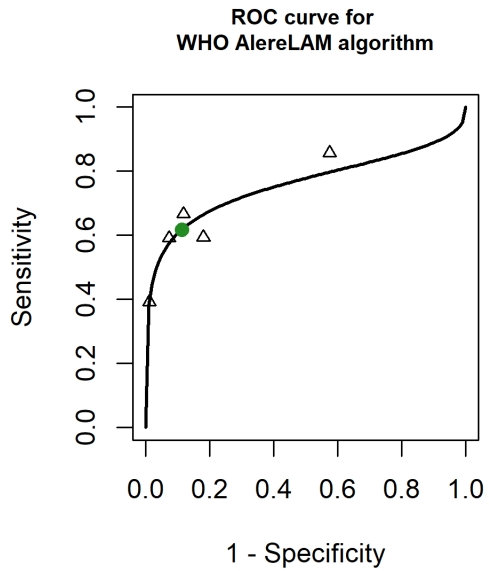

ROC curve for
W4SS or CD4 count <=200 cells/µL
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity

Figure S4B - Summary receiver operating characteristics curves for each LF-LAM strategy using culture or Xpert as a reference standard



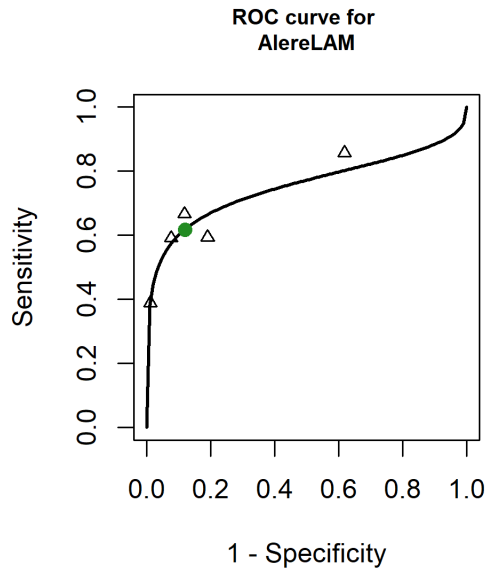
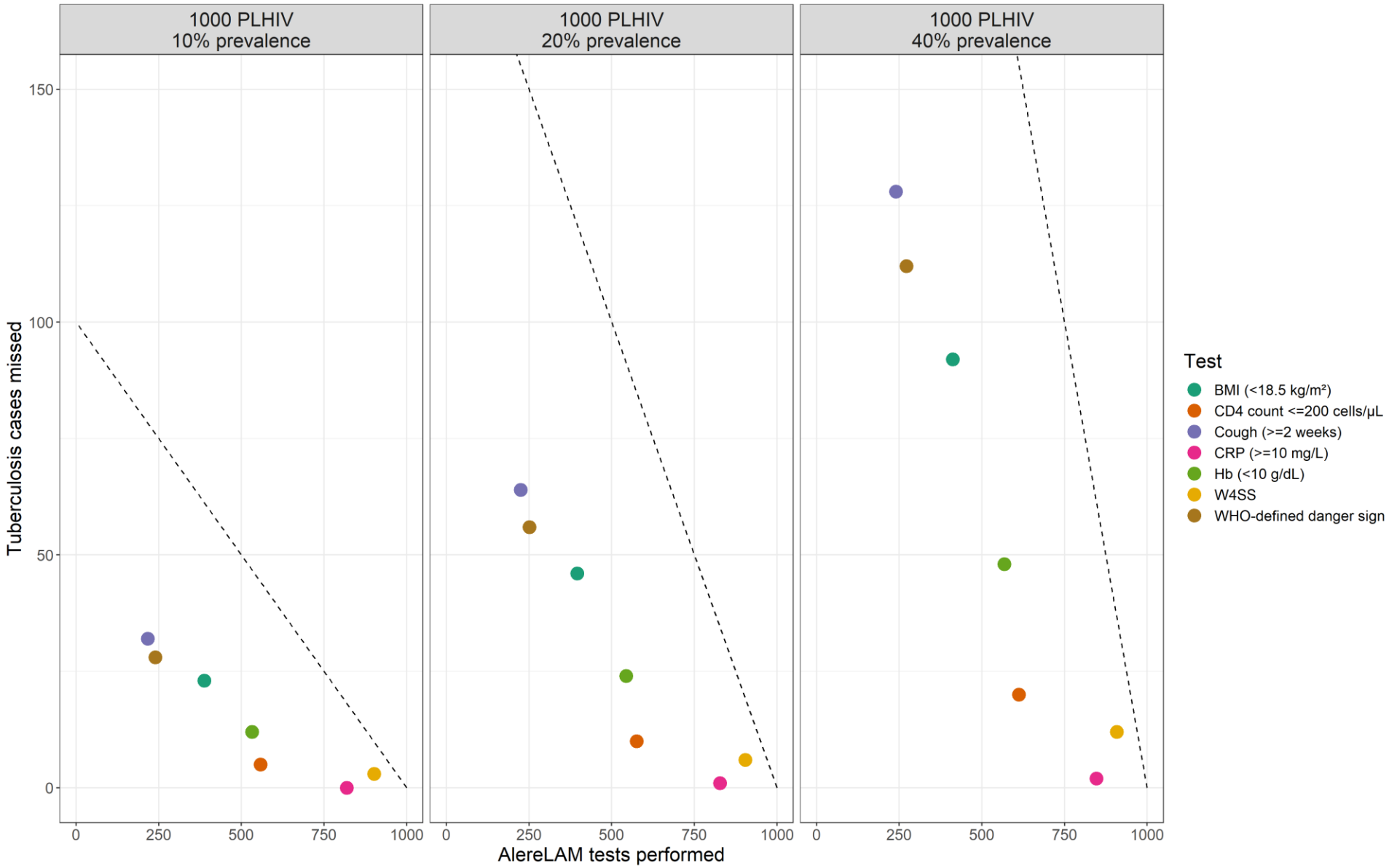


Figure S5 - Plot comparing number of AlereLAM positive tuberculosis cases missed with number of AlereLAM tests required for different tuberculosis screening tests when screening a population of 1000 persons†


1000 PLHIV
10% prevalence
1000 PLHIV
20% prevalence
1000 PLHIV
40% prevalence
Tuberculosis cases missed
AlereLAM tests performed
0
50
100
150
0
250
500
750
1000
Test
BMI (<18.5 kg/m²)
CD4 count <=200 cells/µL
Cough (>=2 weeks)
CRP (>=10 mg/L)
Hb (<10 g/dL)
W4SS
WHO-defined danger sign

†Using a reference standard of AlereLAM. The dashed line represents the number of AlereLAM positive tuberculosis cases diagnosed when applying x AlereLAM tests at random among 1000 PLHIV. Tests closer to the bottom left corner would offer a better trade-off between tuberculosis cases missed and AlereLAM tests required

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, Hb = haemoglobin, PLHIV = people living with HIV, W4SS = WHO four-symptom screen

References

1. Bjerrum S, Kenu E, Lartey M, et al. Diagnostic accuracy of the rapid urine lipoarabinomannan test for pulmonary tuberculosis among HIV-infected adults in Ghana-findings from the DETECT HIV-TB study. *BMC Infect Dis* 2015; **15**: 407.
2. Gupta-Wright A, Corbett EL, van Oosterhout JJ, et al. Rapid urine-based screening for tuberculosis in HIV-positive patients admitted to hospital in Africa (STAMP): a pragmatic, multicentre, parallel-group, double-blind, randomised controlled trial. *Lancet* 2018; **392**(10144): 292-301.
3. Huerga H, Mathabire Rucker SC, Bastard M, et al. Urine Lipoarabinomannan Testing for All HIV Patients Hospitalized in Medical Wards Identifies a Large Proportion of Patients With Tuberculosis at Risk of Death. *Open Forum Infect Dis* 2021; **8**(2): ofaa639.
4. Lawn SD, Kerkhoff AD, Burton R, et al. Rapid microbiological screening for tuberculosis in HIV-positive patients on the first day of acute hospital admission by systematic testing of urine samples using Xpert MTB/RIF: a prospective cohort in South Africa. *BMC Med* 2015; **13**: 192.
5. Thit SS, Aung NM, Htet ZW, et al. The clinical utility of the urine-based lateral flow lipoarabinomannan assay in HIV-infected adults in Myanmar: an observational study. *BMC Med* 2017; **15**(1): 145.