The evaluation of different strategies to improve the diagnosis of tuberculosis in people living with HIV in resource-limited settings

Ashar Vijay Dhana

MBBCh (WITS) MPH (Harvard) MMed (UCT) SCE-Dermatology (UK) FC Derm (SA)

Thesis presented for the degree of

DOCTOR OF PHILOSOPHY (MEDICINE)

in the Department of Medicine,

Faculty of Health Sciences

University of Cape Town

October 2022

Supervisors:

Professor Gary Maartens

Professor Graeme Meintjes

Dr David Adam Barr

TABLE OF CONTENTS

NOTICE OF COPYRIGHT

DECLARATION

I, Ashar Vijay Dhana, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publications in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publications:

1. **Dhana AV**, Hamada Y, Kengne AP, Kerkhoff AD, Rangaka MX, Kredo T, Baddeley A, Miller C, Singh S, Hanifa Y, Grant AD, Fielding K, Affolabi D, Merle CS, Wachinou AP, Yoon C, Cattamanchi A, Hoffmann CJ, Martinson N, Mbu ET, Sander MS, Balcha TT, Skogmar S, Reeve BWP, Theron G, Ndlangalavu G, Modi S, Cavanaugh J, Swindells S, Chaisson RE, Ahmad Khan F, Howard AA, Wood R, Thit SS, Kyi MM, Hanson J, Drain PK, Shapiro AE, Kufa T, Churchyard G, Nguyen DT, Graviss EA, Bjerrum S, Johansen IS, Gersh JK, Horne DJ, LaCourse SM, Al-Darraji HAA, Kamarulzaman A, Kempker RR, Tukvadze N, Barr DA, Meintjes G, Maartens G.

   "Tuberculosis screening among ambulatory people living with HIV: a systematic review and individual participant data meta-analysis"

   The Lancet Infectious Diseases. 2022 Apr;22(4):507-518

   https://doi.org/10.1016/S1473-3099(21)00387-X
2. **Dhana AV**, Hamada Y, Kengne AP, Kerkhoff AD, Rangaka MX, Kredo T, Baddeley A, Miller C, Gupta-Wright A, Fielding K, Wood R, Huerga H, Rücker SCM, Heidebrecht C, Wilson D, Bjerrum S, Johansen IS, Thit SS, Kyi MM, Hanson J, Barr DA, Meintjes G, Maartens G.

   "Tuberculosis screening among HIV-positive inpatients: a systematic review and individual participant data meta-analysis"

   The Lancet HIV. 2022 Apr;9(4):e233-e241

   https://doi.org/10.1016/S2352-3018(22)00002-9
3. **Dhana AV**, Hamada Y, Kengne AP, Kerkhoff AD, Broger T, Denkinger CM, Rangaka MX, Gupta-Wright A, Fielding K, Wood R, Huerga H, Rücker SCM, Bjerrum S, Johansen IS, Thit SS, Kyi MM, Hanson J, Barr DA, Meintjes G, Maartens G

"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"
The Journal of Infection. Jul;85(1):40-48.
https://doi: 10.1016/j.jinf.2022.05.010.

4. **Dhana AV**, Gupta RK, Hamada Y, Kengne AP, Kerkhoff AD, Yoon C, Cattamanchi A, Reeve BW, Theron G, Wood R, Drain PK, Calderwood CJ, Noursadeghi M, Boyles T, Meintjes G, Maartens G, Barr DA.
"Clinical utility of WHO-recommended screening tools and development and validation of novel clinical prediction models for tuberculosis screening among outpatient people living with HIV: an individual participant data meta-analysis"
Formatted for submission to the European Respiratory Review.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature:

Ashar Vijay Dhana (DHNASH004)

Date: 25th October 2022

ACKNOWLEDGEMENTS

First, I am incredibly grateful to have Prof Gary Maartens as my primary supervisor and mentor. Gary invested an enormous amount of time guiding and supporting me throughout this PhD and always provided insightful feedback and advice. He also submitted the initial proposal that led the World Health Organization to award the University of Cape Town the contract to update the HIV-associated tuberculosis screening guidelines – a project that was a privilege to be a part of.

I am also very grateful to my co-supervisors, Graeme Meintjes and David Barr, for their guidance and support throughout my PhD. I am particularly thankful for their expert feedback during protocol development and manuscript preparation. Graeme provided me with several opportunities to help compete my PhD, and David's statistical expertise played a huge role in helping me to develop analyses plans and run analyses.

I wish to thank the investigators of the primary studies for contributing their data and the patients who participated in these studies. I am grateful to several other people. I thank the core team for their assistance (Yohhei Hamada, Andrew Kerkhoff, Andre Kengne, and Molebogeng Rangaka). Yohhei was instrumental to the completion of these studies, assisting as a second reviewer. I thank Rishi Gupta for his expert advice and for the many hours he set aside to assist with the protocol and analysis of the final paper. I thank the staff at the World Health Organization (Annabel Baddeley, Cecily Miller, Tamara Kredo, and Dennis Falzon) and Tamra Kredo for their expert advice and guidance. I also wish to thank Kathryn Wood for her assistance with research administration.

Finally, I am grateful to my family for all their support and guidance, as without them none of this would have been possible.

ABBREVIATIONS

ACF	Active case-finding
AIDS	Acquired immunodeficiency syndrome
AlereLAM	Alere Determine TB-LAM
ART	Antiretroviral therapy
BMI	Body-mass index
CD4	Cluster of differentiation 4
CI	Confidence interval
C-statistic	Concordance statistic
CPM	Clinical prediction model
CRP	C-reactive protein
CXR	Chest X-ray
DOTS	Directly Observed Treatment, Short-Course
FujiLAM	Fujifilm SILVAMP TB-LAM
Hb	Hemoglobin
HR	Hazard ratio
HIV	Human immunodeficiency virus
ICF	Intensified tuberculosis case-finding
IECV	Internal external cross-validation
IPD	Individual participant data
IPDMA	Individual participant data meta-analysis
IPT	Isoniazid preventative therapy
IQR	Interquartile range
LAM	Lipoarabinomannan
LAMP	Loop-mediated isothermal amplification
LF-LAM	Lateral-flow lipoarabinomannan
LTBI	Latent tuberculosis infection
OR	Odds ratio
PCF	Passive case-finding
PLHIV	People living with HIV
ROC	Receiver-operating characteristic
RR	Risk ratio
TB	Tuberculosis

TPP	Target product profiles
TPT	Tuberculosis preventative therapy
Truenat	Truenat MTB, Truenat MTB Plus, and Truenat MTB-RIF Dx
W4SS	WHO four-symptom screen
WHO	World Health Organization
Xpert	Xpert MTB/RIF
Xpert Ultra	Xpert MTB/RIF Ultra

ABSTRACT

Background

The 2019 WHO screening and diagnostic algorithm for tuberculosis in people living with HIV (PLHIV) has 2 components: the WHO Xpert MTB/RIF (Xpert) algorithm and WHO Alere Determine TB-LAM (AlereLAM) algorithm. According to the WHO Xpert algorithm, WHO recommends that PLHIV be routinely screened for tuberculosis with the WHO four-symptom screen (W4SS; comprising any one of current cough, fever, night sweats, or weight loss) and, if the screen is positive, receive Xpert or Xpert MTB/RIF Ultra (Xpert Ultra) confirmatory testing. According to the WHO AlereLAM algorithm, WHO also recommends that PLHIV be routinely screened for tuberculosis using screening criteria and, if the screen is positive, receive urine lateral-flow lipoarabinomannan (LF-LAM) confirmatory testing with AlereLAM.

We aimed:

i. To determine the diagnostic accuracy of the W4SS and alternative screening tools and strategies in ambulatory PLHIV, including key subgroups, and to compare the diagnostic accuracy of the WHO Xpert algorithm with Xpert confirmatory testing for all ambulatory PLHIV
ii. To determine the performance of the W4SS and alternative screening tools and strategies in HIV-positive inpatients and to compare the diagnostic accuracy of the WHO Xpert algorithm with Xpert confirmatory testing for all HIV-positive inpatients
iii. To determine the performance of WHO screening criteria and alternative screening tools and strategies to guide LF-LAM testing in HIV-positive inpatients and to compare the performance of the WHO AlereLAM algorithm with AlereLAM and Fujifilm SILVAMP TB-LAM (FujiLAM; a novel LF-LAM test) confirmatory testing in all HIV-positive inpatients.
iv. To develop and validate novel clinical prediction models (CPMs) for tuberculosis screening in outpatient PLHIV and to determine the clinical utility of these CPMs and WHO-recommended screening tools

Methods

We conducted a systematic review and individual participant data (IPD) meta-analysis. We updated a search of PubMed (MEDLINE), Embase, Cochrane Library, and conference abstracts for publications from Jan 1, 2011, to March 12, 2018, done in a previous systematic

review to include the period up to August 2, 2019 (objectives i and iv) and March 1, 2020 (objectives ii and iii). We also screened reference lists of identified pieces and contacted experts in the field.

We included prospective cross-sectional studies, observational studies, and randomized trials that enrolled adult and adolescent (age ≥10 years) PLHIV irrespective of symptoms and signs of tuberculosis. We also included studies that enrolled outpatient PLHIV with a positive W4SS (objective iv only). We extracted study-level data using a standardized data extraction form, and we requested IPD from study authors. The reference standards were culture (objectives i, ii, and iv) and culture or Xpert (objective iii). For screening tools and strategies, we also used separate reference standards of Xpert (objective i and ii), AlereLAM (objective iii), and FujiLAM (objective iii). We selected these confirmatory tests as reference standards since these tests are the most likely confirmatory tests used in practice.

We obtained pooled proportion estimates with a random-effects model, assessed diagnostic accuracy (i.e., sensitivity and specificity) by fitting random-effects bivariate models, and assessed diagnostic yield (i.e., proportion of total tuberculosis cases with a positive confirmatory test) descriptively. For CPMs, we first used logistic regression, allowing for non-linear relations, to develop an extended CPM (using backwards selection of C-reactive protein [CRP] and other predictors) and a CRP-only CPM (which only included CRP along with spline transformations); we then used internal-external cross-validation to evaluate discrimination, calibration, and clinical utility (i.e., decision curve analysis) of both CPMs and other screening strategies. Decision curve analysis plots net benefit across a range of risk thresholds. This systematic review has been registered with PROSPERO, CRD42020155895.

Results

i. We obtained data for 22 of 25 studies (n= 15,666 participants; 4,347 on antiretroviral therapy [ART]). W4SS sensitivity was 82% (95% CI 72, 89) and specificity was 42% (29, 57). CRP (≥10 mg/L) had similar sensitivity (77% [61, 88]), but higher specificity (74% [61, 83]; n=3571). Cough (lasting ≥2 weeks), haemoglobin (<10 g/dL), body mass index (<18.5 kg/m²), and lymphadenopathy had high specificities (80–90%) but low sensitivities (29–43%). The WHO Xpert algorithm had a sensitivity of only 58% (50,66) and a specificity of 99% (98, 100); Xpert for all had a sensitivity of 68% (57–76) and similar specificity. In the only study that compared both tests, the

sensitivity of sputum Xpert Ultra was higher than sputum Xpert (73% [62, 81] vs 57% [47, 67]) and specificities were similar.

Among outpatients on ART, W4SS sensitivity was 53% (35, 71) and specificity was 71% (51, 85). In this population, a parallel strategy (two or more screening tests offered at the same time) of W4SS with any chest X-ray abnormality had higher sensitivity (89% [70, 97]) and lower specificity (33% [17, 54]; n=2,670) than W4SS alone; at a 5% tuberculosis prevalence, this strategy would require 379 more Xpert tests per 1,000 PLHIV than W4SS but detect 18 more cases. Among outpatients not on ART, W4SS sensitivity was 85% (76, 91) and specificity was 37% (25, 51). CRP (≥10 mg/L) had a similar sensitivity (83% [79, 86]), but higher specificity (67% [60, 73]; n=3,187) and a sequential strategy (second screening test offered only if first screening test is positive) of W4SS then CRP (≥5 mg/L) also had similar sensitivity (84% [75, 90]) but higher specificity (64% [57, 71]; n=3187); at 10% tuberculosis prevalence, these CRP-based strategies would require 272 and 244 fewer Xpert tests per 1,000 PLHIV than W4SS but miss two and one more cases, respectively.

ii. We obtained data for all six eligible studies (n=3,660 participants). The pooled proportion of inpatients eligible for Xpert was 90% (89, 91; n=3,658). Among screening tools to guide Xpert testing, W4SS and CRP (≥5 mg/L) had highest sensitivities (≥96%) but low specificities (≤12%); cough (≥2 weeks), haemoglobin (<8 g/dL), body mass index (<18.5 kg/m^2), and lymphadenopathy had higher specificities (61–90%) but low sensitivities (12–57%). The WHO Xpert algorithm had sensitivity of 76% (67, 84) and specificity of 93% (88, 96; n=637). Xpert for all had similar accuracy to the WHO Xpert algorithm: sensitivity was 78% (69, 85) and specificity was 93% (87, 96; n=639).

iii. We obtained data from all 5 identified studies (n=3,504). The pooled proportion of inpatients eligible for AlereLAM testing using WHO criteria was 93% (91, 95). Among screening tools to guide LF-LAM testing, WHO criteria, CRP (≥5 mg/L), and CD4 count (<200 cells/ μL) had high sensitivities but low specificities; cough (≥2 weeks), hemoglobin (< 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 (62% [47, 75]) and specificity (88% [64, 97]) as WHO AlereLAM algorithm. Sensitivities of FujiLAM and AlereLAM were 69% and 48%, while specificities were 88% and 96%, respectively. In 2 studies that

collected sputum and non-sputum samples for Xpert and/or culture, diagnostic yield of sputum Xpert was 40–41%, 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 61%, 81%, and 92% of all cases, respectively.

iv. We obtained data from all 6 identified studies (8 cohorts [n=4,315 participants]). The extended CPM had a C-statistic of 0.81; the CRP-only CPM had similar discrimination (C-statistic 0.79). The C-statistics for CRP (≥5 mg/L; 0.70) and W4SS (0.57) were lower. For clinical utility, both CPMs had equivalent or higher net benefit compared with WHO-recommended tools. Compared with both CPMs, CRP (≥5 mg/L) had equivalent net benefit across a clinically useful range of threshold probabilities, while W4SS had lower net benefit. The W4SS would capture 91% of cases and require confirmatory testing for 78% of participants. CRP (≥5 mg/L), the extended CPM (4.2% threshold), and the CRP-only CPM (3.6% threshold) would capture similar percentage of cases but reduce confirmatory tests required by 24%, 27%, and 36%, respectively.

Conclusion

These findings informed the updated 2021 WHO guidelines on tuberculosis screening in PLHIV. Among outpatient PLHIV, the WHO-recommended W4SS has suboptimal diagnostic accuracy and clinical utility. CRP reduces the need for further Xpert confirmatory testing compared with W4SS without compromising sensitivity and has been included in the updated WHO tuberculosis screening guidelines. CRP also shows utility when used in a CPM. However, CRP data were scarce for outpatients on ART, necessitating future research on the accuracy of CRP in this subgroup. Chest X-ray can be useful in outpatients on ART when combined with W4SS. The WHO Xpert algorithm has suboptimal sensitivity; Xpert for all offers slight sensitivity gains and may be considered if resources permit.

Among HIV-positive inpatients, WHO screening criteria and other potential screening tools to guide Xpert and AlereLAM testing have suboptimal performance. Based on these findings, WHO now strongly recommends Xpert testing in all medical HIV-positive inpatients in settings where tuberculosis prevalence is higher than 10%. The findings in this thesis also support that AlereLAM testing be implemented in all HIV-positive medical inpatients.

Routine FujiLAM testing in all HIV-positive medical inpatients may substantially improve tuberculosis diagnosis, but prospective evaluation of this novel assay is required.

LIST OF TABLES

LIST OF FIGURES

STRUCTURE OF THESIS

This thesis synthesises evidence using systematic review, individual participant meta-analysis (IPDMA), and clinical prediction model (CPM) methodology to inform global policy recommendations for tuberculosis screening in people living with HIV (PLHIV).

In Chapter 1, I provide a background for the aims and objectives of this thesis. First, I review the epidemiology of tuberculosis and HIV-associated tuberculosis. Second, I describe the clinical features of HIV-associated tuberculosis. Third, I review diagnosis of tuberculosis in PLHIV, focusing on confirmatory tests and screening tools for HIV-associated tuberculosis. This subsection of the chapter not only summarizes the current literature on screening for and diagnosis of tuberculosis in PLHIV, but also highlights limitations and gaps in evidence. Fourth, I discuss the value of IPDMA with emphasis on its use in diagnostic test accuracy and CPM research. Finally, I summarize the main aim of this thesis along with the accompanying four objectives:

1. To determine the diagnostic accuracy of the WHO four-symptom screen (W4SS) and alternative screening tools in ambulatory PLHIV, including key subgroups
2. To determine the performance of the W4SS and alternative screening tools and strategies in HIV-positive inpatients
3. To determine the performance of WHO screening criteria and alternative screening tools to guide lateral-flow lipoarabinomannan (LF-LAM) testing in HIV-positive inpatients
4. To develop and validate novel CPMs for pulmonary tuberculosis screening in outpatient PLHIV and to determine the clinical utility of these CPMs and WHO-recommended screening tools.

In Chapters 2 to 5, I present the research papers for each of the four objectives. To address these objectives, I collected individual-level data from multiple studies and conducted analyses using an IPDMA framework.

In Chapter 2, I addressed screening strategies for outpatient PLHIV. I show that C-reactive protein (CRP) alone or combined with W4SS reduces the need for further confirmatory testing with Xpert MTB/RIF (Xpert) in outpatients not on antiretroviral therapy (ART) compared with W4SS alone, without compromising sensitivity. Although CRP data in outpatients on ART were scarce, I show that chest X-ray could be combined with the W4SS, depending on available resources, because this strategy detects more cases than does the W4SS alone. I also demonstrate that the WHO Xpert algorithm (W4SS followed by Xpert if

the W4SS is positive) has suboptimal sensitivity in outpatients, and Xpert confirmatory testing for all outpatients (i.e., no use of a screening tool) offers small improvements in sensitivity. These findings informed the updated 2021 WHO guidelines on tuberculosis screening in PLHIV and led to 4 new and updated WHO recommendations. These findings were also presented at the WHO HIV-TB Implementation for Impact meeting, South Africa Tuberculosis Screening Guidelines Task Team Meeting, and 52nd Union World Conference on Lung Health. This work has been published in *The Lancet Infectious Diseases,* and the paper was selected for the 2021 UCT best publication award in the public health category.

Having synthesised evidence on tuberculosis screening in outpatient PLHIV, I then addressed screening strategies for HIV-positive inpatients in Chapters 3 and 4. Since it was anticipated that screening performance would be different in inpatients due to differences in case-mix and prior probability of symptoms, a separate analysis was performed for this subgroup.

In Chapter 3, I show that the W4SS and other potential screening tools or strategies to guide Xpert confirmatory testing have suboptimal diagnostic accuracy in HIV-positive inpatients. Thus, Xpert confirmatory testing should be performed in all HIV-positive inpatients. These findings also highlight the need for more accurate screening tools to guide confirmatory testing in HIV-positive inpatients. This work informed the updated 2021 WHO guidelines on tuberculosis screening in HIV-positive inpatients and led to a new WHO recommendation. These findings were also presented at the WHO HIV-TB Implementation for Impact meeting and 9th Annual UCT Research Day, resulting in the 2021 UCT prize for research from a full-time clinician. This work has been published in *The Lancet HIV*.

In Chapter 4, I show that WHO screening criteria and other potential screening tools or strategies to guide Alere Determine TB-LAM (AlereLAM) confirmatory testing in HIV-positive inpatients have suboptimal diagnostic accuracy. The WHO screening criteria to guide AlereLAM testing may complicate the WHO tuberculosis screening and diagnostic algorithm, potentially serving as a barrier to the widespread use of AlereLAM. These findings support that AlereLAM confirmatory testing be implemented in all HIV-positive inpatients alongside routine Xpert confirmatory testing. However, I also show that a negative Xpert and AlereLAM confirmatory test still does not rule out tuberculosis in this population. Finally, I demonstrate that routine Fujifilm SILVAMP TB-LAM (FujiLAM), a novel confirmatory Lateral-flow lipoarabinomannan (LF-LAM) test, may substantially improve the

diagnosis of tuberculosis in this population. These findings were presented at the 9th Annual UCT Research Day. This work has been published in the *Journal of Infection*.

Next, I revisit screening strategies for outpatient PLHIV, focusing on the use of CPMs as a screening tool and summarizing the clinical utility of screening strategies.

In Chapter 5, I develop and validate novel CPMs (that incorporate CRP) for pulmonary tuberculosis screening in outpatient PLHIV and determine the optimal screening approach by comparing the performance and clinical utility of these novel CPMs with WHO-recommended screening tools, i.e., W4SS and CRP (≥5 mg/L) (at the time of the study WHO had recommended CRP based on the findings of Chapter 3). My findings demonstrate that CRP (≥5 mg/L) shows clinical utility for tuberculosis screening among outpatient PLHIV. I also show that the CRP-based CPMs add value if resources permit more confirmatory tests per diagnosed case or if resources only allow fewer confirmatory tests per diagnosed case. Thus, CRP-based screening strategies set the standard for tuberculosis screening among outpatient PLHIV. Furthermore, I show that a 'confirmatory testing for all' strategy might be considered if a setting is able to offer many confirmatory tests per diagnosed case. Finally, I show that the W4SS has suboptimal performance and utility compared with other screening tools and strategies. This work has been submitted for publication and is under peer review.

Finally, I summarize the findings of the four research papers in Chapter 6. I discuss limitations, implications for global tuberculosis programmes, and identify areas for future research.

FUNDING

I am grateful to have received funding for this PhD from 1) the Fogarty HIV-associated TB Training Program Clinical Science PhD Fellowship awarded through the Fogarty International Center of the US National Institutes of Health (D43 TW010559); 2) the National Research Foundation Doctoral Bursary awarded through the National Research Foundation of South Africa; 3) the UCT Pharmacology Bursary awarded through the Division of Clinical Pharmacology at the University of Cape Town; and 4) the Harry Crossley Clinical Research Fellowship awarded through the Harry Crossley Foundation. The studies described in chapters 2 and 3 were funded by the World Health Organization.

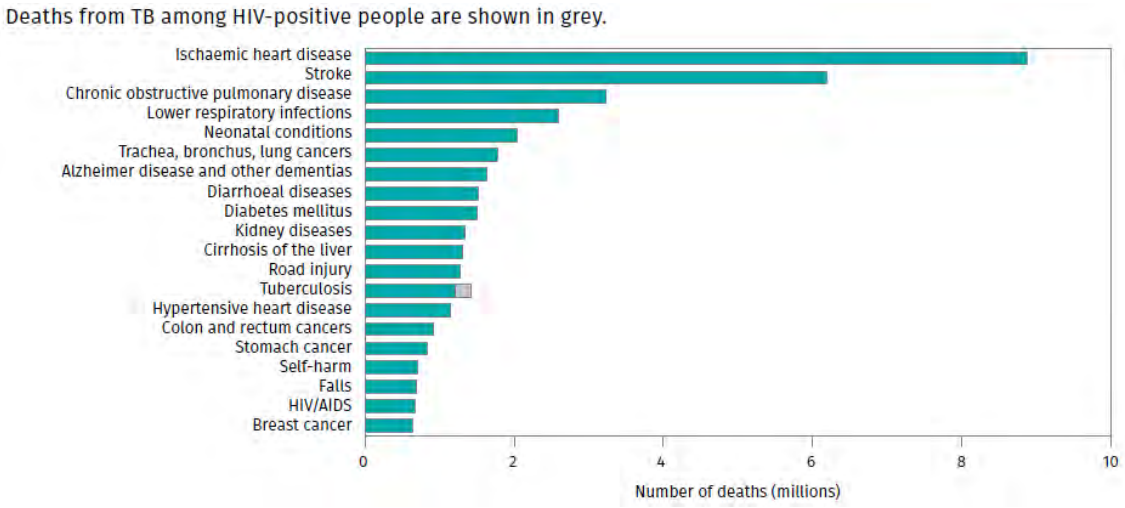
CHAPTER 1: INTRODUCTION

1.1 Tuberculosis epidemiology

1.1.1 Global burden of tuberculosis

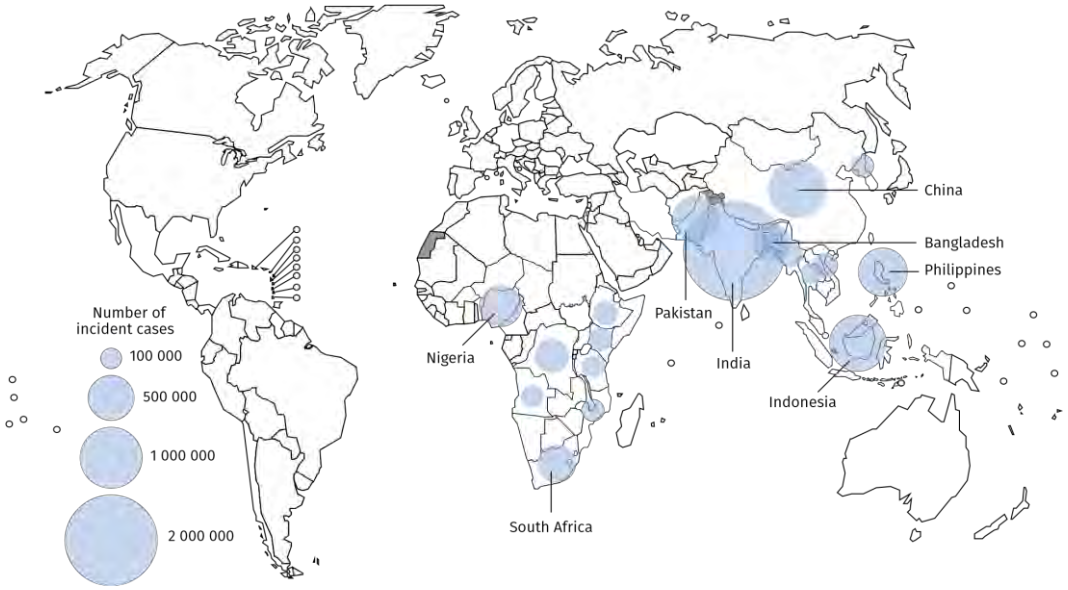
Tuberculosis is a major global health issue. Before the coronavirus pandemic, tuberculosis was the leading infectious cause of death and the 13th leading cause of death worldwide (Figure 1-1).[1] In 2020, there were an estimated 10 million tuberculosis cases and 1.5 million tuberculosis deaths worldwide.[1] Although the global case fatality ratio (i.e., the estimated percentage of people with tuberculosis who die from the disease) has been declining, the global case fatality ratio is still 15%.[1]

Figure 1-1: Leading causes of death in 2019[1]



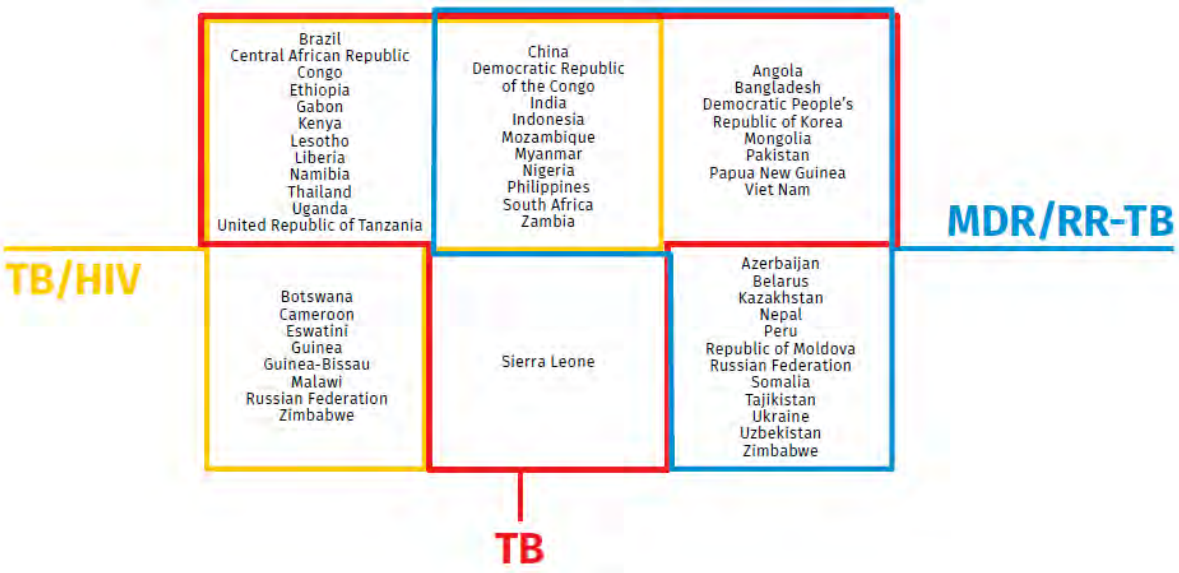
The burden of tuberculosis varies substantially worldwide. Tuberculosis disproportionately affects regions with lower socioeconomic status. Of all tuberculosis cases, an estimated 43% occurred in the WHO South-East Asia Region, 25% in the WHO African Region, and 18% in the WHO Western Pacific Region (Figure 1-2).[1]

Figure 1-2: Tuberculosis incidence by region in 2020[1]



WHO have identified 20 high burden countries for tuberculosis and a further 10 countries with high estimated incidence rate per 100,000 population (Figure 1-3).[1] In 2020, these 30 countries accounted for an estimated 86% of all tuberculosis cases worldwide.[1] Eight countries account for almost 70% of global cases.[1] Lesotho and South Africa are countries with the highest tuberculosis cases per 100,000 population per year.[1]

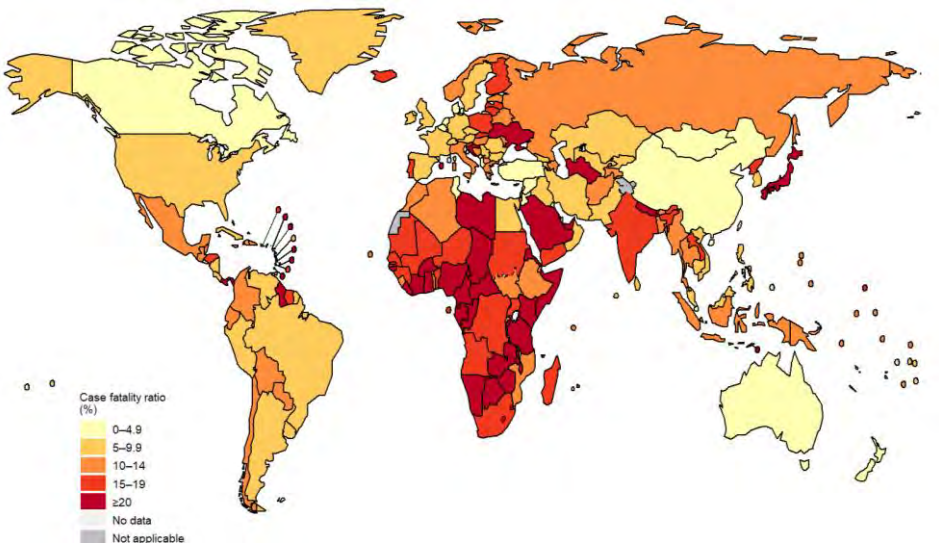
Figure 1-3: Areas of overlap of WHO global lists of high-burden countries for TB, TB/HIV, and MDR/RR-TB for the period 2021 to 2025[1]



TB = tuberculosis, TB/HIV = HIV-associated tuberculosis, and MDR/RR-TB = multidrug-resistant tuberculosis/rifampicin-resistant tuberculosis

In terms of mortality, the tuberculosis case fatality ratio also varies substantially worldwide (Figure 1-4). An estimated 85% of all tuberculosis deaths occurred in the WHO African and WHO South-East Asia regions.[1] Some regions with lower socioeconomic status have a case fatality that is ≥20% .

Figure 1-4: Tuberculosis case fatality ratio by region in 2020[1]



1.1.2 Global response to tuberculosis

The WHO declared tuberculosis a global emergency in 1993. In 1995, WHO responded by recommending that the Directly Observed Treatment Short-Course (DOTS) strategy be implemented worldwide to control the epidemic.[2] In 2000, the Millennium Development Goals were established and included several tuberculosis targets for the year 2015.[3] The targets were to reduce the incidence of tuberculosis each year by an average of 1.5% and to reduce tuberculosis mortality and prevalence by 50% compared with baseline. In 2005, WHO updated the DOTS strategy with the Stop TB Strategy, which was aligned with the Millennium Development Goals and implemented for the period of 2006 to 2015.[4] In 2015, the Millennium Development Goals target to reduce the incidence of tuberculosis was met. The other targets to reduce tuberculosis mortality and prevalence were nearly met; mortality was reduced by 47% while prevalence was reduced by 42%.

In 2016, the Sustainable Development Goals were established and included new targets with the aim to end the global tuberculosis epidemic by 2030.[5] At the same time, WHO updated the Stop TB Strategy with the WHO End TB Strategy for the period 2016 to 2035.[6] The WHO End TB Strategy also aims to end the global tuberculosis epidemic. The WHO End TB

Strategy is aligned with the Sustainable Development Goals. It includes several targets for 2035 such as reducing the number of tuberculosis deaths by 95% and the tuberculosis incidence rate by 90% compared with 2015 (Table 1-1). In 2018, the United Nations held its first ever high-level meeting on tuberculosis.[7] The resulting political declaration by member states not only reaffirmed targets set out on the Sustainable Development Goals and WHO End TB Strategy but also included additional diagnosis, treatment, and funding targets (Table 1-1).

Table 1-1: Targets and milestones as part of the WHO End TB Strategy and the 2018 United Nations political declaration[1,7]

WHO End TB Strategy	**Vision**	**A world free of tuberculosis - zero deaths, disease and suffering due to tuberculosis**			
	Goal	**End the global tuberculosis pandemic**			
		Milestones		**Targets**	
	Indicators	**2020**	**2025**	**2030**	**2035**
	Percentage reduction in the absolute number of tuberculosis deaths (compared with 2015)	35%	75%	90%	95%
	Percentage reduction in the tuberculosis incidence rate (compared with 2015)	20%	50%	80%	90%
	Percentage of tuberculosis-affected households facing catastrophic costs due to tuberculosis (2015 level unknown)	0%	0%	0%	0%
UN high-level meeting on TB, 2018	**Targets**				
	40 million people treated for tuberculosis from 2018 to 2022, including: -3.5 million children -1.5 million people with drug-resistant tuberculosis, including 115 000 children				
	30 million people provided with tuberculosis preventive treatment from 2018 to 2022, including: -6 million people living with HIV -4 million children aged under 5 years and 20 million people in other age groups, who are household contacts of people affected by tuberculosis				
	Funding of at least US$ 13 billion per year for universal access to tuberculosis prevention, diagnosis, treatment, and care by 2022				
	Funding of at least US$ 2 billion per year for tuberculosis research from 2018 to 2022				

The WHO End TB Strategy 2020 interim milestones were to reduce number of tuberculosis deaths by 35% and tuberculosis incidence rate by 20% compared with 2015.[1] However, these milestones were not achieved as mortality was only reduced by 9% while prevalence was reduced by only 11%.[8] Furthermore, although the interim milestone was to have 0% of

tuberculosis-affected households facing catastrophic costs due to tuberculosis, almost half of tuberculosis-affected households still face catastrophic costs due to the disease.[8]

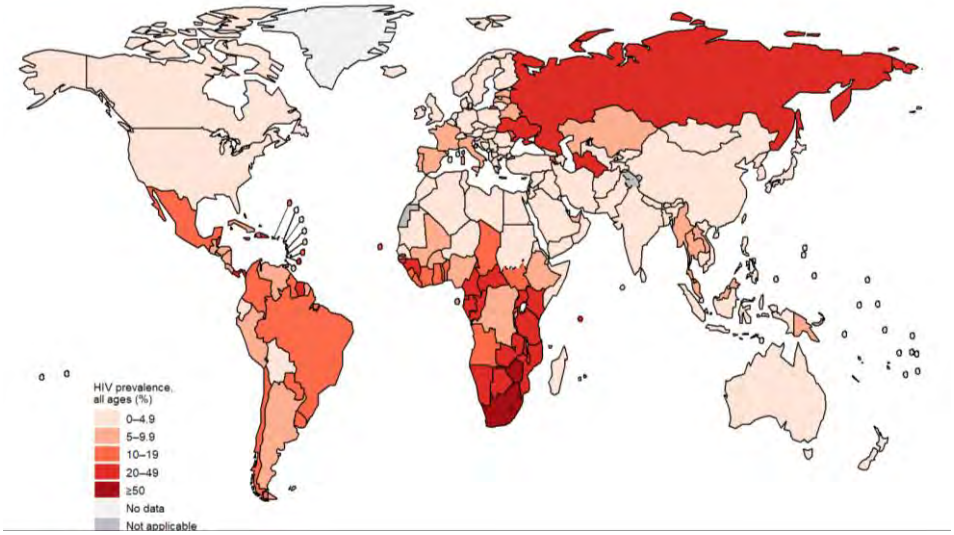
1.1.3 Global burden of HIV-associated tuberculosis

Addressing the burden of tuberculosis in key subgroups, such as PLHIV, is a crucial component of the global response to tuberculosis. HIV is considered the strongest risk factor for tuberculosis. Compared with the general population, PLHIV have a 20 to 37 times increased risk of tuberculosis.[9] PLHIV have an increased risk of both tuberculosis infection and tuberculosis disease with the latter developing from either progression of infection to active disease or reactivation of latent tuberculosis infection.[10] As CD4 cell counts drop and immunosuppression advances, the risk of tuberculosis increases.[11]

In 2020, there were almost 1 million tuberculosis cases and 214,000 tuberculosis deaths among PLHIV.[1] Although HIV-associated tuberculosis accounts for only 8% of all tuberculosis cases worldwide, approximately 15% of all tuberculosis deaths worldwide occur in PLHIV.[1] Indeed, tuberculosis is the leading cause of death in PLHIV.[12] The number of deaths due to tuberculosis in PLHIV is likely underestimated, since tuberculosis often goes undiagnosed in this subpopulation. For example, in facility-based post-mortem studies among PLHIV, tuberculosis was estimated to be the cause of death in 37% of all deaths, but in almost half of all these deaths tuberculosis went undiagnosed at time of death.[12]

Like the global burden of tuberculosis, the global burden of HIV-associated tuberculosis also varies substantially worldwide. The HIV-associated tuberculosis epidemic disproportionately affects the WHO African region, which accounted for an estimated 74% of all cases of HIV-associated tuberculosis in 2020 and which has 23 countries on WHO's global list of 30 countries with a high burden of HIV-associated tuberculosis.[1,13] Within the WHO African region, the burden is especially high in Southern Africa where HIV-associated tuberculosis accounts for more than 50% of all new tuberculosis cases (Figure 1-5).[1] In fact, Lesotho and South Africa are the two countries with the highest number of HIV-associated tuberculosis cases per 100,000 population per year.[1] Of all HIV-associated tuberculosis deaths, around 80% also occurred in the WHO Africa region.[1] The WHO African region has a HIV-associated tuberculosis mortality rate per 100,000 population that is 6 times the global HIV-associated tuberculosis mortality rate per 100,000 population.[1]

Figure 1-5: HIV prevalence in all new and relapse tuberculosis cases by region in 2020[1]



1.2 Clinical features of tuberculosis and HIV-associated tuberculosis

Tuberculosis can affect almost any anatomical site.[14] Pulmonary disease usually presents with cough (which may be chronic) or with haemoptysis,[15] while extrapulmonary tuberculosis has clinical manifestations that depend on anatomical site of disease. The most commonly involved anatomical sites other than the lungs are the lymph nodes, pleura, genitourinary system, musculoskeletal system, central nervous system, gastrointestinal system, and pericardium.[14] All patients may also present with non-specific systemic features such as fever, night sweats, and weight loss.[15]

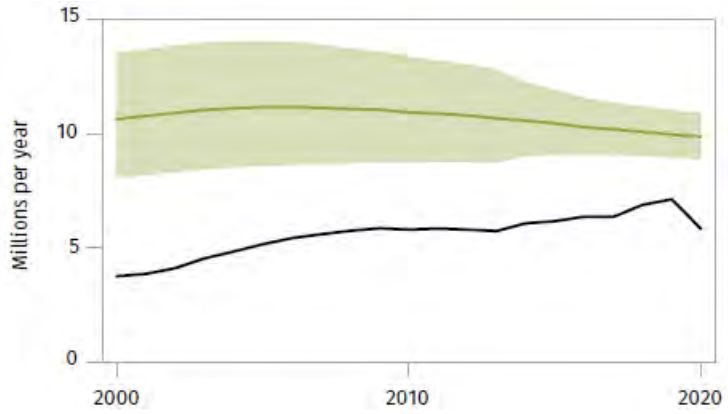
HIV-associated tuberculosis presents variably, depending on the degree of immunosuppression.[16] At higher CD4 cell counts, clinical presentation is like that in HIV-negative individuals. At lower CD4 cell counts, and in HIV-positive inpatients, presentation is often non-specific or atypical, sputum is difficult to produce for diagnostic testing, disease progression is rapid, and extrapulmonary and disseminated disease is common.[17-20] Extra-pulmonary tuberculosis may occur from initial infection at extrapulmonary sites or from tuberculosis that has spread from the lungs where advancing immunosuppression has resulted in uncontrolled disease.[11] On the other hand, disseminated tuberculosis results from continual lymphatic or hematogenous spread and is defined as involvement of two or more non-contiguous sites.[21] In a meta-analysis of facility-based autopsy studies among PLHIV, an estimated 88% of deaths due to HIV-associated tuberculosis were disseminated;[12] the lung, spleen, liver, and lymph nodes were the most common organs involved. Finally, HIV-associated tuberculosis may also present with no symptoms (i.e., subclinical tuberculosis). In

a meta-analysis of PLHIV who were recruited irrespective of tuberculosis symptoms and signs, 11% of PLHIV not on ART and 49% or those on ART had bacteriologically confirmed tuberculosis but were asymptomatic.[22]

1.3 Diagnosis and screening of HIV-associated tuberculosis

In 2020, the gap between the total number tuberculosis cases and number of tuberculosis cases reported was 4.1 million worldwide (Figure 1-6).[1] The gap is a result of both underreporting of people diagnosed with tuberculosis and underdiagnosis of tuberculosis. Although this gap was becoming smaller between 2017 and 2019, the gap widened from 2019 to 2020 because the COVID-19 pandemic resulted in a reduction in the estimated total number of tuberculosis cases reported by 18%.[1] In a meta-analysis of facility-based autopsy studies among PLHIV, tuberculosis was estimated to be the cause of death in 37% of all deaths, but 46% of these tuberculosis cases went undiagnosed at time of death.[12] Therefore, undiagnosed tuberculosis potentially accounts for approximately 20% of all facility-based deaths in PLHIV.

Figure 1-6: Global number of tuberculosis cases reported (black) and estimated number of total tuberculosis cases (green) from 2000 to 2020 (shaded area represents uncertainty interval)[1]



HIV-associated tuberculosis is challenging to diagnose. PLHIV with severe immune suppression typically have disseminated or extrapulmonary tuberculosis, a non-specific clinical presentation, and produce paucibacillary specimens that reduce the accuracy of diagnostic tests.[23,24] Furthermore, a large proportion have difficulty producing sputum for diagnostic testing, especially in certain subgroups such as inpatients.[18,19,24,25]

Strategies as part of the global response to tuberculosis, from the DOTS strategy to the STOP TB Strategy and now the WHO End TB Strategy, have placed a strong emphasis on early and accurate diagnosis of HIV-associated tuberculosis.[26] The current WHO End TB Strategy includes 10 components housed within 3 strategic pillars: 1) integrated, patient-centred tuberculosis care and prevention; 2) bold policies and supportive systems; and 3) intensified research and innovation (Table 1-2).[26] The first pillar has 4 components. The first component is early diagnosis of tuberculosis, including systematic screening of high-risk groups such as PLHIV. The third component is collaborative tuberculosis/HIV activities and management of co-morbidities. For this component, WHO has developed a 12-point package on collaborative TB/HIV activities.[27] One goal of the package is to reduce the burden of tuberculosis in PLHIV using the "Three I's" strategy, which involves intensified tuberculosis case-finding (ICF; which includes systematic screening) and high-quality treatment; isoniazid preventative therapy (IPT) and early ART; and tuberculosis infection, prevention, and control. Effective tuberculosis screening and diagnosis are crucial to ensure effective implementation of each of the "Three I's". The third pillar, which includes 2 components, emphasizes intensified research and innovation, including new tools and strategies to improve screening for and diagnosis of tuberculosis.

Table 1-2: Pillars and components of the WHO End TB Strategy[6]

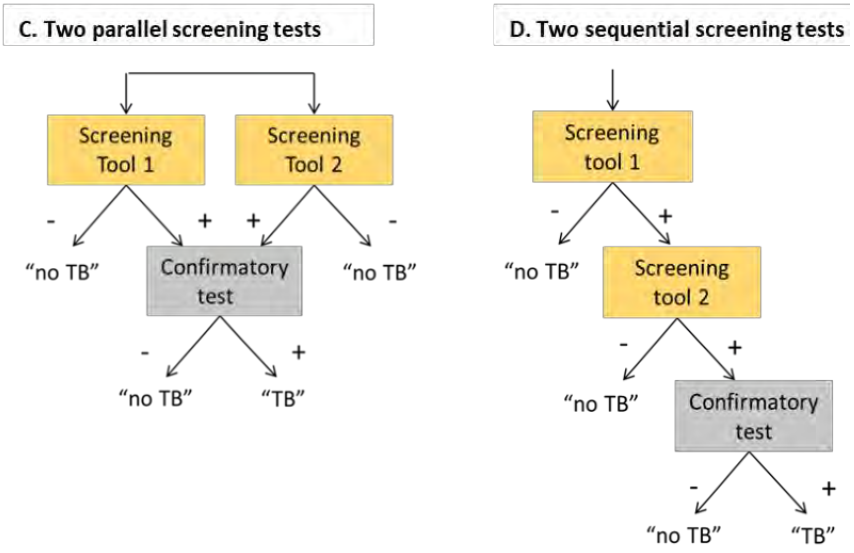
Pillar 1 - Integrated, patient-centred care and prevention
1. Early diagnosis of tuberculosis including universal drug-susceptibility testing, and systematic screening of contacts and high-risk groups
2. Treatment of all people with tuberculosis including drug-resistant tuberculosis, and patient support
3. Collaborative tuberculosis and HIV activities, and management of comorbidities
4. Preventive treatment of persons at high risk, and vaccination against tuberculosis
Pillar 2 - Bold policies and supportive systems
1. Political commitment with adequate resources for tuberculosis care and prevention
2. Engagement of communities, civil society organizations, and public and private care providers
3. Universal health coverage policy, and regulatory frameworks for case notification, vital registration, quality and rational use of medicines, and infection control
4. Social protection, poverty alleviation and actions on other determinants of tuberculosis
Pillar 3 - Intensified research and innovation
1. Discovery, development and rapid uptake of new tools, interventions and strategies
2. Research to optimize implementation and impact, and promote innovations

In general, there are two approaches to diagnosis of tuberculosis: active case-finding (ACF) and passive case-finding (PCF). PCF involves a person with symptoms suggestive of tuberculosis seeking care and a health worker who is able to correctly identify that the symptoms may be a result of tuberculosis and that the person requires diagnostic evealutation.[28,29] PCF is mainly patient-initiated.[28]

ACF is an alternative approach. ACF is often used synonymously with systematic screening and is mainly provider initiated. Systematic screening for tuberculosis is defined as "the systematic identification of people at risk for tuberculosis disease, in a predetermined target group, by assessing symptoms and using tests, examinations, or other procedures that can be applied rapidly".[28] The aims of screening programmes for both individuals and health systems are to reduce mortality, incidence, and severity, as well as to increase choice by identifying disease earlier when more treatment options are available.[30] Screening also has greater economic benefits. For example, screening may reduce costs by preventing disability.[30] Conversely, screening can lead to harms. False positive results can lead to anxiety and individual complications from further investigations.[30] Screening may also be resource intensive, particularly when there are many false positive results.[30] False negative results may also affect the health system by resulting in legal claims. Furthermore, screening can also result in overdiagnosis and overtreatment.[30]

ACF involves identifying those who need confirmatory tuberculosis testing (i.e., those who screen positive) from those who do not (i.e., those who screen negative). Thus, ACF for tuberculosis involves two steps: screening followed confirmatory testing (and other diagnostic evaluation) for those with a positive screen.[28] A single screening tool may be used or screening tools may be combined as a parallel strategy (e.g., two or more screening tests offered at the same time) to improve sensitivity or as a sequential strategy (e.g., second screening test offered only if first screening test is positive) to improve specificity (Table 1-7).[31]

Figure 1-7: Screening algorithms that combine screening tools in a parallel or sequential strategy[31]



ACF can be performed for the whole population or for high-risk groups, such as PLHIV, and can be performed in individuals who seek care (with or without tuberculosis symptoms and signs) or in individuals who do not seek care (e.g., community-based screening).[28] The goal of ACF is to identify more PLHIV with tuberculosis at an earlier stage to reduce morbidity, mortality, and community transmission and to identify PLHIV without tuberculosis who would benefit from IPT/tuberculosis preventative therapy (TPT).[27,32]

The 2011 WHO guidelines for ICF and IPT/TPT in PLHIV recommends that PLHIV be systematically screened for active tuberculosis at each visit to a health facility.[27] The 2013 WHO guidelines on systematic screening for active tuberculosis in PLHIV reaffirm this recommendation.[28] The WHO screening and diagnostic algorithm for tuberculosis in PLHIV has 2 components: the WHO Xpert algorithm and the WHO AlereLAM algorithm (Table 1-

3). According to the WHO Xpert algorithm, WHO recommends screening PLHIV at each encounter for the presence of a positive WHO four symptom screen (W4SS) (comprising any one of current cough, fever, night sweats, or weight loss). In those with a positive screen, Xpert MTB/RIF (Xpert) or Xpert MTB/RIF Ultra (Xpert Ultra) confirmatory testing should be performed. According to the WHO AlereLAM algorithm, WHO recommends also screening for the presence of a positive W4SS, CD4 cell count ≤200 cells/μL (in inpatients) or CD4 cell count ≤100 cells/μL (in outpatients), WHO stage 3 or 4, or positive WHO-defined danger sign. In those with a positive screen, Alere Determine TB-LAM (AlereLAM) confirmatory testing should be performed. In those with a negative screen using both algorithms, IPT/TPT should be provided.

Table 1-3 Summary of initial steps in 2019 WHO tuberculosis screening and diagnostic algorithm for PLHIV[33]

WHO Xpert algorithm
Assess for tuberculosis signs and symptoms* → if positive, perform confirmatory testing with Xpert Ultra
WHO AlereLAM algorithm
Assess for tuberculosis signs and symptoms*, CD4 cell count ≤ 200 cells/μL (inpatients) or CD4 cell count ≤ 100 cells/μL (outpatients), WHO stage 3 or 4, or a WHO-defined danger sign (i.e., meets seriously ill criteria) † → if any positive, perform urine AlereLAM

*Using the WHO four symptom screen, defined as any one of current cough, fever, night sweats, or weight loss.

†WHO-defined danger signs are respiratory rate >30 breaths/min, body temperature >39°C, heart rate >120 beats/min, or unable to walk unaided

The following sections summarize confirmatory tests and screening tools that may be used for active tuberculosis in PLHIV.

1.4 Confirmatory tests for active tuberculosis

Diagnostic tools for tuberculosis include confirmatory tests and non-confirmatory tests. Confirmatory tests usually involve the use of microbiological tests to identify *Mycobacterium tuberculosis* in a specimen (e.g., Xpert or culture). Non-confirmatory tests are typically advised for use later in WHO screening and diagnostic algorithms when initial confirmatory tests are negative and include response to broad-spectrum antibiotics, clinical assessment, and radiological tests (e.g., chest X-ray and abdominal ultrasound).[34] The following section

focuses on confirmatory tests for tuberculosis, which are advised earlier in WHO screening and diagnostic algorithms.

1.4.1 Smear microscopy

Smear microscopy involves identifying acid-fast bacilli and has been used since the 19th century to diagnose tuberculosis. It can be performed at more decentralized levels of care since it is inexpensive, rapid, and requires minimal technical expertise. However, sensitivity in PLHIV is suboptimal even with light-emitting diode fluorescence microscopy, which is 10% more sensitive than conventional microscopy.[35,36] Specificity may also be affected by the presence of non-tuberculous mycobacteria.[36] Furthermore, microscopy is unable to identify drug resistance. Although sensitivity is low, microscopy is still a commonly used confirmatory test in resource-limited settings.

1.4.2 Culture

Mycobacterial culture is the gold-standard for confirmatory diagnosis of tuberculosis.[36] It allows for mycobacterial speciation and drug susceptibility testing. Solid culture has been increasingly replaced by liquid culture, which is faster and has higher sensitivity.[37] However, liquid culture can still take weeks for a result, is more prone to contamination, requires specialist laboratory facilities and highly trained staff, and is expensive.[36] Although, culture plays a limited role in tuberculosis diagnosis, particularly in resource-limited settings, it is commonly used for drug susceptibility testing.

1.4.3 Xpert MTB/RIF and Xpert MTB/RIF Ultra

In 2010, WHO recommended the Xpert assay, a molecular diagnostic test, to replace microscopy as the first-line confirmatory test for pulmonary tuberculosis.[38] A polymerase chain reaction test, Xpert is a semi-automated cartridge-based system on the GeneXpert platform that can detect tuberculosis in less than 2 hours along with the presence of rifampicin-resistance.[35] The polymerase chain reaction targets an 81-bp region of the *rpoB* gene of the bacillus.[36]

In a Cochrane systematic review, Xpert showed good accuracy for the diagnosis of pulmonary tuberculosis in PLHIV with a sensitivity of 79% (95% CI 70, 86) and specificity of 99% (98, 99).[39] For the diagnosis of extrapulmonary tuberculosis, specificity is high for most anatomical sites, but sensitivity is generally lower than for pulmonary tuberculosis and varies by anatomical site.[40] In 2013, WHO also recommended Xpert for use on

extrapulmonary samples.[22] Xpert had also been shown to improve health outcomes in PLHIV. In an individual participant data meta-analysis (IPDMA) of randomized-controlled trials, Xpert reduced mortality compared with microscopy among PLHIV (HR 0.76 [0.60, 0.97]).[41]

In 2017, WHO recommended Xpert Ultra, which has an updated cartridge, as a confirmatory test.[42] Compared with Xpert, Xpert Ultra cartridges have a larger chamber for DNA amplification and, in addition to the *rpoB* target, two multicopy amplification targets (*IS6110* and *IS1081)*.[35] Xpert Ultra can therefore diagnose tuberculosis from specimens with fewer bacilli and has higher sensitivity compared with Xpert. In a Cochrane systematic review, for the diagnosis of pulmonary tuberculosis in PLHIV, the sensitivity was 88% (75, 94) for sputum Xpert Ultra and 75% (59, 86) for sputum Xpert.[43] However, specificity of sputum Xpert Ultra was 93% (82, 97), which was lower than the specificity of sputum Xpert (100% [99, 100]). The lower specificity may be a result of a higher number of false positive tests due to the detection of non-viable bacilli. The lower specificity is a challenge in settings with a high HIV burden and where a history of tuberculosis is common.[44] For the diagnosis of extrapulmonary tuberculosis, the sensitivity of Xpert Ultra – like the sensitivity of Xpert – is generally lower than for pulmonary tuberculosis and varies by anatomical site.[45]

Urine Xpert may be useful for the diagnosis of tuberculosis, especially in those unable to produce sputum. In outpatient PLHIV who were recruited irrespective of tuberculosis symptoms and signs, the sensitivity of urine Xpert for the diagnosis of pulmonary tuberculosis was only 19% (11, 29) but increased to 44% (22, 69) in those with a CD4 cell count ≤ 50 cells/μL.[46] In HIV-positive inpatients with clinically suspected tuberculosis, sensitivity of urine Xpert for the diagnosis of pulmonary and extrapulmonary tuberculosis was 48% (39, 57), which was the same as that of AlereLAM.[47] In HIV-positive inpatients who were recruited irrespective of tuberculosis symptoms and signs, yield of urine Xpert (i.e., proportion of total tuberculosis cases with a positive test) was 59%, but only 27% and 38% for sputum Xpert and AlereLAM, respectively.[25] However, in another study of HIV-positive inpatients who were recruited irrespective of tuberculosis symptoms and signs, yield of urine Xpert was only 35%, while yield of sputum Xpert (40%) and AlereLAM (75%) were higher.[18] Urine Xpert Ultra may show improved sensitivity over urine Xpert, but data is limited. In a recent study of inpatient and outpatient PLHIV undergoing evaluation for pulmonary tuberculosis, sensitivity of urine Xpert Ultra was double that of AlereLAM (33% vs 16%).[48]

Xpert has limitations. Xpert positivity depends on mycobacterial load,[49] meaning that its sensitivity may be lower in screening settings where PLHIV may be relatively well. Although Xpert increases the proportion of treated participants with bacteriologically confirmed tuberculosis, a large proportion are still treated without bacterial confirmation.[50] Xpert can provide a result in <2 hours, but in the real world a result takes several days and sometimes even longer than microscopy.[51] Furthermore, Xpert is not a true point-of-care test since it has significant infrastructure, maintenance, and technical needs,[52] meaning it is generally implemented at district or subdistrict levels.[35] Xpert is also relatively expensive for resource-limited settings. The cost is subsidized at $10 per cartridge in the public sector and much higher in the private sector, a place where up to 60% of all tuberculosis patients initiate care.[53]

Scale-up and implementation of Xpert has been slow in resource-limited settings. Its availability and accessibility are particularly low at the primary health care level because of infrastructure and technical needs. In a 2016 survey of countries with a high HIV-associated tuberculosis burden, investigators found that Xpert was broadly available at tertiary level, but it was only available in 35% and 50% of those countries at the primary health care and district levels, respectively.[54] Furthermore, an observational cohort of tuberculosis patients from 18 countries evaluated implementation of Xpert at hospital and primary care facilities from 2012-2016.[55] Although 63% of facilities had access to Xpert, only 4% of all PLHIV diagnosed with tuberculosis were tested using Xpert. Despite its poor sensitivity, smear microscopy is often still used as a confirmatory test for tuberculosis.

1.4.4 Other molecular-based tests

There are limited data on the use of other molecular-based tests in PLHIV, especially when used in a screening setting.

Truenat assays (Truenat MTB, Truenat MTB Plus, and Truenat MTB-RIF Dx) are molecular-based assays that are chip-based and use a real-time micro-polymerase chain reaction to detect tuberculosis, as well as rifampicin resistance, in 1 hour.[56] In a large multicentre study of 1807 participants with signs and symptoms of pulmonary tuberculosis, Truenat MTB and Truenat MTB Plus had sensitivities of 73% (67, 78) and 80% (75, 84) and specificities of 98% (97, 99) and 96% (95, 97), respectively.[57] However, only 48 participants were PLHIV. In 2020, WHO recommended Truenat assays for the diagnosis of pulmonary tuberculosis in adults with symptoms and signs of tuberculosis.[58] Although there was insufficient data in

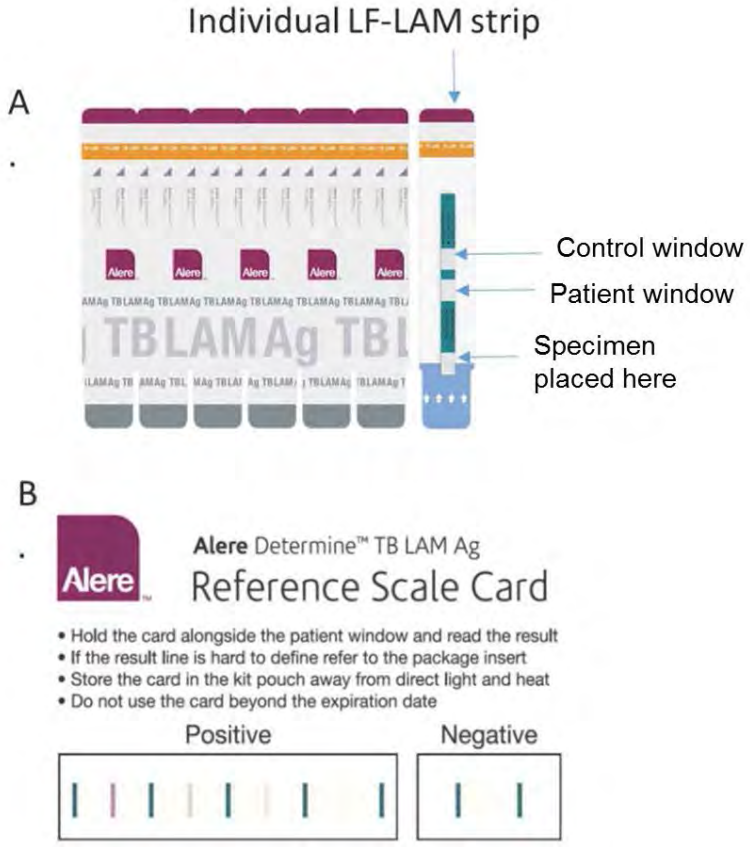
PLHIV, the recommendation was extended to include PLHIV because of indirect data on diagnostic accuracy in smear negative individuals.[59]

Loop-mediated isothermal amplification (LAMP) is an isothermal DNA amplification technique.[60] TB-LAMP is a molecular-based test that uses LAMP. It is a manually performed assay that can be read with the naked eye under ultraviolet light and provides a result in less than an hour.[58] In 2016, WHO recommended that TB-LAMP may be used as a replacement for microscopy in adults with symptoms and signs consistent with tuberculosis.[58] However, this recommendation does not extend to PLHIV because of suboptimal sensitivity and inability to detect rifampicin resitance.[61] In a systematic review and meta-analysis of 4 studies (271 PLHIV with symptoms and signs of tuberculosis), the sensitivity of TB-LAMP was similar to the sensitivity of microscopy (64% vs 62%) for the diagnosis of pulmonary tuberculosis.[62]

1.4.5 Urine lipoarabinomannan-based tests

Lipoarabinomannan (LAM) is a cell wall lipopolysaccharide found in mycobacteria. LAM is released from mycobacteria and is subsequently filtered by the kidney and can therefore be detected in urine.[63] The first commercially available urine LAM test was the laboratory-based urine Clearview TB-ELISA.[64] Subsequently, the AlereLAM assay, a urine point-of-care lateral-flow LAM (LF-LAM) assay, was developed as a confirmatory test for LAM. The AlereLAM assay is currently the only commercially available LF-LAM assay. It costs only US $3.50, gives a result in only 25 minutes or less, poses minimal biohazard risk, and is easy to perform since the test requires minimal technical expertise and urine is readily available and easy to collect (Figure 1-8). In 2015, WHO recommended the use of AlereLAM in symptomatic inpatient or outpatient PLHIV who have a CD4 cell count ≤100 cells/μL or who are ‘seriously ill’ (i.e., respiratory rate > 30/minute, temperature >39°C, heart rate > 120/minute, or unable to walk unaided).[65] In 2019, WHO broadened its recommendation for inpatient PLHIV to those who are symptomatic, have a CD4 cell count ≤ 200 cells/μL, have advanced HIV disease, or who are ‘seriously ill’.[66]

Figure 1-8: AlereLAM test strip (A). Urine is applied to the test strip and the result is read 25 minutes later. Reference card (B). The reference card is used to determine if a test is positive and to "grade" the test result.[67]



WHO has only recommended AlereLAM in subgroups, because AlereLAM has suboptimal sensitivity in all PLHIV. Sensitivity of AlereLAM is highly dependent on setting and degree of immunodeficiency (Table 1-4). In a 2019 Cochrane systematic review, the pooled sensitivity of AlereLAM among outpatient PLHIV irrespective of tuberculosis symptoms and signs was only 31% (18, 47) while specificity was 95% (87, 99);[33] in symptomatic outpatient PLHIV, sensitivity was similar (29% [17, 47]). The sensitivity of AlereLAM was higher in inpatient PLHIV irrespective of tuberculosis symptoms and signs (62% [41, 83]) and in symptomatic HIV-positive inpatients (52% [40, 64]), but specificity was only 84% (48, 96) and 87% (78, 93), respectively. Among symptomatic PLHIV, the sensitivity in those with CD4 cell count >100 cells/µL and CD4 cell count >200 cells/mm was only 17% (10, 27) and 16% (8, 31), respectively, but higher in those with CD4 cell count ≤100 cells/µL (54% [38, 69) and CD4 cell count ≤200 cells/µL (45% [31, 61]), respectively. Data on diagnostic accuracy by CD4 cell count among PLHIV irrespective of tuberculosis symptoms and signs was limited.

Table 1-4: Pooled sensitivity and specificity of AlereLAM for the diagnosis of tuberculosis in PLHIV overall, as well as by setting and CD4 cell count[33]

Setting	Symptomatic PLHIV		PLHIV irrespective of tuberculosis symptoms and signs	
	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)
Overall accuracy				
All participants	42% (31%, 55%)	91% (85%, 95%)	35% (22%, 50%)	95% (89%, 98%)
By setting				
Inpatients	52% (40%, 64%)	87% (78%, 93%)	62% (41%, 83%)	84% (48%, 96%)
Outpatients	29% (17%, 47%)	96% (91%, 99%)	31% (18%, 47%)	95% (87%, 99%)
By CD4 cell count				
CD4 > 200	16% (8%, 31%)	94% (81%, 97%)	Not applicable	Not applicable
CD4 ≤ 200	45% (31%, 61%)	89% (77%, 94%)	26% (9%, 56%)	96% (87%, 98%)
CD4 > 100	17% (10%, 27%)	95% (89%, 98%)	20% (10%, 35%)	98% (95%, 99%)
CD4 ≤ 100	54% (38%, 69%)	88% (77%, 94%)	47% (40%, 64%)	90% (77%, 96%)
CD4 101-199	24% (14%, 38%)	90% (77%, 96%)	Not applicable	Not applicable

The higher sensitivity in inpatients and those with more advanced immunodeficiency is likely because these groups have higher mycobacterial burden and higher rates of haematogenous dissemination of tuberculosis with subsequent renal involvement.[68] The low specificity of AlereLAM in some studies is likely a result of an imperfect microbiological reference standard, since most studies did not collect multiple samples for culture and/or Xpert from both pulmonary and extra-pulmonary sites.[69] Disseminated nontuberculous mycobacteria may also reduce specificity but are uncommon.[70,71]

Although AlereLAM has lower sensitivity compared with sputum Xpert,[67] diagnostic yield (i.e., proportion of total tuberculosis cases with a positive confirmatory test) is higher in some populations, such as inpatients, because urine is more readily available. For example, in 2 cohorts of HIV-positive inpatients who were enrolled regardless of tuberculosis symptoms and signs, only 57% and 63% of inpatients were able to produce sputum for confirmatory testing, respectively, while >99% were able to produce urine for AlereLAM testing.[18,25] In 1 cohort, sputum Xpert diagnosed 27% of all tuberculosis cases (vs 38% for AlereLAM),[25]

while in the other cohort sputum Xpert diagnosed 40% of all tuberculosis cases (vs 75% for AlereLAM).[18]

AlereLAM rapidly identifies those at high risk of mortality who may benefit from prompt treatment. The risk of mortality in PLHIV with LAM positive tuberculosis was 2.3 times that of PLHIV with LAM negative tuberculosis.[72] Furthermore, two randomised trials have demonstrated a reduction in all-cause mortality among HIV-positive medical inpatients with the use of AlereLAM in addition to routine diagnostics (pooled RR 0.85 [0.76, 0.94]).[18,66,73] One trial assessed HIV-positive medical inpatients irrespective of tuberculosis symptoms and signs,[18] while the other assessed medical inpatients with a positive W4SS (who typically comprise >90% of all HIV-positive medical inpatients).[18,73] In a subgroup analysis, the trial conducted in HIV-positive inpatients irrespective of tuberculosis signs and symptoms found that AlereLAM reduced mortality in 3 pre-specified subgroups: those with a CD4 cell count <100 cells/μL, severe anaemia, and clinically suspected tuberculosis.[18]

Recently, a novel urine-based LF-LAM test has been developed – the Fujifilm SILVAMP TB LAM (FujiLAM).[74] In an IPDMA of 5 studies, the pooled sensitivity was 71% (59, 81) for FujiLAM but only 35% (20, 51) for AlereLAM.[75] Compared with the sensitivity of AlereLAM, the sensitivity of FujiLAM was 28 and 43 percentage points higher in outpatients and inpatients, respectively. FujiLAM showed slightly lower specificity compared with AlereLAM. Since FujiLAM detects lower concentrations of LAM,[76] the higher false positive results with FujiLAM may reflect the inability of the reference standard to correctly classify active tuberculosis at a lower mycobacterial burden. Nontuberculous mycobacteria may also reduce the specificity of FujiLAM but were found in only 4% of participants with a false-positive FujiLAM test.[75] Although studies in the IPDMA used bio-banked urine samples, these samples showed similar results compared with fresh samples.[74] In two recent large multicentre diagnostic accuracy studies, FujiLAM sensitivities were 55% (49, 60) and 60% (51, 69).[77,78] However, accuracy varied significantly by lot number. This variability likely needs to be addressed before FujiLAM can be commercially available.

1.5 Screening tools for active tuberculosis

The aim of a screening tool for tuberculosis in PLHIV is to distinguish those with a higher risk of tuberculosis, who should undergo further confirmatory testing, from those with lower risk of tuberculosis, who should be given IPT/TPT. In 2014, WHO developed target product profiles (TPPs) to provide direction for novel tests for tuberculosis.[79] The high priority areas

were to develop a rapid biomarker-based non-sputum test for tuberculosis diagnosis, a sputum-based test for pulmonary tuberculosis diagnosis at the microscopy-centre level, drug-susceptibility test at the microscopy-centre level, and a screening tool (described as a triage test in the meeting report) to identify people suspected of having tuberculosis who should undergo further confirmatory testing.[79] According to WHO, the minimal performance requirement for such a screening tool is a sensitivity of >90% and specificity of >70% when compared with the confirmatory test, while the optimal performance requirement is a sensitivity of >95% and specificity of >80%.[80] Since the screening tool must aim to miss few cases of tuberculosis (i.e., produce few false negative test results), its sensitivity is prioritized. However, specificity is also important. A screening tool with low specificity will lead to many false-positive results and subsequently many unnecessary follow-up confirmatory tests, which are typically expensive. A screening tool should also meet WHO's operational characteristics; it should be non-sputum based, rapid, inexpensive, and require minimal training and infrastructure needs.[81] The following section focuses on screening tools for tuberculosis in PLHIV.

1.5.1 WHO-recommended four symptom screen

Before 2010, there was no standardized screening tool for tuberculosis in PLHIV. As a result, in 2010, WHO commissioned a systematic review and IPDMA to develop a standardized tuberculosis screening rule for PLHIV in resource-limited settings.[82] The results of the IPDMA were based on 9 studies that systematically collected sputum specimens for ≥1 culture in PLHIV irrespective of symptoms and signs of tuberculosis. The reference standard was culture of any specimen, although most studies collected only sputum specimens, meaning the results were mostly applicable to pulmonary tuberculosis. The final analysis included 8,148 PLHIV not on ART, although only 25% of participants were derived from clinical settings while the remainder were derived from community (59%) and mining (17%) settings. The diagnostic accuracy of 23 different combinations of 5 symptoms were assessed, and the authors selected the most sensitive rule, which became known as the W4SS and comprised four symptoms: current cough, fever, night sweats, and weight loss. If any of these 4 symptoms were present, the screening tool was considered positive, indicating the need for further confirmatory testing. Overall, the sensitivity of the W4SS was 79% (58, 91) and specificity was 50% (29, 70) (Table 1-5).[82] The sensitivity was higher (90%) in clinical settings and lower (67%) in community settings; corresponding specificities by setting were not reported. At a tuberculosis prevalence of 1%, 5%, and 10%, the negative predictive

values were 99.6%, 97.7%, and 95.3%, respectively. The results of the IPDMA led WHO to recommend that PLHIV be regularly screened for tuberculosis with the W4SS.[27] If a PLHIV screens positive using the W4SS, then he or she should undergo confirmatory testing (e.g., Xpert) as part of a diagnostic evaluation for tuberculosis. If a PLHIV screens negative, he or she should be offered IPT/TPT.

Table 1-5: Diagnostic accuracy of W4SS by ART status from the initial[82] and updated[83] meta-analyses

ART status	Number of studies	Sample size	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)
Initial meta-analysis				
PLHIV not on ART	9	8,148	79% (58%, 91%)	50% (29%, 70%)
Updated meta-analysis*				
PLHIV not on ART	16	8,664	89% (83%, 94%)	28% (19%, 40%)
PLHIV on ART	7	4,640	51% (28%, 73%)	71% (48%, 86%)

***Does not include studies from the initial meta-analysis**

In 2018, WHO commissioned a systematic review to update its latent tuberculosis infection (LTBI) guidelines and reviewed the accuracy of the W4SS to rule out tuberculosis prior to initiation of IPT/TPT. Unlike the initial meta-analysis, this study also examined the diagnostic accuracy of the W4SS in PLHIV on ART.[83] The results of the IPDMA were based on 8,664 PLHIV not on ART and 4,640 PLHIV on ART from 18 studies. The sensitivity of the W4SS was lower in those on ART (51% [28, 73]) than in those not on ART (89% [83, 94]) (Table 1-5). The specificity was higher in those on ART (71% [48, 86]) than in those not on ART (28% [19, 40]). At a tuberculosis prevalence of 1%, 5%, and 10%, the negative predictive values among PLHIV on ART were 99.3%, 96.5%, and 92.8%, respectively.

The W4SS has several limitations. It has heterogenous diagnostic accuracy, even when results are stratified by ART status, [82,83] meaning that generalisability of the W4SS is unclear. The subjective nature of the W4SS likely contributes to its heterogenous accuracy. Furthermore, the W4SS has low sensitivity in PLHIV on ART.[83] The W4SS also has low specificity in PLHIV not on ART and HIV-positive inpatients, meaning that the tool may pose a challenge for resource-limited settings, since many PLHIV will screen positive and require unnecessary confirmatory testing with Xpert, which is expensive.[20,82,83] The low specificity also means that IPT/TPT may be delayed. The W4SS is more applicable to pulmonary tuberculosis, since most studies in both meta-analyses did not collect

extrapulmonary samples. [82,83] The W4SS is also more applicable to PLHIV who can produce sputum, since most studies in both meta-analyses excluded those unable to produce sputum. [82,83]

1.5.2 C-reactive protein

C-reactive protein (CRP) may also be useful for tuberculosis screening. CRP is an acute phase reactant that is found in serum and a non-specific marker of inflammation. CRP screening can be done by staff with minimal training, in a POC manner using capillary blood, at a cost of only $2 per test, and with results available in <3 minutes (Figure 1-9).

Figure 1-9: The i-CHROMA POC CRP test. The device automatically calculates and presents CRP results for concentrations ranging from 2.5mg/L to 300mg/L[84]

A 2017 meta-analysis assessed the diagnostic accuracy of CRP for tuberculosis in PLHIV either irrespective of symptoms and signs of tuberculosis or with symptoms suggestive of tuberculosis.[85] Among 936 outpatient PLHIV from 5 studies, the sensitivity of CRP was 93% (85, 97) and specificity was 64% (42, 81). Two out of these 5 studies (n=697) were conducted in PLHIV irrespective of symptoms and signs of tuberculosis; [86,87] sensitivities were 80% to 85% and specificities were 58% to 87%. Among 287 inpatient PLHIV from 3 studies, the sensitivities of CRP were high, ranging from 89% to 100%, but specificities were low, ranging from 0% to 40%. These findings suggested that CRP may be a useful screening tool in PLHIV, particularly in outpatient PLHIV.

Subsequently, in a large study that recruited 1,237 PLHIV initiating ART irrespective of symptoms and signs of tuberculosis, the investigators evaluated the use of CRP for tuberculosis screening, finding a sensitivity of 89% (83, 93) and specificity of 72% (69, 75) at

a 10mg/L cutoff.[88] Therefore, CRP approached the WHO TPP for a screening test of 90% sensitivity and 70% specificity.[79] The W4SS had slightly higher sensitivity (96%) but low specificity (14%), which was almost 58 percentage points lower than that of CRP. At a 5mg/L cut-off, CRP had a sensitivity that was 93% and specificity that was 60%. The investigators also assessed CRP as a screening tool in diagnostic algorithms,[89] finding that CRP, as opposed to W4SS, could reduce the number of patients needing confirmatory Xpert testing by 50%. Furthermore, the resources saved by using CRP as a screening tool could be used to provide confirmatory testing with Xpert, culture, and AlereLAM at a cost per tuberculosis case diagnosed less than that of using W4SS as a screening tool followed by only Xpert confirmatory testing. In another study of 425 PLHIV initiating ART irrespective of symptoms and signs of tuberculosis, the sensitivity of CRP at a 5mg/L cut-off and W4SS was similar (91%) but specificity was significantly higher for CRP (59% vs 37%).[90]

Table 1-6 Studies comparing the diagnostic accuracy of CRP (≥10mg/L) with W4SS in PLHIV irrespective of tuberculosis symptoms and signs

			CRP (10mg/L cutoff)		W4SS	
Study	**Participants**	**Sample size**	**Pooled sensitivity (95% CI)**	**Pooled specificity (95% CI)**	**Pooled sensitivity (95% CI)**	**Pooled specificity (95% CI)**
Lawn, 2013[83,86]	Outpatient PLHIV not on ART	496	85% (75%, 92%)	58% (53%, 62%)	81% (71%, 89%)	33% (29%, 38%)
Yoon, 2018[88]	Outpatient PLHIV not on ART	1237	93% (88%, 96%)	60% (57%, 63%)	96% (91%, 98%)	14% (12%, 17%)
Shapiro, 2018[90]	Outpatient PLHIV not on ART	425	91% (77%, 97%)	59% (53%, 64%)	91% (77%, 97%)	37% (32%, 42%)
Gersh, 2018[91]	Outpatient PLHIV on ART	383	40% (5%, 85%)	79% (75%, 83%)	0% (0%, 52%)	87% (83%, 90%)

CRP also has some limitations. There is limited data on CRP for screening outpatient PLHIV on ART. In 1 recent study (n=382) with only 5 culture positive tuberculosis cases, the sensitivity of CRP (5mg/L cut-off) was 40% (vs 0% for W4SS) and specificity was 79% (vs 87% for W4SS).[91] Most studies that have evaluated CRP have excluded participants with no sputum culture results or not collected extrapulmonary samples. Finally, the availability of POC CRP-based tests is currently limited.

1.5.3 Chest X-ray

Chest X-ray is a useful tool for both tuberculosis screening and diagnosis. In particular, digital chest X-ray is becoming more available. It is less costly, can be done using a portable system, has improved image quality, and is safer because a lower radiation dose is needed.[92] Two chest X-ray classification systems have been used for screening purposes.[93] The first system only distinguishes between any abnormal findings and normal findings. It is aimed to be used by health workers with no specialist background (e.g., medical officers or radiographers). The second system distinguishes between abnormal findings suggestive of tuberculosis and abnormal findings not suggestive of tuberculosis or normal findings. It is aimed to be used by health workers with a specialist background (e.g., pulmonologists or radiologists).

In the 2011 IPDMA that was used to develop the W4SS,[82] the authors also evaluated the accuracy of combing abnormal chest X-ray findings to the W4SS. Data were available for 2,805 PLHIV not on ART from 4 studies. The addition of abnormal chest X-ray findings to the W4SS increased sensitivity by 12% (91% versus 79%) compared with W4SS alone, but reduced specificity by 11% (50% versus 39%) (Table 1-7).[82] Although the authors did not conduct a meta-analysis of the diagnostic accuracy of abnormal chest X-ray findings alone, sensitivities ranged from only 26% to 71% in the 4 individual studies and specificities ranged from 48% 99%.[59,94-96] Thus, the low sensitivity of abnormal chest X-ray findings means that tuberculosis often presents with a normal chest X-ray in PLHIV. The low specificity suggests that other diseases may mimic tuberculosis on a radiograph, as does scarring due to previous tuberculosis disease.

Table 1-7: Diagnostic accuracy of combining abnormal chest X-ray findings to the W4SS by ART status from the initial[82] and updated[83] meta-analyses

ART status	Number of studies	Sample size	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)
Initial meta-analysis				
PLHIV not on ART	4	2,805	91% (67%, 98%)	39% (13%, 73%)
Updated meta-analysis*				
PLHIV not on ART	5	1,801	94% (76%, 99%)	20% (8%, 44%)
PLHIV on ART	2	646	85% (70%, 93%)	30% (26%, 34%)

***Does not include studies from the initial meta-analysis**

In the 2018 WHO commissioned updated systematic review and meta-analysis, the authors examined the diagnostic accuracy of combining abnormal chest X-ray findings to the W4SS

by ART status.[83] For PLHIV not on ART, data were available for 1,801 participants from 5 studies; for PLHIV on ART, data were available for 646 participants from 2 studies. The sensitivity of abnormal chest X-ray findings combined with W4SS was high in those on ART (85% [70, 93]) and in those not on ART (94% [76, 99]), but specificity was low in both those on ART (30% [26, 34]) and those not on ART (20% [8, 44]) (Table 1-7). In PLHIV on ART, the sensitivity of abnormal chest X-ray findings combined with W4SS was higher than W4SS (85% vs 51%), but specificity was lower (30% vs 71%). However, these comparisons were indirect (i.e., based on all studies that assessed at least one test of interest) and are thus prone to bias.

Besides suboptimal sensitivity, chest X-ray has several other limitations. Chest X-ray findings, like clinical findings, vary depending on degree of immunosuppression.[97,98] For example, cavitation - a more typical radiographic finding - is less common at lower CD4 cell counts.[97] Thus, evaluation of chest X-ray findings can be difficult without knowing the CD4 cell count.[97] A normal chest X-ray is also common in those with pulmonary tuberculosis. In an IPDMA of PLHIV not on ART, over a third of culture positive tuberculosis cases - which were predominately sputum culture positive - had a normal chest X-ray.[82] Furthermore, among PLHIV with tuberculosis, several studies have shown that a normal chest X-ray is more common in those with advanced immunodeficiency.[98-100] Chest X-ray also requires trained readers and has variable inter- and intra-reader agreement.[92,101] In a 2016 survey of 14 countries with a high HIV-associated tuberculosis burden, chest X-ray as a screening tool was available in only 14% of primary health care centres, while it was available at district and tertiary level in almost all countries.[54] Chest X-ray also has logistical and operational challenges. Given its negative impact on infrastructure, human resources, and cost, chest X-ray may pose a significant burden in resource-limited settings. Computer-aided detection software may overcome some of these limitations. Although an IPDMA of participants presenting with symptoms of tuberculosis showed that computer-aided detection software can be a high sensitivity rule-out test,[102] sensitivity and specificity were modified by HIV status, and only ~600 PLHIV were included in the analysis. Furthermore, there was no study that enrolled PLHIV irrespective of symptoms and signs of tuberculosis.

1.5.4 Clinical prediction models for tuberculosis screening

Clinical prediction models (CPMs) may also be used as a screening tool to identify PLHIV who are at increased risk of tuberculosis and need further confirmatory testing. CPMs (also

known as clinical prediction rules, risk models or risk scores) predict the risk of disease (e.g., tuberculosis) for a person using the combination of multiple predictors.[103,104] These predictors include clinical information (e.g., age, sex, symptoms, and signs) and laboratory tests (e.g., CD4 cell count and CRP). CPMs require appropriate development and validation to be clinically useful.

The development of CPMs require several steps.[105] Once the dataset is prepared, the CPM is typically developed using multivariable regression methods such as logistic regression. This process requires several decisions: identifying candidate predictors, determining the functional form of predictors, assessing sample size relative to number of model parameters, handling missing data using a suitable method, and identifying an appropriate strategy for predictor selection.[103] A full model approach, in which all candidate predictors are included, is one strategy for predictor selection.[106] This approach reduces overfitting, but requires extensive knowledge about the most relevant predictors. Conversely, an alternative predictor selection approach chooses predictors by backward or forward selection.[106] Forward selection begins with an empty model to which predictors are added, while backward selection begins with all predictors and removes predictors.[106] Backward selection is preferred as all predictors are assessed at the same time. Univariable selection is also commonly used in the literature but is not recommended since important predictors may be rejected.[107]

The validation of a CPM involves internal validation (using data from the same population) and external validation (using data from a different population). At validation, the performance of the CPM is evaluated by determining the model's discrimination and calibration. Discrimination refers to how well the model can differentiate between patients that have active tuberculosis and those that do not.[105] Discrimination is assessed using the concordance statistic (C-statistic). A C-statistic of ≥ 0.7 and ≥ 0.8 is considered as acceptable and excellent performance, respectively.[108] Calibration refers to agreement between expected and observed outcomes.[105] Calibration is assessed using calibration-in-the-large (a value of 0 indicating perfect calibration), calibration slope (a value of 1 indicating perfect calibration), and calibration plots. The clinical utility of a CPM to improve decision making should also be evaluated. Clinical utility can be evaluated using decision curve analysis, which shows the net benefit (i.e., benefit versus harm) over a range of clinically relevant threshold probabilities.[109] The model or test with the greatest net benefit for a particular threshold is considered to have the most clinical value.[109]

The TRIPOD statement was developed to provide recommendations on the development and validation of CPMs.[107] However, despite the availability of the TRIPOD statement,[107] as well as other guidance and methodological frameworks,[104,105] CPMs are still poorly developed and validated.[110] There have been few CPMs developed for tuberculosis screening either in PLHIV irrespective of signs and symptoms of tuberculosis or in PLHIV who have a positive W4SS. [111-116] However, the methodology used to develop and/or validate these CPMs is inadequate.

Auld et al developed a CPM for active tuberculosis in outpatient PLHIV not on ART irrespective of symptoms and signs of tuberculosis.[111] The CPM included W4SS symptoms, sex, smoking status, temperature, body-mass index (BMI), and haemoglobin as predictors. The analysis involved splitting a cohort of PLHIV not on ART enrolled in Botswana by geographic region into development and internal validation datasets. No sample size calculations were provided. Although the development dataset had 189 tuberculosis cases, 15 predictors were assessed including 6 continuous variables that were assessed for nonlinearity, meaning that overfitting of the model was a possibility. In the development dataset, sputum was only collected in those with a positive W4SS (for Xpert testing), and sputum culture was only performed if 4 sputum samples were collected. The final CPM was converted to a simplified score, which included categorization of continuous variables. This approach is known to lead to a loss of discriminative ability. Only the simplified score was externally validated in 3 outpatient cohorts from South Africa. The 3 cohorts differed from the derivation cohort in that one included a high percentage on ART, another included those with low CD4 cell counts, and the third included PLHIV not on ART derived from a background population with high tuberculosis prevalence. The final model showed excellent and acceptable discrimination in the derivation (C-statistic: 0.82) and internally validated (C-statistic: 0.77) datasets. However, the CPM had suboptimal and variable discrimination in the 3 external validation datasets with C-statistics of 0.63, 0.71, and 0.79. Furthermore, at a cut-off that provided similar sensitivity to W4SS, the score did not improve specificity. In 2 of the 3 external validation datasets, 45% and 29% of participants were also excluded because of missing data, respectively. There was no assessment of the clinical utility of the CPM.

Baik et al developed a CPM for tuberculosis in outpatients who were both HIV-positive and HIV-negative and who were symptomatic (i.e., W4SS positive) using a dataset from 28 clinics in South Africa.[112] The CPM included age, sex, HIV status, diabetes, W4SS symptoms, and cough ≥2 weeks as predictors. The CPM was converted to a simplified score

following categorization of continuous predictors and externally validated using a dataset from 4 clinics in Uganda. In the development and external validation datasets, the definition of tuberculosis was only a positive sputum Xpert. Discrimination was acceptable in the validation cohort (C-statistic: 0.75). However, discrimination was not assessed specifically in those who were HIV-positive. Decision curve analysis was performed but the score was not compared to other screening tools such as the W4SS. Because the authors used a random sample of those without tuberculosis, spectrum bias is a concern.

Hanifa et al developed a CPM among outpatient PLHIV with a positive W4SS who were drawn from a larger cohort of PLHIV enrolled irrespective of symptoms and signs of tuberculosis in South Africa.[113] The CPM included ART status, CD4 cell count, BMI, and W4SS as predictors. The analysis involved splitting the cohort by time into development and internal validation datasets. The development dataset had 52 tuberculosis cases. However, 11 predictors were assessed including nonlinear terms for continuous variables and interaction terms, meaning that overfitting of the model was likely. The final CPM was converted to a simplified score following categorization of continuous predictors. The score showed acceptable discrimination during internal validation (C-statistic: 0.72). Boyles et al externally validated the full CPM among outpatient PLHIV who were enrolled from a PCF setting in South Africa and found adequate calibration but suboptimal discrimination (C-statistic: 0.65), suggesting that performance would be even lower for the simplified score that was derived from the full CPM.[115]

Balcha et al developed a CPM among outpatient PLHIV not on ART with a positive W4SS who were drawn from a larger cohort enrolled irrespective of symptoms and signs of tuberculosis in Ethiopia.[114] The CPM incorporated cough, lymphadenopathy, haemoglobin, Karnofsky score, and mid-upper arm circumference (MUAC). During model development, initial predictors were selected using their univariable associations with tuberculosis – a procedure that may falsely exclude important predictors. Continuous variables were also dichotomized. The development dataset had 137 tuberculosis cases, but since 25 predictors were assessed, overfitting of the model is likely. Furthermore, the authors excluded those unable to produce a sputum sample and those with a clinical diagnosis of tuberculosis. The final CPM was converted to a simplified score, which showed acceptable discrimination in the derivation cohort (C-statistic: 0.75). However, the CPM and simplified score were not internally or externally validated. The CPM and simplified score are also complex, requiring

examination for lymphadenopathy, measurement of haemoglobin concentration, and assessment of Karnofsky Performance.

Boyles et al developed 2 CPMs for active tuberculosis in outpatient PLHIV with a positive W4SS in South Africa based on first visit and return visit, respectively.[115] The CPM based on first visit included ART status, number of W4SS symptoms, duration of W4SS symptoms, and temperature as predictors. The CPM based on return visit included change in symptoms after antibiotics, CRP at return visit, number of W4SS symptoms, duration of W4SS symptoms, and ART status. Both models were developed and internally validated according to the TRIPOD principles.[107] During internal validation, the CPM based on first visit showed suboptimal discrimination (C-statistic: 0.68), while the CPM based on return visit showed acceptable discrimination (C-statistic: 0.76). However, externally validation and clinical utility of both models has not yet been assessed.

Nanta et al developed a CPM among PLHIV who were enrolled regardless of symptoms and signs of tuberculosis from an ART clinic, tuberculosis clinic, and outpatient and inpatient departments in Thailand.[116] The CPM included BMI ≤19 kg/m^2, cough >2 weeks, shaking chills ≥1 week, ART status, CD4 cell count ≤200 cells/µl, and history of tuberculosis. However, the study had several limitations. During model development, the authors selected predictors using their univariable associations with tuberculosis. Overfitting was likely, because 43 predictors were assessed, but there were only 66 cases of tuberculosis. The CPM has not been validated (either internally or externally), and clinical utility has not yet been assessed.

In summary, although there are several CPMs for tuberculosis screening in PLHIV enrolled irrespective of tuberculosis symptoms and signs or in PLHIV with a positive W4SS, they have several limitations. Current CPMs have been developed using many predictors relative to number of events,[111,113,114,116] univariable selection of predictors,[114,116] or categorization of continuous variables.[111-114] Current CPMs have also not been internally validated,[114,116] shown suboptimal performance at external validation,[111,113] have not been externally validated or extensively externally validated,[112-116] or have not been assessed for clinical utility.[111,113-116]

1.5.5 Other laboratory tests and biomarkers for tuberculosis screening

Several other laboratory tests and biomarkers are predictors of tuberculosis and may therefore be useful for tuberculosis screening in PLHIV.

Low CD4 cell count plays a large role in the increased risk of tuberculosis in PLHIV. *Mycobacterium tuberculosis* infects macrophages, which require $CD4^+$ lymphocytes to augment clearance.[11] A recent meta-analysis showed that there was a 1.43 (1.16, 1.88) fold increase in tuberculosis incidence for every 100 cells/μL decrease in CD4 cell count.[80] Clinical presentation varies depending on the degree of immunosuppression, as does diagnostic accuracy of tests.[16] For example, AlereLAM and FujiLAM have significantly lower sensitivities at higher CD4 cell counts.[33,75]

Low blood haemoglobin concentration (i.e., anaemia) is a predictor of tuberculosis. Although anaemia may predispose to the development of tuberculosis, it is more likely that anaemia is an early marker of tuberculosis that develops in the period before clinical disease is apparent.[117] Anaemia is largely a result of anaemia of chronic disease.[118] According to WHO, anaemia is classified as mild (11.0 to 12.9 g/dL for men and 11.0 to 11.9 g/dL for women), moderate (8.0 to 10.9 g/dL), or severe (<8.0 g/dL).[119] In a recent systematic review, anaemia was significantly associated with increased risk of tuberculosis in 12 out of 14 studies conducted among PLHIV.[120] The authors also found a significant dose-response relationship between risk of tuberculosis and severity of anaemia in 6 out of 7 studies.[120] In 1 study of outpatient PLHIV not on ART, sensitivities of assays for the diagnosis of tuberculosis were significantly higher in participants with moderate to severe anaemia than in those with no or mild anaemia using AlereLAM (55% vs 0%) and Xpert (74% vs 41%).[121] These results suggest that anaemia may have utility in tuberculosis screening and that confirmatory testing be routinely performed in those with anaemia.

Biomarker-based screening tools that have been discovered in different "omics" levels (e.g., genomics, transcriptomics, proteomics, and metabolomics) might also be useful for tuberculosis screening in PLHIV. Several blood transcriptional biomarkers have been discovered for active pulmonary tuberculosis.[122] In one study that recruited PLHIV irrespective of tuberculosis symptoms and signs from a community setting, the sensitivity of an 11-gene blood transcriptional signature (called RISK11) was 88% (58, 100) and specificity was 66% (63, 69), while the sensitivity of W4SS was only 30% (CI 0, 63) and specificity was 94% (93, 96).[123]

1.5.6 Other clinical features for tuberculosis screening

Several other clinical features are predictors of tuberculosis and may therefore be useful for tuberculosis screening in PLHIV.

ART leads to rising CD4 cell counts, and ART is associated with a markedly reduced incidence of tuberculosis among PLHIV (HR 0.35 [0.28, 0.44]).[124] However, incidence of tuberculosis remains high among PLHIV on ART compared with those who do not have HIV.[125,126]

Common anthropometric measures of malnutrition include BMI and MUAC. Low BMI has been shown to impair the innate and adaptive immune responses, predisposing to the development of tuberculosis.[127,128] However, a bidirectional relationship likely exists, since weight loss is a common symptom of tuberculosis.[129] Low BMI is classified as a BMI <18.5 kg/m^2.[130] In a recent systematic review and meta-analysis of 104,387 PLHIV, low BMI was significantly associated with increased risk of tuberculosis in 7 out of 10 studies (pooled HR 2.1 [1.6, 2.7]).[131] MUAC is a less commonly used anthropometric measure, but low MUAC (<200 mm) was found to be predictive of tuberculosis in outpatient PLHIV initiating ART.[114]

Lymphadenopathy is a clinical sign that may indicate tuberculosis disease. Tuberculosis may present with lymphadenopathy either as tuberculous adenitis (a form of extra-pulmonary tuberculosis) or reactive lymphadenitis.[132] Other infections and malignancies may also present with lymphadenopathy.[132] In one study of outpatient PLHIV not on ART and enrolled irrespective of tuberculosis signs and symptoms,[133] lymphadenopathy had moderate sensitivity (67% [53, 79]) and specificity (55% [50, 60]) for the diagnosis of tuberculosis. The combination of cough and lymphadenopathy had high sensitivity (93% [82, 98]), but low specificity (25% [21, 29]).

The WHO defines a PLHIV as 'seriously ill' if he or she has at least one WHO danger sign (respiratory rate >30 breaths/min, body temperature >39 °C, heart rate >120 beats/min, or is unable to walk unaided).[134] In the 2007 WHO tuberculosis guideline to diagnose smear-negative tuberculosis,[134] WHO danger signs were developed to determine if referral to a higher level facility was needed for a patient suspected of having tuberculosis in a HIV-prevalent setting. However, the use of danger signs was largely developed based on expert opinion. WHO danger signs have limited utility in determining hospital admission. For example, in the STAMP trial, 79% of all HIV-positive medical inpatients had no WHO danger signs but still required hospital admission.[18] In the 2015 WHO LF-LAM tuberculosis guideline and 2016 WHO ART guideline, [34,65] WHO danger signs are included as part of screening criteria to determine eligibility for AlereLAM testing. However, WHO danger

signs were not evaluated in the review that led to the recommendation. Thus, the role of danger signs in WHO tuberculosis screening guidelines is unclear.

1.5.7 No screening tool (i.e., confirmatory testing for all PLHIV)

The authors of two studies conducted among PLHIV initiating ART have argued that the preferred strategy should be confirmatory testing for all PLHIV, rather than confirmatory testing only in those who have a positive result on a screening tool (e.g., a positive W4SS).[135,136] In both studies, the W4SS was negative in 16% and 22% of participants with confirmed tuberculosis, respectively, meaning that these cases would have been missed by screening.[135,136] Routine confirmatory testing may be especially useful in certain subgroups of PLHIV such as those not on ART, those with low CD4 cell counts, and inpatients. These subgroups have a high pre-test probability of tuberculosis, and the consequence of missing a case is also high. In a study of inpatient PLHIV who were enrolled regardless of signs and symptoms, the W4SS had very low specificity (11% [8, 15]) and tuberculosis prevalence was very high (33% [28, 37], indicating that routine confirmatory testing in all inpatients may be the preferred strategy in this subgroup.[20] Although a "confirmatory testing for all PLHIV" approach could optimise diagnostic yield, cost and capacity issues may restrict its implementation in resource-poor settings.

1.6 Diagnostic test accuracy and CPM research using IPDMA

A conventional meta-analysis involves statistical analyses of aggregate data from multiple studies to provide a summary pooled estimate (e.g., sensitivity and specificity).[137] On the other hand, an IPDMA uses the individual-level raw data of each participant from multiple studies to provide a summary pooled estimate.

IPDMA of diagnostic test accuracy offers several advantages over a conventional meta-analysis.[138] First, some studies may not be included in a meta-analysis because the study collected data on the index test but did not publish that data. Some studies may also publish data on an index test but do not report sufficient data on the index test to allow for inclusion of the study in a meta-analysis. Therefore, obtaining IPD would allow investigators to include more eligible studies in the analysis, increasing sample size and reducing publication bias. Second, for tests that use a continuous measure (e.g., CRP), some studies may not report aggregate data at a threshold of interest. IPDMA allows investigators to use the same threshold in multiple studies. IPDMA also allows meta-analysis at multiple thresholds. Third,

IPDMA enables investigators to standardize index tests and reference standards across multiple studies. Fourth, IPDMA allows investigators to evaluate different combinations of index tests. Finally, since test accuracy may vary by key subgroups (e.g., by ART status), IPDMA allows investigators to examine test accuracy across those subgroups.

Similarly, IPDMA of CPMs offers several advantages for both CPM development and validation.[138,139] For CPM development, IPDMA enables investigators to increase sample size, reducing the possibility of overfitting. For CPM validation, IPDMA allows investigators to assess generalizability of a CPM across several settings and subgroups. If performance is found to be suboptimal in certain settings or subgroups, IPDMA can improve performance using various updating strategies.[140] Finally, if 2 or more CPMs for a target population are described in the literature, an IPDMA also allows investigators to compare performance of the CPMs overall and across different settings.[138]

An IPDMA also has disadvantages. IPDMA involves collecting and checking data for multiple studies – procedures that are time consuming. An IPDMA is also complex, requiring greater statistical expertise to perform analyses. Some studies may also not provide data, potentially increasing the risk of selection bias and decreasing precision.

1.7 Thesis rationale

Early and accurate diagnosis of HIV-associated tuberculosis is an important component of the WHO End TB Strategy.[26] However, WHO-recommended screening tools and strategies to guide confirmatory testing with Xpert and LF-LAM (e.g., AlereLAM) have suboptimal diagnostic accuracy in PLHIV.

For the WHO Xpert algorithm (Table 1-3), the W4SS has low specificity in those not on ART, leading to many unnecessary and expensive follow-up confirmatory tests with Xpert in this population.[82,83] Furthermore, the W4SS might have low specificity in inpatients since HIV-related opportunistic diseases often present with one or more of the W4SS symptoms.[141] The W4SS also has reduced sensitivity in some subgroups such as those who are on ART. The entire WHO Xpert algorithm (W4SS followed by Xpert) may also have low sensitivity since overall sensitivity depends on the combined sensitivity of the W4SS and Xpert. Alternative screening tools and strategies to guide Xpert confirmatory testing need to be explored. A ‘confirmatory testing for all’ strategy with Xpert (i.e., no use of a screening tool) also needs to be explored.

For the WHO AlereLAM inpatient algorithm, WHO screening criteria to guide AlereLAM may be challenging to implement in busy inpatient settings.[142] The diagnostic accuracy of WHO screening criteria is also unknown. Since the W4SS was developed among ambulatory PLHIV,[82] it may have low specificity in inpatients who are frequently symptomatic. Furthermore, inpatients typically have advanced immunodeficiency, meaning that CD4 cell count may also have low specificity. CD4 cell count is also often not rapidly available. Furthermore, WHO has recommended that WHO-defined danger signs be used as a screening tool to guide AlereLAM confirmatory testing, but the diagnostic accuracy of WHO-defined danger signs was not assessed in the review that led to the recommendation.[66] Alternative screening tools to guide AlereLAM testing need to be explored. A 'confirmatory testing for all' strategy with AlereLAM (i.e., no use of screening tools or criteria) in an inpatient setting may be more appropriate than AlereLAM confirmatory testing only if screening criteria are met.

Alternative screening tools and strategies are therefore needed to meet the ambitious WHO End TB Strategy targets. Improving tuberculosis screening may 1) increase the number of tuberculosis cases captured; 2) reduce the number of expensive follow-up confirmatory tests required; 3) lead to more effective and timely implementation of IPT/TPT since screening is required to rule out active tuberculosis prior to provision of IPT/TPT; 4) prevent inadvertent IPT/TPT in those who need tuberculosis treatment; 5) reduce diagnostic complexity, which is an important consideration in busy clinical settings; and 6) reduce community transmission.

By using individual-level data, an IPDMA provides a unique opportunity to overcome several limitations of conventional meta-analyses and address some of the evidence gaps that exist in WHO tuberculosis screening and diagnostic guidelines for PLHIV.

1.8 Aim and objectives

1.8.1 Aim

The aim of this thesis is to improve the diagnosis of HIV-associated tuberculosis by investigating alternative tuberculosis screening tools and strategies within an IPDMA framework.

1.8.2 Objectives

There are four separate objectives of this thesis:

1. To determine the diagnostic accuracy of the W4SS and alternative screening tools and strategies in ambulatory PLHIV, including key subgroups.
    a. First, I compared diagnostic accuracy of different tuberculosis screening tools and strategies with diagnostic accuracy of the W4SS to guide confirmatory testing.
    b. Second, I compared the diagnostic accuracy of the WHO-recommended Xpert algorithm (i.e., W4SS followed by Xpert) with the diagnostic accuracy of Xpert confirmatory testing for all ambulatory PLHIV.
2. To determine the performance of the W4SS and alternative screening tools and strategies in HIV-positive inpatients
    a. First, I determined the proportion of inpatients eligible for Xpert confirmatory testing using the WHO Xpert algorithm (W4SS followed by Xpert).
    b. Second, I compared diagnostic accuracy of different tuberculosis screening tools and strategies with diagnostic accuracy of the W4SS to guide confirmatory testing.
    c. Third, I compared diagnostic accuracy of the WHO Xpert algorithm with diagnostic accuracy of Xpert confirmatory testing for all inpatients.
    d. Fourth, I determined the diagnostic yield of Xpert (i.e., proportion of total tuberculosis cases with a positive Xpert test)
3. To determine the performance of WHO screening criteria and alternative screening tools and strategies to guide LF-LAM testing in HIV-positive inpatients
    a. First, I determined the proportion of inpatients eligible for AlereLAM confirmatory testing using the WHO AlereLAM algorithm (i.e., WHO screening criteria followed by AlereLAM).
    b. Second, I compared diagnostic accuracy of different tuberculosis screening tools and strategies with diagnostic accuracy of the WHO screening criteria to guide confirmatory AlereLAM testing.
    c. Third, I compared diagnostic accuracy of the WHO-recommended AlereLAM algorithm with diagnostic accuracy of AlereLAM or FujiLAM testing in all inpatients.
    d. Fourth, I determined the diagnostic yield of tuberculosis confirmatory testing (i.e., proportion of total tuberculosis cases with a positive sputum or urine Xpert, AlereLAM, and/or FujiLAM).

e. Fifth, I evaluated diagnostic accuracy of WHO-defined danger signs for tuberculosis.

4. To develop and validate novel CPMs for pulmonary tuberculosis screening in outpatient PLHIV and to determine the clinical utility of these CPMs and WHO-recommended screening tools.
    a. First, I developed and validated novel CPMs that combined CRP with several routinely available clinical predictors for active pulmonary tuberculosis among outpatient PLHIV.
    b. Second, I determined the optimal tuberculosis screening approach by comparing the performance and clinical utility of these novel CPMs with that of WHO-recommended tests: W4SS and CRP (when this analysis began, WHO had recommended CRP for tuberculosis screening in PLHIV based on the findings of the first objective)

Chapter 2: TUBERCULOSIS SCREENING AMONG AMBULATORY PEOPLE LIVING WITH HIV: A SYSTEMATIC REVIEW AND INDIVIDUAL PARTICIPANT DATA META-ANALYSIS

This study is published in *The Lancet Infectious Diseases*:

Dhana AV, Hamada Y, Kengne AP, Kerkhoff AD, Rangaka MX, Kredo T, Baddeley A, Miller C, Singh S, Hanifa Y, Grant AD, Fielding K, Affolabi D, Merle CS, Wachinou AP, Yoon C, Cattamanchi A, Hoffmann CJ, Martinson N, Mbu ET, Sander MS, Balcha TT, Skogmar S, Reeve BWP, Theron G, Ndlangalavu G, Modi S, Cavanaugh J, Swindells S, Chaisson RE, Ahmad Khan F, Howard AA, Wood R, Thit SS, Kyi MM, Hanson J, Drain PK, Shapiro AE, Kufa T, Churchyard G, Nguyen DT, Graviss EA, Bjerrum S, Johansen IS, Gersh JK, Horne DJ, LaCourse SM, Al-Darraji HAA, Kamarulzaman A, Kempker RR, Tukvadze N, Barr DA, Meintjes G, Maartens G.

"Tuberculosis screening among ambulatory people living with HIV: a systematic review and individual participant data meta-analysis"
The Lancet Infectious Diseases. 2022 Apr;22(4):507-518
https://doi.org/10.1016/S1473-3099(21)00387-X

2.1 Abstract

Background: The WHO-recommended tuberculosis screening and diagnostic algorithm in ambulatory people living with HIV is a four-symptom screen (known as the WHO-recommended four symptom screen [W4SS]) followed by a WHO-recommended molecular rapid diagnostic test (eg Xpert MTB/RIF [hereafter referred to as Xpert]) if W4SS is positive. To inform updated WHO guidelines, we aimed to assess the diagnostic accuracy of alternative screening tests and strategies for tuberculosis in this population.

Methods: In this systematic review and individual participant data meta-analysis, we updated a search of PubMed (MEDLINE), Embase, the Cochrane Library, and conference abstracts for publications from Jan 1, 2011, to March 12, 2018, done in a previous systematic review to include the period up to Aug 2, 2019. We screened the reference lists of identified pieces and contacted experts in the field. We included prospective cross-sectional, observational studies and randomised trials among adult and adolescent (age ≥10 years) ambulatory people living with HIV, irrespective of signs and symptoms of tuberculosis. We extracted study-level data using a standardised data extraction form, and we requested individual participant data from study authors. We aimed to compare the W4SS with alternative screening tests and strategies and the WHO-recommended algorithm (ie, W4SS followed by Xpert) with Xpert for all in terms of diagnostic accuracy (sensitivity and specificity), overall and in key subgroups (eg, by antiretroviral therapy [ART] status). The reference standard was culture. This study is registered with PROSPERO, CRD42020155895.

Findings: We identified 25 studies, and obtained data from 22 studies (including 15 666 participants; 4347 [27·7%] of 15 663 participants with data were on ART). W4SS sensitivity was 82% (95% CI 72–89) and specificity was 42% (29–57). C-reactive protein (≥10 mg/L) had similar sensitivity to (77% [61–88]), but higher specificity (74% [61–83]; n=3571) than, W4SS. Cough (lasting ≥2 weeks), haemoglobin (<10 g/dL), body-mass index (<18·5 kg/m²), and lymphadenopathy had high specificities (80–90%) but low sensitivities (29–43%). The WHO-recommended algorithm had a sensitivity of 58% (50–66) and a specificity of 99% (98–100); Xpert for all had a sensitivity of 68% (57–76) and a specificity of 99% (98–99). In the one study that assessed both, the sensitivity of sputum Xpert Ultra was higher than sputum Xpert (73% [62–81] *vs* 57% [47–67]) and specificities were similar (98% [96–98] *vs* 99% [98–100]). Among outpatients on ART (4309 [99·1%] of 4347 people on ART), W4SS sensitivity was 53% (35–71) and specificity was 71% (51–85). In this population, a parallel

strategy (two tests done at the same time) of W4SS with any chest x-ray abnormality had higher sensitivity (89% [70–97]) and lower specificity (33% [17–54]; n=2670) than W4SS alone; at a tuberculosis prevalence of 5%, this strategy would require 379 more rapid diagnostic tests per 1000 people living with HIV than W4SS but detect 18 more tuberculosis cases. Among outpatients not on ART (11 160 [71·8%] of 15 541 outpatients), W4SS sensitivity was 85% (76–91) and specificity was 37% (25–51). C-reactive protein (≥10 mg/L) alone had a similar sensitivity to (83% [79–86]), but higher specificity (67% [60–73]; n=3187) than, W4SS and a sequential strategy (both test positive) of W4SS then C-reactive protein (≥5 mg/L) had a similar sensitivity to (84% [75–90]), but higher specificity than (64% [57–71]; n=3187), W4SS alone; at 10% tuberculosis prevalence, these strategies would require 272 and 244 fewer rapid diagnostic tests per 1000 people living with HIV than W4SS but miss two and one more tuberculosis cases, respectively.

Interpretation: C-reactive protein reduces the need for further rapid diagnostic tests without compromising sensitivity and has been included in the updated WHO tuberculosis screening guidelines. However, C-reactive protein data were scarce for outpatients on ART, necessitating future research regarding the utility of C-reactive protein in this group. Chest x-ray can be useful in outpatients on ART when combined with W4SS. The WHO-recommended algorithm has suboptimal sensitivity; Xpert for all offers slight sensitivity gains and would have major resource implications.

2.2 Introduction

Tuberculosis is the leading cause of death among PLHIV and often goes undiagnosed.[12,143] One approach to reduce this tuberculosis burden involves systematic screening as part of an intensified case-finding strategy. WHO recommends a tuberculosis screening and diagnostic algorithm in PLHIV at each clinical encounter using the W4SS (comprising any one of current cough, fever, night sweats, or weight loss) followed by confirmatory testing using a WHO-recommended molecular rapid diagnostic test such as Xpert or Xpert Ultra for those with a positive W4SS.[82,144] However, the W4SS has low specificity, meaning many people require unnecessary and expensive confirmatory testing with a rapid diagnostic test.[82,83] Furthermore, the W4SS has reduced sensitivity in specific subgroups (eg, those who are on ART, are pregnant, or have high CD4 counts).[82,83,145] The entire algorithm might also have low sensitivity,[89] because overall sensitivity depends on the combined sensitivity of the W4SS and the rapid diagnostic test.

Alternative screening tests to the W4SS need to be explored. According to WHO, a screening test should have a sensitivity of more than 90% and a specificity of more than 70%.[79] Several studies have shown that CRP has improved diagnostic accuracy compared with W4SS.[86,88,90] CRP assays as point-of-care assays are easy to use, inexpensive (approximately US$2 per test), and provide rapid results (<3 min). One study among PLHIV initiating ART found that replacing the W4SS with CRP (10 mg/L) could halve the number of Xpert tests performed.[89] Chest X-ray might also be useful for tuberculosis screening, especially when combined with the W4SS in PLHIV on ART;[83] however, it is often unavailable and resource intensive. Haemoglobin, BMI, and lymphadenopathy are other predictors of tuberculosis,[113,114] but their diagnostic accuracy is unclear. The authors of some studies among PLHIV initiating ART have argued that Xpert for all, rather than Xpert only for those who are positive on the W4SS, should be the preferred strategy.[135,136] This approach could optimise diagnostic yield, but cost and capacity issues could restrict its implementation in resource-poor settings. The AlereLAM lateral flow urine assay for screening outpatients living with HIV has been recently reviewed and has a sensitivity of 31% and specificity of 95%;[33] next-generation assays based on detection of lipoarabinomannan (eg, Fujifilm SILVAMP TB-LAM) have higher sensitivity (e.g., 71%).[75] WHO recommends the AlereLAM assay if an outpatient has a positive W4SS, CD4 count of 100 cells per μL or lower, is WHO clinical stage 3 or 4, or has a WHO-defined danger sign.[58]

We did a systematic review and IPDMA to provide a more detailed and precise analysis of the accuracy of different tuberculosis screening tests and strategies compared with W4SS among ambulatory PLHIV, including key subgroups. We also assessed the accuracy of the WHO-recommended screening and diagnostic algorithm (W4SS followed by Xpert) and compared its accuracy with Xpert for all as the first screening test.

2.3 Methods

Search strategy and selection criteria

In this systematic review and IPDMA, we updated the systematic review done by Hamada and colleagues,[83] who searched PubMed (MEDLINE), Embase, the Cochrane Library, and conference abstracts (from the Conference on Retroviruses and Opportunistic Infections, AIDS/International AIDS Society, and International Union Against TB and Lung Diseases conferences) without language or geographical restrictions from Jan 1, 2011, to March 12, 2018. The start date restrictions correspond to the year WHO issued recommendations on the W4SS. We rescreened all potential full texts identified via Hamada and colleagues' search to identify eligible studies. Additionally, we applied the same search strategy to the same databases for publications between March 12, 2018, and Aug 2, 2019. We also screened reference lists of reviews and included articles and contacted field experts. Detailed search terms are in the appendix (Table 8-1).

Two authors (AD and YHam) independently screened titles and abstracts from the search and subsequently screened the full texts of potentially eligible articles. For abstracts that were not in English, we used Google Translate to translate the abstracts before screening. We included prospective cross-sectional studies, prospective observational studies, and randomized trials that collected at least one sputum sample for tuberculosis culture from adult and adolescent (i.e., aged ≥10 years) ambulatory PLHIV regardless of signs and symptoms of tuberculosis. We excluded case-control studies, general community or household contact-screening studies, and studies that involved PLHIV who were already on tuberculosis treatment or had a current tuberculosis diagnosis.

The target condition was active tuberculosis (i.e., we excluded articles on latent tuberculosis infections). The reference standard for confirmed tuberculosis was bacteriological confirmation of *Mycobacterium tuberculosis* using culture of a sputum sample or other samples, or both.

We included primary datasets that had sufficient data to allow us to compare the W4SS with alternative screening tests or strategies and the WHO-recommended algorithm (W4SS followed by Xpert) with Xpert for all. We examined several systematically performed screening tests: CRP, chest X-ray, Xpert or Xpert Ultra, haemoglobin, BMI, lymphadenopathy (on examination), and cough (lasting ≥2 weeks). A positive chest X-ray was defined by the authors of the included studies and categorised as any abnormality or abnormality suggestive of tuberculosis. We were primarily interested in any abnormality on chest X-ray because identification of features suggestive of tuberculosis on chest X-ray requires a skilled reader. For CRP, we primarily focused on the 10 mg/L threshold, which is considered the upper limit of normal.[146,147] We also explored a 5 mg/L threshold to maximize sensitivity and an 8 mg/L threshold because a previous study found that this cutoff met WHO's minimum sensitivity (≥90%) and specificity (≥70%) targets.[79,88] Finally, we examined several parallel strategies (two screening tests offered at the same time) to improve sensitivity and sequential strategies (second screening test offered only if first screening test is positive) to improve specificity.

We have reported our findings according to the PRISMA-IPD and PRISMA-DTA statements.[148,149] This study was registered with PROSPERO (CRD42020155895).

Data extraction, study quality, and IPD synthesis

Using a standardized data extraction form, two authors (AD and YHam) independently extracted study-level information on first author, publication year, study period, country, setting (e.g., HIV clinic, hospital clinic, prison clinic), exclusion criteria, study design, type of participants (e.g., all PLHIV, only pregnant people), and method of tuberculosis diagnosis. Two authors independently (AD and YHam) assessed study quality using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[150]

We invited authors of eligible datasets by email to contribute IPD. We prespecified variables to be collected after consultation with WHO and our study group (appendix Table 8-2). We standardized IPD, then synthesized a single dataset with study-level data. Study participants younger than 10 years were excluded, and contaminated cultures were considered negative. To ensure integrity of the IPD, we checked information against study publications and did checks on each dataset for missing, duplicate, invalid, and implausible items.[151,152] We resolved discrepancies by contacting the corresponding author.

Statistical analysis

We did analyses overall and in key subgroups, comprising outpatient clinic attendees (on ART *vs* not on ART), CD4 count (≤200 *vs* >200 cells per μL), and pregnancy. To analyze IPD we used a two-stage approach. IPD were first analyzed separately in each study using an appropriate statistical method (accounting for the design of data collection) and reduced to aggregate data, which were then synthesized using meta-analytical techniques.

In the first stage, we estimated tuberculosis prevalence, positivity rate (proportion of screen-positive participants), and measures of diagnostic performance (including sensitivity and specificity) by screening test or strategy. In the second stage, we pooled tuberculosis prevalence and positivity rates using a generalized linear mixed model with logit transformation[153] in preference to the protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation. We assessed heterogeneity using Cochran's Q test and the I^2 statistic.[154] We pooled absolute accuracy measures (sensitivity, specificity) in a bivariate generalised linear mixed model.[137] In the case of non-convergence, we assumed no correlation between measures of sensitivity and specificity to simplify the model.[155] When data were sparse, we did not do a meta-analysis (e.g., for CRP [n=62] and lymphadenopathy [n=34] in pregnant participants). We illustrated the absolute pooled sensitivity and specificity using summary receiver-operating characteristic (ROC) curves.[156] To compare the accuracy of screening tests and strategies, we did both indirect and direct comparisons. Direct comparisons were based on studies that assessed both tests of interest; indirect comparisons were based on all studies that assessed at least one test of interest. We did a bivariate meta-regression with test type as a covariate and used likelihood ratio tests to assess the significance of differences in sensitivity and specificity. We explored study-level characteristics (tuberculosis prevalence and reference standard) as potential sources of heterogeneity. Accounting for the variation of tuberculosis prevalence across studies and their pooled values, we applied pooled accuracy estimates to a hypothetical cohort of 1000 individuals to show the consequences of using each screening test and strategy, which included calculating negative and positive predictive values using Bayes' theorem. We also calculated predictive values using a trivariate generalised linear mixed model that jointly models predictive values and test prevalence.[157]

We did several sensitivity analyses. We assessed diagnostic accuracy using a prespecified second reference standard of culture or Xpert. This analysis included one additional study of outpatients living with HIV (not on ART and on ART) that did not meet our primary reference

standard criterion.[113] We also assessed diagnostic accuracy using a reference standard of Xpert alone because it is one of the molecular rapid diagnostic tests recommended by WHO. Finally, we did a direct comparison of the accuracy of W4SS followed by Xpert with the accuracy of CRP (≥10 mg/L) followed by Xpert.

We assessed publication bias with funnel plots (for analyses with ten or more studies) and applied Egger's test. Although Deeks' test might be more appropriate, most methods to test for publication bias in studies of test accuracy have limitations.[158] Therefore, we also applied the trim-and-fill method to provide bias-adjusted estimates.[159]

We selected a p value threshold of 0.05 to characterise statistically significant findings. We did all meta-analyses using *lme*, *altmeta*, *meta*, *metafor,* and *mada* packages in R (version 3.6.1). The substantive protocol deviations were that we did not perform a leave-one-out sensitivity analysis and did not compare IPD results with aggregate data for which IPD were not obtained because we obtained more than 90% of requested data.

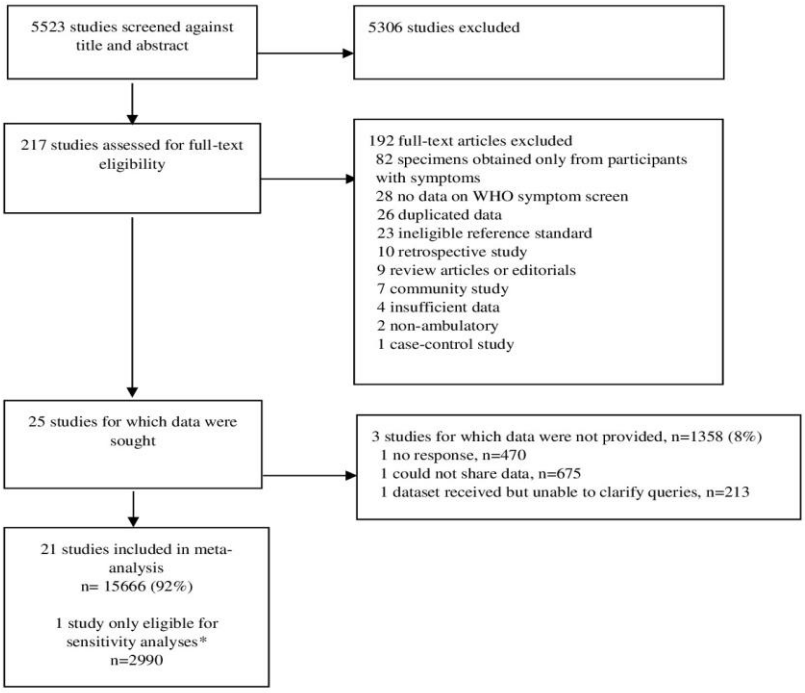
Role of the funding source

The funder had a role in the study design, data collection, data analysis, data interpretation, and writing of the report.

2.4 Results

Of 5523 potentially eligible publications, 25 were eligible (Figure 2-1). IPD were provided for 22 studies (including one study[113] that was eligible only for sensitivity analyses).[88,90,91,114,133,145,160-174] IPD were not provided for three studies.[175-177] Hence, we obtained IPD for 15 666 (92%) of 17 024 participants identified. The characteristics of included studies are shown in the appendix (Table 8-3). The studies collected data from 2007 to 2020. 18 studies were done in sub-Saharan Africa. Two studies included only pregnant women, and one study included only people living in prison. Overall, we judged studies as low risk of bias in most QUADAS-2 domains (appendix Figure 8-1), but six studies had high applicability concerns for participant selection (e.g., selected only PLHIV with advanced immunosuppression). Missing data by study are shown in the appendix (Table 8-4).

Figure 2-1: Study selection



W4SS=WHO-recommended four symptom screen. *One study (Hanifa and colleagues[113]) was incorporated into sensitivity analyses because the study's reference standard made it ineligible for the main analyses.

Participant characteristics overall are shown in Table 2-1 and by study are shown in the appendix (Table 8-5). 10 388 (66.3%) of 15 666 participants were female, and 4347 (27.8%) of 15 663 with available data were on ART. W4SS was positive in 8028 (51.3%) of 15 625 participants, and CRP was elevated (≥10 mg/L) in 1259 (35.1%) of 3582 participants. CRP was measured with a point-of-care assay (2695 participants) or laboratory assay (887 participants) in five studies. The median CD4 count was 269 cells per μL (IQR 142–439; in 15 281 participants).

Table 2-1: Summary of main characteristics for all participants†

Variable	All	N‡
Participants	15666 (100)	
Clinical setting		15666
Outpatient	15541 (99.2)	
Other setting*	125 (0.8)	
Age (years)	34 (28-42)	15666
Female	10388 (66.3)	15666
On ART	4347 (27.8)	15663
CD4 count (cells/μL)	269 (142-439)	15281
History of tuberculosis	1955 (17.5)	11148
W4SS	8028 (51.3)	15652
Cough	4629 (29.6)	15623
Fever	3391 (21.7)	15631
Weight loss	5575 (35.7)	15602
Night sweats	3270 (20.9)	15630
Cough >= 2 weeks	2205 (20.2)	10919
Lymphadenopathy	374 (15.6)	2394
CXR (suggests tuberculosis)	1296 (21.0)	6177
CXR (any abnormality)	2158 (34.7)	6222
Total Xpert positive**	616 (7.1)	8625
BMI (kg/m²)	22 (19-26)	12704
CRP (mg/L)§	4 (2-21)	3582
CRP (>=10 mg/L)	1259 (35.1)	3582
Hb (g/dL)	12 (10-13)	5118
Hb (<10 g/dL)	1093 (21.4)	5118

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

‡Participants with data available for variable

*One study among a prison population

**Sputum and/or non-sputum Xpert result

§Measured with a point-of-care assay (n=2695) or laboratory assay (n=887)

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

The pooled tuberculosis prevalence was 7.7% (95% CI 5.7–10.4) using culture as a reference standard (Table 2-2). The pooled prevalence of tuberculosis in outpatients not on ART was 9.3% (7.0–12.1) compared with 3.3% (2.2–4.8) among outpatients on ART. For participants with a CD4 count of 200 cells per μL or less, the prevalence of tuberculosis was 13.7% (11.1–16.7) and among those with a CD4 count of more than 200 cells per μL it was 4.9% (3.6–6.6; Table 2-2). Heterogeneity of tuberculosis prevalence was high. The pooled tuberculosis prevalences were slightly higher using a reference standard of either culture or Xpert than with a reference standard of culture alone, but subgroup comparisons remained qualitatively similar (appendix Table 8-6).

Table 2-2: Prevalence of tuberculosis in all participants and by subgroup (using culture as a reference standard)

					Heterogeneity			
Subgroup§	**No studies**	**N**	**No tuberculosis**	**Prevalence % (95% CI)†**	**I² (95% CI)**	**P-value**	**Egger's test (p-value)**	**Subgroup analysis (p-value)††**
All	**21**	**15,611**	**1,347**	**7.7 (5.7-10.4)**	**95 (94-96)**	**<0.0001**	**0.02**	**-**
All (setting and ART status)	**21**	**15,608**	**1,347**	**7.7 (5.7-10.4)**	**95 (94-96)**	**<0.0001**	**0.02**	**-**
Outpatients (on ART)*	9	4,309	137	3.3 (2.2-4.8)	81 (65-90)	<0.0001	0.79	<0.0001
Outpatients (not on ART)	20	11,174	1,195	9.3 (7.0-12.1)	92 (89-94)	<0.0001	0.05	-
Other setting**	1	125	15	12.0 (7.4-19.0)	- (-)	-	-	-
All (CD4 count)	**21**	**15,227**	**1,320**	**7.8 (5.8-10.4)**	**95 (94-96)**	**<0.0001**	**0.02**	**-**
CD4 count <=200 cells/μL	21	5,622	866	13.7 (11.1-16.7)	84 (77-89)	<0.0001	0.03	<0.0001
CD4 count >200 cells/μL	21	9,605	454	4.9 (3.6-6.6)	88 (84-92)	<0.0001	0.22	-
All (pregnancy status)***	**21**	**10,351**	**701**	**6.4 (4.7-8.7)**	**91 (88-94)**	**<0.0001**	**0.15**	**-**
Pregnant	8	1,938	53	2.7 (2.1-3.6)	0 (0-60)	<0.0001	0.04	<0.0001
Non-pregnant	19	8,413	648	7.3 (5.4-9.8)	90 (85-93)	<0.0001	0.21	-

§Subgroup in bold is the overall comparator. For example, all (setting and ART status) contains combined subgroups outpatients (on ART), outpatients (not on ART), and other setting

†Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

††P-value of between subgroups heterogeneity statistic Q (based on random effects model)

*P(subgroup) compares outpatients (on ART) with outpatients (not on ART)

**One study among a prison population

***Pregnancy status unavailable for some studies, female participants in those studies categorized as non-pregnant

Definition of abbreviations: ART = antiretroviral therapy

Plots of sensitivity and specificity for each test in all participants and each subgroup are shown in the appendix (Figure 8-2). Indirect comparisons between each test and W4SS in all participants are shown in Table 2-3 and each subgroup are shown in the appendix (Table 8-7). Among 15 597 participants with available culture results, the sensitivity of W4SS was 82% (95% CI 72–89) and specificity was 42% (29–57; Table 2-3; appendix Figure 8-2). The sensitivity of CRP (≥10 mg/L) was similar to, and its specificity was higher than, that of W4SS (sensitivity 77% [95% CI 61–88; p=0.71], specificity 74% [61–83; p=0.041]; Table 2-3; Figure 2-2). The sensitivity of chest X-ray (with any abnormality) was 72% (65–78) and specificity was 62% (51–71; Table 2-3; appendix Figure 8-2). Cough (lasting ≥2 weeks), haemoglobin (<10 g/dL), BMI (<18.5 kg/m²), and lymphadenopathy had high specificities but low sensitivities, making them unsuitable to be explored further as screening tests.

Table 2-3: Indirect comparisons between each test and W4SS for the detection of tuberculosis in all participants (using culture as a reference standard)

					Difference from W4SS††	
Test	**No studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Sensitivity (p-value)**	**Specificity (p-value)**
W4SS	21	15597	82 (72-89)	42 (29-57)	-	-
CRP (>=10 mg/L)	5	3571	77 (61-88)	74 (61-83)	0.706	0.041
CRP (>=8 mg/L)	5	3571	81 (68-89)	70 (57-81)	0.913	0.071
CRP (>=5 mg/L)	5	3571	87 (77-93)	60 (48-71)	0.512	0.268
CXR (abnormal)	8	6195	72 (65-78)	62 (51-71)	0.261	0.129
CXR (suggests tuberculosis)	8	6150	63 (57-70)	78 (67-86)	0.071	0.005
Cough (any)	21	15568	56 (48-63)	72 (65-79)	<0.0001	0.001
Cough (>=2 weeks)	17	10906	38 (29-49)	84 (77-90)	<0.0001	<0.0001
Hb (<10 g/dL)	9	5116	43 (33-54)	80 (73-85)	0.001	0.001
Hb (<8 g/dL)	9	5116	12 (9-16)	96 (93-97)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	18	12650	29 (22-38)	89 (84-92)	<0.0001	<0.0001
Lymphadenopathy	4	2391	31 (14-55)	90 (75-96)	0.002	0.002
W4SS or CRP (>=10 mg/L)¶	5	3571	88 (63-97)	31 (13-57)	0.358	0.456
W4SS or CXR (abnormal)¶	8	6186	94 (89-97)	20 (10-37)	0.008	0.066
W4SS then CRP (>=5 mg/L)¶	5	3571	70 (31-92)	75 (53-88)	0.546	0.04
W4SS then Xpert*§	12	8557	58 (50-66)	99 (98-100)	-	-
Xpert alone*§	12	8570	68 (57-76)	99 (98-99)	0.094#	0.397#

†Using culture as a reference standard. Indirect comparisons are based on all studies that evaluated at least one of the W4SS or relevant screening tests

††For Xpert alone, the comparator is W4SS then Xpert

¶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

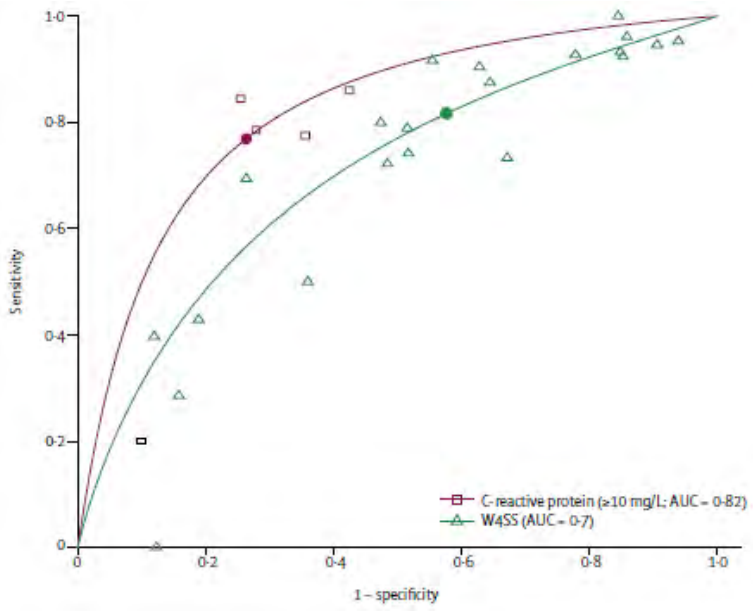
*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result). Alternative algorithms are W4SS then single sputum Xpert (12 studies; 8556 participants; sensitivity 55 (48-63), specificity 99 (99-100) and single sputum Xpert alone (12 studies; 8569 participants; sensitivity 64 (53-74), specificity 99 (98-99).

§One study assessed Xpert and Xpert Ultra among 733 participants. The accuracy of sputum Xpert was: sensitivity 57 (47-67), specificity 99 (98-100); sputum Xpert Ultra: sensitivity 73 (62-81), specificity 98 (96-98); urine Xpert Ultra: sensitivity 27 (19-38), specificity 98 (96-99); sputum and urine Xpert Ultra: sensitivity 75 (65-83), specificity 95 (94-97)

#Bivariate model did not converge; results 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

Figure 2-2: Summary ROC curves comparing C-reactive protein (≥10 mg/L) with W4SS in all participants*



AUC=area under the ROC. ROC=receiver operating characteristic. W4SS=WHO-recommended four-symptom screen. *Data were extrapolated beyond observed datapoints.

Parallel strategies that combined W4SS with either chest X-ray (with any abnormality) or CRP (≥10 mg/L) had higher sensitivities and lower specificities than W4SS alone (Table 2-3). A sequential strategy of W4SS followed by CRP (≥5 mg/L) had a lower sensitivity but higher specificity than W4SS alone. A sequential strategy of W4SS followed by chest X-ray (with any abnormality) had a sensitivity of 63% (54–71) and specificity of 73% (62–82); we did not assess this strategy further because of reduced sensitivity compared with W4SS alone.

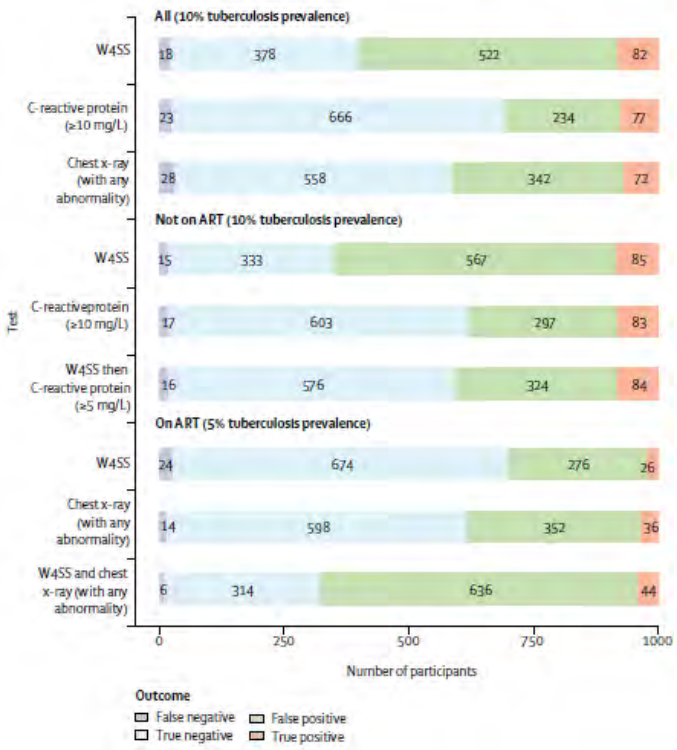
The sensitivity of W4SS followed by Xpert was 58% (95% CI 50–66; Table 2-3). The sensitivity of Xpert for all was 68% (95% CI 57–76). The specificities of both strategies -W4SS followed by Xpert and Xpert for all - were 99% (Table 2-3). The sensitivity of sputum Xpert Ultra was higher than that of sputum Xpert (73% [95% CI 62–81] *vs* 57% [47–67]) and specificities were similar (98% [96–98] and 99% [98–100]) in the only study (unpublished) that compared both tests.[173]

Direct and indirect comparisons of individual tests were largely similar (appendix Table 8-8); however, the lower sensitivity and higher specificity of chest X-ray (with any abnormality) than with W4SS were more pronounced in the direct comparison. Forest plots and summary

ROC curves for all tests and screening strategies are provided in the appendix (Figure 8-3 and Figure 8-4). The point estimates for the specificities of CRP (≥10 mg/L cutoff) were numerically higher than those of W4SS in each individual study that had these data (appendix Figure 8-3). Additional diagnostic accuracy measures are shown in the appendix (Table 8-9).

We assessed how estimates for each test or strategy affected detection rates in a hypothetical cohort of 1000 PLHIV at different tuberculosis prevalences (appendix Table 8-10). At a tuberculosis prevalence of 10%, the W4SS would result in 604 rapid diagnostic tests being needed; CRP (≥10 mg/L) would reduce the number of rapid diagnostic tests needed by 293 but miss five additional tuberculosis cases, and chest X-ray (with any abnormality) would reduce the number of rapid diagnostic tests needed by 190, but miss ten additional tuberculosis cases (Figure 2-3; appendix Table 8-10). At 10% prevalence, the WHO-recommended algorithm (W4SS followed by Xpert) would result in 604 Xpert tests, and Xpert for all would increase the number of Xpert tests needed by 396 (ie, because all 1000 people would receive an Xpert test), but it would detect ten additional tuberculosis cases (appendix Table 8-10).

Figure 2-3: Screening outcomes for selected screening tests and strategies in a hypothetical cohort of 1000 people living with HIV at 10% (all and not on ART) and 5% (on ART) tuberculosis prevalence



ART=antiretroviral therapy. W4SS=WHO-recommended four-symptom screen.

Indirect comparisons by ART status are shown in the appendix (Table 8-7, Figure 8-2). Most tests, except chest X-ray and haemoglobin, had lower sensitivity and higher specificity in outpatients on ART than in outpatients not on ART. In outpatients on ART, a parallel strategy of W4SS and chest X-ray (with any abnormality) had higher sensitivity than W4SS alone (89% [95% CI 70–97] *vs* 53% [35–71]) but lower specificity (33% [17–54] *vs* 71% [51–85]; appendix Table 8-7). In a hypothetical cohort of 1000 outpatients on ART with 5% tuberculosis prevalence, this strategy would increase the number of rapid diagnostic tests needed by 378 compared with W4SS alone but detect 18 additional tuberculosis cases (Figure 2-3; appendix Table 8-10).

In outpatients not on ART, sensitivities for CRP (≥10 mg/L) alone (83% [95% CI 79–86]) and a sequential strategy of W4SS then CRP (≥5 mg/L; 84% [75–90]) were similar to the sensitivity of W4SS alone (85% [76–91]), but their specificities were higher (67% [60–73] for CRP alone; 64% [57–71] for sequential strategy) than with W4SS alone (37% [25–51]; appendix Table 8-7). In a hypothetical cohort of 1000 outpatients not on ART with 10% tuberculosis prevalence, compared with use of W4SS alone, use of CRP (≥10 mg/L) would reduce the number of rapid diagnostic tests needed by 272 but miss two additional tuberculosis cases, and use of the sequential strategy of W4SS then CRP (≥5 mg/L) would reduce the number of rapid diagnostic tests needed by 244 but miss one additional tuberculosis case (Figure 3; appendix Table 8-10).

Indirect comparisons between each test and W4SS by CD4 cell count are shown in the appendix (Table 8-7, Figure 8-2). Most tests, except chest X-ray, had lower sensitivity and higher specificity in participants with CD4 counts of more than 200 cells per μL than those with CD4 counts of 200 cells per μL or lower. Similarly, most tests had lower sensitivity and higher specificity in pregnant women living with HIV than in the overall population (appendix Table 8-7, Figure 8-2); however, these estimates had suboptimal precision. Indirect and direct comparisons for the subgroups were largely similar (appendix Tables 8-7 and 8-8). However, among outpatients on ART, the slightly higher sensitivity of chest X-ray (both with any abnormality and suggestive of tuberculosis) than of W4SS alone in indirect comparisons was attenuated in direct comparisons (appendix Table 8-8). Only one study (n=381) among outpatients on ART assessed CRP (≥10 mg/L), for which there was a similar sensitivity and specificity compared with W4SS alone (appendix Table 8-8).[91]

We did sensitivity analyses using two alternative reference standards: culture or Xpert, and Xpert alone (appendix Table 8-11). Results were largely similar to the main analyses, although

sensitivities were slightly higher for the reference standard of Xpert alone than for the main reference standard of culture. In sensitivity analyses directly comparing W4SS followed by Xpert with CRP (≥10 mg) followed by Xpert, both strategies had similar sensitivities and specificities (appendix Table 8-12).

Egger's test and meta-regression results are provided in the appendix (Table 8-9), as well as funnel plots (appendix Figure 8-5). We found no evidence of publication bias (Egger's test p>0.05) for most tests. Meta-regression showed that prevalence explained some heterogeneity in the analyses for several tests, but reference standard type generally did not.

2.5 Discussion

In this systematic review and IPDMA, we found that the sensitivity of CRP (≥10 mg/L) was similar to that of W4SS alone, but its specificity was higher (74% *vs* 42%). Chest x-ray (with any abnormality) had lower sensitivity than W4SS alone in direct comparisons, making it less suitable than a standalone screening test. Cough (lasting ≥2 weeks), haemoglobin (<10 g/dL), BMI (<18.5 kg/m²), and lymphadenopathy had high specificities (>80%), but their low sensitivities also made them less suitable as screening tests than W4SS. The WHO-recommended algorithm of W4SS then Xpert had a sensitivity of only 58% (95% CI 50–66), and Xpert for all had a slightly higher sensitivity of 68% (57–76). In one unpublished study, Xpert Ultra improved sensitivity over Xpert (73% [62–81] *vs* 57% [47–67]).[173]

Among outpatients on ART, the sensitivity of a parallel strategy of W4SS with chest x-ray (any abnormality) was higher than that of W4SS alone, but its specificity was lower. At 5% tuberculosis prevalence, this strategy was estimated to require more than double the number of rapid diagnostic tests needed compared with W4SS alone but would detect 70% more tuberculosis cases. Among outpatients not on ART, the sensitivities of CRP (≥10 mg/L) and a sequential strategy of W4SS then CRP (≥5 mg/L) were similar to W4SS alone, but specificities were higher. At 10% tuberculosis prevalence, these strategies would reduce the number of rapid diagnostic tests needed by 42% for CRP (≥10 mg/L) and 37% for W4SS then CRP (≥5 mg/L) compared with W4SS alone but would miss a similar number of tuberculosis cases.

We found that CRP (≥10 mg/L) approached the WHO-defined minimum thresholds for a screening test (with 83% sensitivity and 67% specificity *vs* WHO's thresholds of 90% sensitivity and 70% specificity) for outpatients not on ART.[79] Efforts to scale-up of access to WHO-recommended, molecular, rapid diagnostic tests have been slow, particularly in

decentralised locations.[54,178] CRP testing could allow for broader implementation of rapid diagnostic tests because its greater specificity means that screening using CRP would require fewer subsequent rapid diagnostic tests than screening with W4SS. The need for fewer tests could also reduce laboratory processing time; Xpert can provide a result in less than 2 h, but a result often takes several days in the real world.[51] The high specificity of CRP would reduce the time to start tuberculosis preventive therapy in PLHIV. Current CRP point-of-care assays have differing complexities, ranging from qualitative lateral-flow assays that do not require a power source or refrigeration to quantitative assays that require a small machine.[179] CRP point-of-care assays can cost approximately US$2 per test, provide results in less than 3 min, and be performed easily with minimal expertise (blood collected by finger prick). Thus, available point-of-care assays have the potential for affordable scale-up.

The sensitivity of a parallel strategy incorporating W4SS and chest x-ray was higher than the sensitivity of other tests or strategies in those on ART; however, the higher number of rapid diagnostic tests needed might pose a substantial cost burden. Furthermore, a 2016 survey of 14 countries with high HIV-associated tuberculosis burdens found that chest x-ray as a screening tool was available at only 14% of primary health-care centres.[54] We found that the sensitivity of chest x-ray was not increased in those not on ART and at lower CD4 cell counts of 200 per μL or lower; the most likely explanation for these findings is that normal chest x-ray images in patients with pulmonary and extra-pulmonary tuberculosis occur more frequently in those with advanced immunosuppression than in other PLHIV.[97,180]

The low sensitivities of haemoglobin, BMI, and lymphadenopathy make them unsuitable as screening tests. However, haemoglobin levels below 10 g/dL, a BMI of less than 18.5 kg/m², and lymphadenopathy in ambulatory PLHIV should prompt a thorough search for tuberculosis, given their high specificities and known association with mortality.[181,182]

We found that the WHO-recommended strategy (W4SS followed by Xpert) would miss approximately 40% of tuberculosis cases. The low yield is a result of the inadequate sensitivities of both the W4SS and Xpert. Approximately 20% of PLHIV with tuberculosis will be missed with W4SS and thus have subclinical tuberculosis, 56–75% of whom will probably progress to symptomatic disease.[183,184] Although Xpert for all would still miss approximately 33% of tuberculosis cases, Xpert Ultra showed improved sensitivity over Xpert in one study.[173] Xpert Ultra costs the same as Xpert, and the point-of-care GeneXpert Omni platform might allow its use at decentralized locations. Further research is needed to assess this approach.

Our study has limitations. First, we did not have adequate precision in some analyses for outpatients on ART and pregnant PLHIV. Specifically, we had little data on CRP in PLHIV on ART. Furthermore, there was a paucity of data on countries other than South Africa, where almost half of all included studies were done, and which might be more urbanised than other low-income and middle-income countries. Second, we largely excluded participants who were unable to produce a sputum sample, meaning our findings might not generalise to this group. Few studies also systematically included extra-pulmonary tuberculosis samples, meaning our results are more applicable to pulmonary tuberculosis. However, pulmonary tuberculosis probably comprises most tuberculosis cases in an ambulatory screening setting. Third, we used an imperfect reference standard, because sputum culture, which was all that was done in most of the included studies, should ideally comprise multiple samples collected in the early morning to maximise sensitivity, but this was not done in any of our included studies. Fourth, although direct comparison minimises confounding, these analyses involved fewer studies and reduced precision. Fifth, we were unable to obtain IPD from three studies. However, these studies comprised only approximately 8% of data. Sixth, only one study assessed Xpert Ultra,[173] and we did not assess non-Xpert nucleic acid amplification tests. Seventh, our study findings might not be generalisable to children with HIV and they might not be generalisable to all settings because most included studies were done in settings with high tuberculosis prevalence. Test performance might also vary in the context of regular screening. Finally, although calculations based on a hypothetical cohort give insight into consequences of testing, they were often based on heterogenous results.

Findings from this study have informed the updated 2021 WHO tuberculosis screening guidelines in PLHIV.[185] Compared with W4SS, CRP reduces the need for additional rapid diagnostic tests without compromising sensitivity, but there was a paucity of data for outpatients on ART. In outpatients not on ART, CRP assays could be used as a standalone screening test or combined with W4SS in a sequential strategy. In outpatients on ART, chest x-ray could be used in parallel with W4SS, depending on available resources, because this strategy detects more tuberculosis cases than does W4SS alone. Overall, the WHO-recommended screening and diagnostic algorithm (W4SS followed by Xpert) has suboptimal sensitivity; Xpert for all offers small improvements in sensitivity and would be resource intensive. Future research is needed to assess the utility of CRP screening in outpatients on ART and Xpert Ultra in all PLHIV, and to investigate the cost-effectiveness of different screening tests and strategies. Because no test or strategy met both WHO-defined minimum

sensitivity and specificity thresholds, improved screening tests for tuberculosis need to be developed for this population.

2.6 Contributors

AD, YHam, APK, ADK, MXR, TKr, AB, CM, SSi, DAB, GMe, and GMa designed the study and protocol and interpreted the results. GMa supervised the study. AD and YHam did the systematic review. ADK, MXR, YHam, ADG, KF, DA, CSM, APW, CY, AC, CJH, NM, ETM, MSS, TTB, SSk, BWPR, GT, GN, SM, JC, SSw, REC, FAK, AAH, RW, SST, MMK, JH, PKD, AES, TKu, GC, DTN, EAG, SB, ISJ, JKG, DJH, SML, HAAA, AK, RRK, NT, and GMe contributed data to the meta-analysis. AD analysed the data with assistance from APK, DAB, and YHam. AD and GMa wrote the first draft of the manuscript, which was revised based on comments from co-authors. AD, DAB, and YHam accessed and verified the data. All authors approved the final version of the manuscript.

CHAPTER 3: TUBERCULOSIS SCREENING AMONG HIV-POSITIVE INPATIENTS: A SYSTEMATIC REVIEW AND INDIVIDUAL PARTICIPANT DATA META-ANALYSIS

This study is published in *The Lancet HIV*:

Dhana AV, Hamada Y, Kengne AP, Kerkhoff AD, Rangaka MX, Kredo T, Baddeley A, Miller C, Gupta-Wright A, Fielding K, Wood R, Huerga H, Rücker SCM, Heidebrecht C, Wilson D, Bjerrum S, Johansen IS, Thit SS, Kyi MM, Hanson J, Barr DA, Meintjes G, Maartens G.

"Tuberculosis screening among HIV-positive inpatients: a systematic review and individual participant data meta-analysis"
The Lancet HIV. 2022 Apr;9(4):e233-e241
https://doi.org/10.1016/S2352-3018(22)00002-9

3.1 Abstract

Background Since 2011, WHO has recommended that HIV-positive inpatients be routinely screened for tuberculosis with the WHO four-symptom screen (W4SS) and, if screened positive, receive a molecular WHO-recommended rapid diagnostic test (eg, Xpert MTB/RIF [Xpert] assay). To inform updated WHO tuberculosis screening guidelines, we conducted a systematic review and individual participant data meta-analysis to assess the performance of W4SS and alternative screening tests to guide Xpert testing and compare the diagnostic accuracy of the WHO Xpert algorithm (ie, W4SS followed by Xpert) with Xpert for all HIV-positive inpatients.

Methods We searched MEDLINE, Embase, and Cochrane Library from Jan 1, 2011, to March 1, 2020, for studies of adult and adolescent HIV-positive inpatients enrolled regardless of tuberculosis signs and symptoms. The separate reference standards were culture and Xpert. Xpert was selected since it is most likely to be the confirmatory test used in practice. We assessed the proportion of inpatients eligible for Xpert testing using the WHO algorithm; assessed the accuracy of W4SS and alternative screening tests or strategies to guide diagnostic testing; and compared the accuracy of the WHO Xpert algorithm (W4SS followed by Xpert) with Xpert for all. We obtained pooled proportion estimates with a random-effects model, assessed diagnostic accuracy by fitting random-effects bivariate models, and assessed diagnostic yield descriptively. This systematic review has been registered on PROSPERO (CRD42020155895).

Findings Of 6162 potentially eligible publications, six were eligible and we obtained data for all of the six publications (n=3660 participants). The pooled proportion of inpatients eligible for an Xpert was 90% (95% CI 89–91; n=3658). Among screening tests to guide diagnostic testing, W4SS and C-reactive protein (≥5 mg/L) had highest sensitivities (≥96%) but low specificities (≤12%); cough (≥2 weeks), haemoglobin concentration (<8 g/dL), body-mass index (<18·5 kg/m2), and lymphadenopathy had higher specificities (61–90%) but suboptimal sensitivities (12–57%). The WHO Xpert algorithm (W4SS followed by Xpert) had a sensitivity of 76% (95% CI 67–84) and specificity of 93% (88–96; n=637). Xpert for all had similar accuracy to the WHO Xpert algorithm: sensitivity was 78% (95% CI 69–85) and specificity was 93% (87–96; n=639). In two cohorts that had sputum and non-sputum samples collected for culture or Xpert, diagnostic yield of sputum Xpert was 41–70% and 61–64% for urine Xpert.

Interpretation The W4SS and other potential screening tests to guide Xpert testing have suboptimal accuracy in HIV-positive inpatients. On the basis of these findings, WHO now strongly recommends molecular rapid diagnostic testing in all medical HIV-positive inpatients in settings where tuberculosis prevalence is higher than 10%.

3.2 Introduction

Tuberculosis is the leading cause of hospital admission and in-hospital deaths in people living with HIV.[186] In a meta-analysis of autopsy studies among people living with HIV, almost 50% of tuberculosis-related deaths were undiagnosed at autopsy.[12] The diagnosis of tuberculosis among HIV-positive inpatients is challenging. HIV-positive inpatients are typically severely immune suppressed with disseminated or extrapulmonary tuberculosis, might have a non-specific clinical presentation, and often produce paucibacillary specimens.[23] Furthermore, a large proportion (31–63%) of inpatients are unable to produce sputum for diagnostic testing.[18,19,25]

Since 2011, WHO has recommended that people living with HIV (including HIV-positive inpatients) be routinely screened for tuberculosis with the WHO four-symptom screen (W4SS; comprising current cough, fever, night sweats, or weight loss);[82] if the W4SS is positive, an inpatient should then receive a molecular WHO-recommended rapid diagnostic test (eg, Xpert MTB/RIF [Xpert] or Xpert MTB/RIF Ultra [Xpert Ultra]).[187]

Rapid tuberculosis diagnostic testing with Xpert in all HIV-positive inpatients in high-burden settings might be more appropriate than pre-screening with the W4SS to assess eligibility for Xpert testing. The W4SS was developed following an individual participant data meta-analysis in ambulatory patients with HIV.[82] However, W4SS might have low specificity in inpatients since HIV-related opportunistic diseases often present with one or more of the W4SS symptoms.[141] Furthermore, the diagnostic accuracies of alternative screening tests or strategies are not well known.

We conducted a systematic review and individual participant data meta-analysis of HIV-positive inpatients admitted to hospital irrespective of signs and symptoms of tuberculosis to inform updated WHO tuberculosis screening guidelines.[185] First, we calculated the proportion of inpatients eligible for rapid tuberculosis diagnostic testing with Xpert using the WHO algorithm (W4SS followed by Xpert). Second, we assessed the diagnostic accuracy of the W4SS and other tuberculosis screening tests or strategies to guide diagnostic testing. Third, we compared the diagnostic accuracy of the WHO Xpert algorithm (W4SS followed by Xpert) with Xpert for all inpatients. Fourth, we calculated the diagnostic yield of Xpert (ie, proportion of total tuberculosis cases with a positive Xpert test).

3.3 Methods

Search strategy and selection criteria

We used similar methods to our recent individual participant data meta-analysis on tuberculosis screening among people living with HIV who were in ambulatory care.[188] Two authors (AD and YH) independently selected studies, extracted data, and assessed study quality. Disagreements were resolved by discussion.

We updated the search of the systematic review by Hamada and colleagues,[83] who searched PubMed (MEDLINE), Embase, Cochrane Library, and conference abstracts from Jan 1, 2011 (the year WHO first recommended the W4SS be used), to March 12, 2018 (appendix Table 8-13). We re-reviewed all potential full-texts from Hamada and colleagues[83] to identify eligible studies. We also applied the same search strategy from Hamada and colleagues[83] for articles published between March 12, 2018, and March 1, 2020. We also reviewed reference lists of reviews and included articles, and we contacted experts for unpublished studies.

We reviewed titles and abstracts from the search and reviewed full-texts of potentially eligible articles. We included cross-sectional studies, observational studies, and randomised trials that collected at least one sputum sample for tuberculosis culture or Xpert among adult and adolescent (aged 10 years or older) inpatients who were HIV-positive and who were enrolled regardless of tuberculosis signs and symptoms (but with data on W4SS). The target condition was active tuberculosis.

The two separate reference standards were bacteriological confirmation of *Mycobacterium tuberculosis* with culture of sputum or other samples, or Xpert of sputum or other samples. Xpert was selected post hoc despite its suboptimal sensitivity because it is recommended by WHO and is the most used confirmatory test in practice (as opposed to culture). WHO recommends assessing screening or triage tests against currently recommended confirmatory tests.[79] Only studies that collected culture contributed to the WHO guidelines[185] with two additional studies added to this meta-analysis after guideline development.[18,189]

We included primary datasets that allowed us to compare the W4SS with alternative screening tests and strategies, and the WHO Xpert algorithm (W4SS followed by Xpert) with rapid tuberculosis diagnostic testing in all HIV-positive inpatients using Xpert. In this Article, a screening test was defined as a test done to assess whether an inpatient requires additional testing for bacteriological confirmation of tuberculosis (eg, with a rapid molecular diagnostic

test or culture) and a diagnostic test was defined as a test that would provide bacteriological confirmation. The systematic screening tests we examined were the W4SS, C-reactive protein concentration (CRP; 10 mg/L [considered the upper limit of normal],[88] 5 mg/L, and 8 mg/L thresholds), chest x-ray, cough lasting 2 weeks or more, haemoglobin concentration (<10 g/dL and <8 g/dL), body-mass index (<18.5 kg/m²), and lymphadenopathy. We also examined several parallel strategies (two screening tests offered at the same time) to improve sensitivity and sequential strategies (second screening test offered only if the first screening test is positive) to improve specificity. Finally, the systematic rapid tuberculosis diagnostic tests that we examined were Xpert and Xpert Ultra (although no included study assessed Xpert Ultra).

We excluded studies that had a case-control design, that only recruited HIV-positive inpatients with presumptive tuberculosis, and that recruited participants who were already on tuberculosis treatment or currently diagnosed with active tuberculosis.

We have reported our findings according to PRISMA-IPD and PRISMA-DTA statements.[148,149] The protocol for this study was approved by the University of Cape Town human research ethics committee. For each included study, participants gave written informed consent and investigators obtained ethics committee approval.

Data extraction, study quality, and individual participant data synthesis

Using a standardised data extraction form, we extracted study-level information on first author, publication year, study period, country, setting, exclusion criteria, study design, type of participants, and method of tuberculosis diagnosis. To assess quality of studies included in proportion meta-analyses, we modified a tool designed to assess study quality in systematic reviews addressing prevalence measures.[190] To assess quality of diagnostic test accuracy studies, we used the Quality Assessment of Diagnostic Accuracy Studies-2 tool to assess patient selection, index test, reference standard, and flow and timing.[150]

We emailed authors of eligible datasets with an invitation to contribute individual participant data. After consultation with WHO and our study group, we prespecified variables to be collected (appendix Table 8-14). We standardised individual participant data and then synthesised a single dataset. We excluded study participants younger than 10 years and considered contaminated cultures as negative. We ensured individual participant data integrity for each dataset by checking information against study publications and checking for

missing, duplicate, invalid, and implausible items.[151,152] We contacted the corresponding authors of each study to resolve discrepancies.

Data analyses

We analysed individual participant data in two-stages. Individual participant data were first analysed separately in each study and reduced to aggregate data, which we then pooled using meta-analytical techniques.

First, we estimated tuberculosis prevalence, proportion of inpatients eligible for Xpert testing according to the WHO algorithm (ie, proportion of inpatients with a positive W4SS), and measures of diagnostic performance (eg, sensitivity and specificity) for individual studies. Second, we pooled tuberculosis prevalence and proportion of inpatients eligible for Xpert testing using a generalised linear mixed model with logit transformation.[153] We assessed heterogeneity with Cochran's Q test and I^2 statistic.[154] We pooled measures of diagnostic performance (sensitivity and specificity) in a bivariate generalised linear mixed model.[137] For these analyses, we excluded HIV-positive inpatients with no data on the reference standard or index test. When there were fewer than four studies or non-convergence, we assumed no correlation between measures of sensitivity and specificity to simplify the model.[155] When all studies had 100% sensitivity or specificity, we computed binomial 95% CIs. We showed the absolute pooled sensitivity and specificity using summary receiver-operating characteristic curves.[156] To compare test accuracy, we did indirect comparisons (based on all studies that evaluated at least one of the tests of interest) and direct comparisons (based on studies that evaluated both tests of interest). For direct comparisons, we did a bivariate meta-regression with test-type as a covariate. Due to the variation of tuberculosis prevalence across studies, we applied pooled accuracy estimates to a hypothetical cohort of 1000 individuals for each screening test or strategy using Bayes' theorem. We also calculated the diagnostic yield of Xpert using a post-hoc analysis; diagnostic yield of Xpert was defined as the proportion of total microbiologically confirmed tuberculosis cases (using culture or Xpert) with a positive diagnostic test.

In sensitivity analyses, we assessed diagnostic accuracy using two other reference standards of combined culture or Xpert, and combined culture or Xpert among datasets that collected sputum for culture. We did not explore heterogeneity with meta-regression or assess for publication bias since few studies were included in each analysis. All meta-analyses were done using *lme4, altmeta, meta, metafor,* and *mada* packages in R (version 3.6.1).

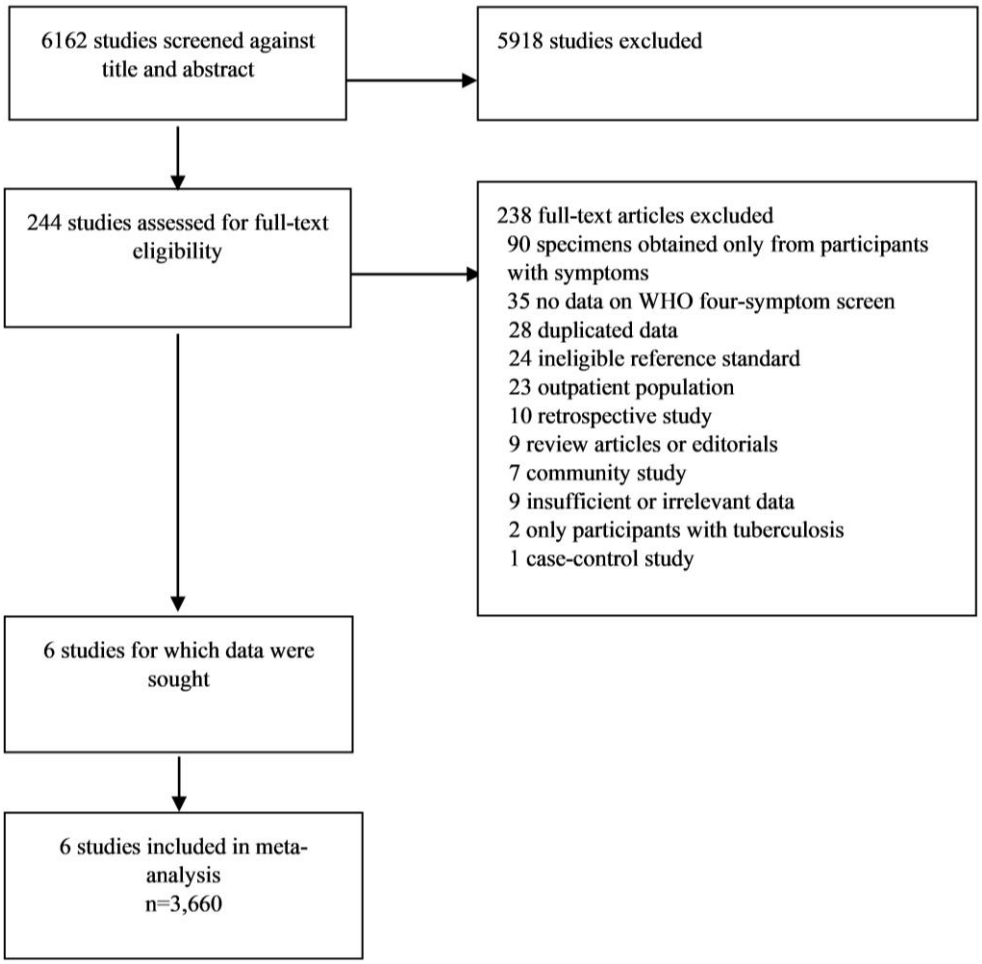
This systematic review has been registered on PROSPERO (CRD42020155895).

Role of the funding source
The funder (WHO) had a role in study design; data collection, analysis, and interpretation; and writing the report.

3.4 Results

Of 6162 publications found, six were eligible, and individual participant data were obtained for all six studies (n=3660; Figure 3-1).[18,20,174,189,191,192] The characteristics of included studies are shown in the appendix (Table 8-15). The included studies collected data from 2012 to 2017. All studies recruited inpatients from medical wards (one study recruited from an infectious disease ward). Five studies were done in sub-Saharan Africa and one in Asia. Studies systematically collected sputum for culture (four studies), sputum for Xpert (six studies), and urine for Xpert (two studies). We judged risk of bias for six studies that contributed to the proportion meta-analysis (appendix Table 8-16). For the response rate domain, risk of bias was judged to be high for two studies that had a response rate of less than 80%. We judged risk of bias for four studies that contributed to the diagnostic meta-analysis with culture as reference standard (appendix Table 8-17). For the reference test domain, risk of bias was judged to be high for three studies that did not obtain extrapulmonary samples for testing. The appendix (Table 8-18) shows missing data by study. In three studies that did not exclude participants who could not produce sputum samples,[18,20,189] sputum Xpert was missing for 35–54% of participants, mostly because inpatients were unable to produce a sputum sample, whereas urine Xpert was missing for 2% or less participants.

Figure 3-1: Study selection



Participant characteristics overall are shown in Table 3-1 and by study in the appendix (Table 8-19). The median age of participants was 37 (IQR 31–45) years, 2104 (58%) of 3659 participants were women, and 2445 (67%) of 3642 participants were receiving ART. The median CD4 count was 205 (IQR 66–408) cells per μL. We did not collect data on ethnicity.

Table 3-1: Summary of main characteristics of participants

Variable	Count (%) or median (IQR)	N†
Participants	3660 (100)	
Demographics		
Age (years)	37 (31-45)	3660
Female	2104 (58)	3659
HIV history		
On ART	2445 (67)	3642
CD4 count (cells/μL)	205 (66-408)	3479
CD4 <=200 cells/μL	1709 (49)	3479
Clinical characteristics		
History of tuberculosis	902 (28)	3268
W4SS*	3306 (90)	3658
Cough	1945 (53)	3655
Fever	1969 (54)	3652
Weight loss	2638 (72)	3651
Night sweats	1490 (41)	3652
Cough >= 2 weeks	765 (24)	3172
Lymphadenopathy	58 (11)	508
Tuberculosis diagnostic tests		
Total Xpert positive**	401 (14)	2957
Total culture positive**	157 (23)	674
Imaging and laboratory tests		
CXR (abnormal)	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

†Participants with data available for variable

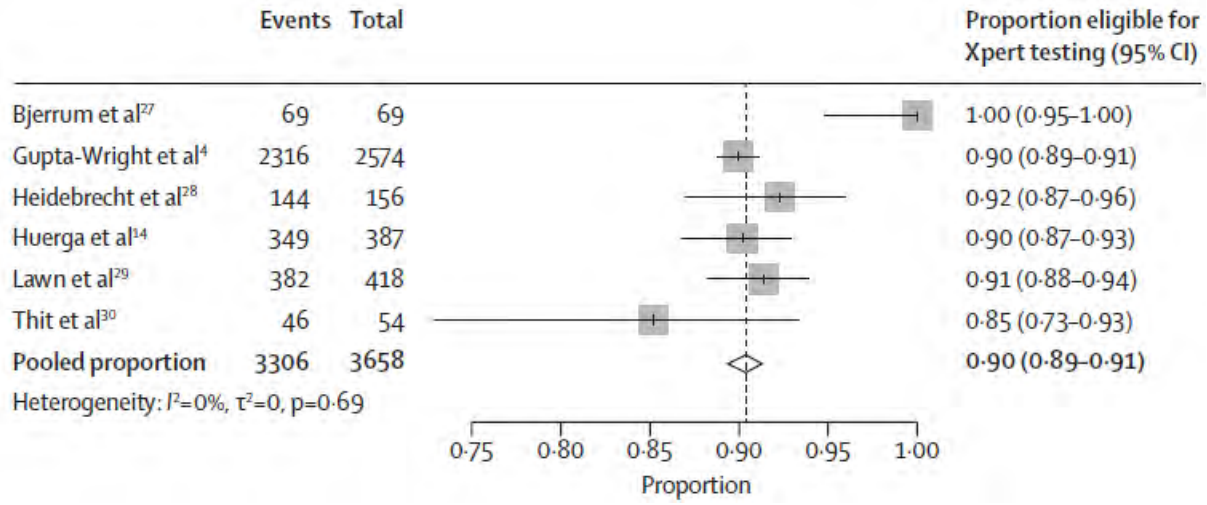
*W4SS defined as one or more of the following: current cough, fever, night sweats, or weight loss

**Sputum and/or non-sputum 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

Among the four studies that collected sputum for culture, the pooled tuberculosis prevalence was 20% (95% CI 13–28; n=674) with culture as a reference standard and 25% (18–33; n=699) with culture or Xpert as a reference standard. Among six studies, the pooled proportion of inpatients with a positive W4SS (i.e., inpatients eligible for Xpert testing according to the WHO algorithm) was 90% (89–91; n=3658); proportion estimates for individual studies ranged from 85% to 100% (Figure 3-2).

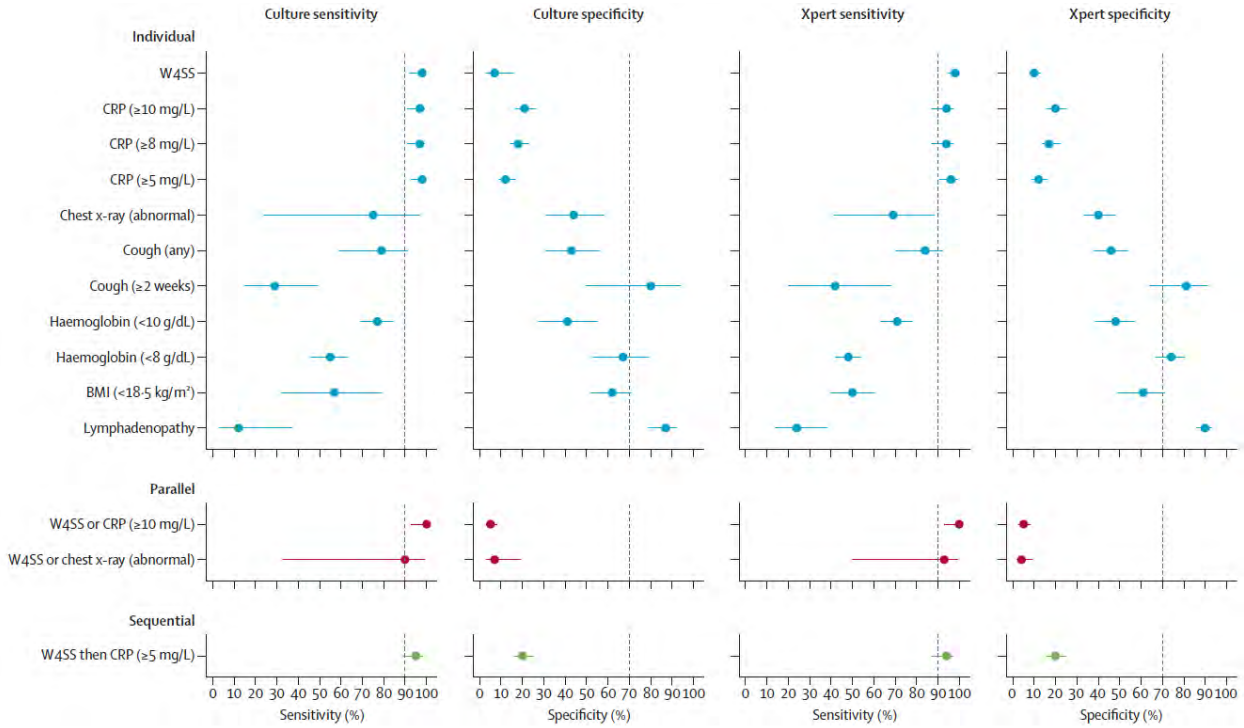
Figure 3-2: Random-effects meta-analysis of proportion of HIV-positive inpatients with positive WHO four symptom screen (ie, proportion eligible for Xpert according to WHO algorithm)



Xpert=Xpert MTB/RIF.

Plots of sensitivity and specificity for each screening test or strategy are shown in Figure 3-3. Indirect comparisons are shown in Table 3-2. For individual tests, the sensitivities of W4SS and CRP (≥5 mg/L) were highest, but the specificities were low. Cough (≥2 weeks), haemoglobin concentration (<8 g/dL), body-mass index (<18·5 kg/m²), and lymphadenopathy had moderate to high specificities, but low sensitivities, making them unsuitable to be explored as screening tests. Data on chest x-ray was sparse. In strategies that combined W4SS with CRP concentration, sensitivities were high, but specificities were low.

Figure 3-3: Pooled sensitivity and specificity along with 95% CIs for each screening test or strategy for the detection of tuberculosis using reference standards of culture or Xpert



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. Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity). BMI=body-mass index. CRP=C-reactive protein. W4SS=WHO four-symptom screen. Xpert=Xpert MTB/RIF.

Table 3-2: Indirect comparisons of the diagnostic accuracy (pooled sensitivity and specificity) for each screening test or strategy for the detection of tuberculosis using reference standards of culture or Xpert

	Culture				Xpert†			
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)
Screening test/strategy								
W4SS	4	672	98 (92-99)	7 (3-16)	6	2176	98 (95-99)	10 (8-13)
CRP (>=10 mg/L)	1	400	97 (91-99)	21 (17-26)	1	395	94 (87-97)	20 (16-25)
CRP (>=8 mg/L)	1	400	97 (91-99)	18 (14-23)	1	395	94 (87-97)	17 (14-22)
CRP (>=5 mg/L)	1	400	98 (93-100)	12 (9-17)	1	395	96 (91-99)	12 (9-16)
CXR (abnormal)	1	52	75 (24-97)	44 (31-58)	2	176	69 (41-88)	40 (33-48)
Cough (any)	4	669	79 (59-91)	43 (31-56)	6	2173	84 (70-92)	46 (38-54)
Cough (>=2 weeks)	3	608	29 (15-49)	80 (50-94)	4	1860	42 (20-68)	81 (64-91)
Hb (<10 g/dL)	3	527	77 (69-84)	41 (28-55)	5	2015	71 (63-78)	48 (39-57)
Hb (<8 g/dL)	3	527	55 (46-63)	67 (53-79)	5	2015	48 (42-54)	74 (67-80)
BMI (<18.5 kg/m²)	2	112	57 (32-79)	62 (52-71)	4	1553	50 (40-60)	61 (49-71)
Lymphadenopathy	2	123	12 (3-37)	87 (79-92)	3	337	24 (14-38)	90 (86-93)
W4SS or CRP (>=10 mg/L)¶	1	399	100 (93-100)	5 (3-8)	1	394	100 (93-100)	5 (3-8)
W4SS or CXR (abnormal)¶	1	52	90 (33-99)	7 (3-19)	2	176	93 (50-99)	4 (2-9)
W4SS then CRP (>=5 mg/L)¶	1	399	95 (89-98)	20 (16-25)	1	394	94 (87-97)	20 (16-25)
Algorithm††								
WHO Xpert algorithm§*	4	637	76 (67-84)	93 (88-96)	-	-	-	-
Xpert alone*	4	639	78 (69-85)	93 (87-96)	-	-	-	-

†In one study by Gupta-Wright (2018), only the intervention arm was included since sputum Xpert and urine Xpert were available, while in the standard of care arm urine Xpert was unavailable and sputum Xpert was only available for 779/1287 (61%) of participants.

¶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

††For Xpert alone, the comparator is the WHO Xpert algorithm

§According to WHO Xpert algorithm, Xpert testing 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).

*Accuracy measures for entire algorithm using sputum and/or urine Xpert result. Alternative algorithms are W4SS then single sputum Xpert (4 studies; 375 participants; sensitivity 78 (57-91), specificity 97 (94-99), single sputum Xpert alone (4 studies; 375 participants; sensitivity 78 (55-91), specificity 97 (93-99), and urine Xpert alone (1 study; 411 participants; sensitivity 59 (50-68), specificity 91 (88-94).

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

Direct comparisons of individual tests were mostly similar to indirect comparisons (appendix Table 8-20). Forest plots and summary receiver operating characteristics curves are provided in the appendix (Figures 8-6 and 8-7). The appendix (Table 8-21) shows how estimates for each test or strategy affect a hypothetical cohort of 1000 HIV-positive inpatients at different tuberculosis prevalences. No individual test offered an optimal trade-off between tuberculosis cases missed and Xpert tests required (appendix Figure 8-8). In sensitivity analyses using alternative reference standards, results were largely similar to the main analyses (appendix Table 8-22).

The sensitivity of the WHO Xpert algorithm (W4SS followed by Xpert) was 76% (95% CI 67–84) and specificity was 93% (88–96; n=637; Table 3-2). The diagnostic accuracy of Xpert

for all was similar to the WHO Xpert algorithm— sensitivity was 78% (69–85) and specificity was 93% (87–96; n=639). In a hypothetical cohort of 1000 HIV-positive inpatients at 20% tuberculosis prevalence, the WHO Xpert algorithm would result in 940 Xpert tests, but miss 48 tuberculosis cases; Xpert for all 1000 HIV-positive inpatients would miss 44 tuberculosis cases (appendix Table 8-21).

The appendix (Table 8-23) shows diagnostic yield using different diagnostic tests and sample types. In one cohort that collected sputum and non-sputum samples for Xpert and culture,[20] sputum Xpert diagnosed only 57 (41%) of all 139 tuberculosis cases (195 [46%] of 420 inpatients were unable to produce a sputum sample), whereas combined concentrated and unconcentrated urine Xpert diagnosed 89 (64%) of all 139 cases; concentrating urine increased diagnostic yield over not concentrating urine, from 42% to 59%. Sputum Xpert combined with urine Xpert diagnosed 116 (83%) of all 139 cases in the same cohort. In one cohort that collected sputum Xpert and concentrated urine Xpert, sputum Xpert diagnosed 85 (70%) of 122 tuberculosis cases and urine Xpert diagnosed 74 (61%) of 122 cases.[18] Across all studies, Xpert was positive in six (2%) of 251 inpatients who had available Xpert results but were ineligible for Xpert testing according to the WHO Xpert algorithm.

3.5 Discussion

In this individual participant data meta-analysis, we found that almost all HIV-positive inpatients in high-burden settings were eligible for Xpert testing using the WHO algorithm. W4SS and CRP concentration (≥5 mg/L) had the highest sensitivities of all screening tests evaluated to guide diagnostic testing, but specificities were low. Other screening tests had low sensitivities or wide 95% CIs. The WHO screening and diagnostic algorithm (ie, W4SS followed by Xpert) had a sensitivity of 76%; Xpert for all inpatients had similar sensitivity (78%). On the basis of these findings, WHO has made a strong recommendation to do molecular rapid diagnostic testing in all HIV-positive inpatients in high-burden settings (>10% tuberculosis prevalence).

We found that all screening tests and strategies to guide additional diagnostic testing fell short of WHO-defined minimum thresholds (90% sensitivity and 70% specificity) in this population.[79] The specificity of W4SS in our study was only 7–10%, which is substantially lower than its specificity among outpatients on ART (71%) and not on ART (37%).[188] The low specificity of CRP concentration for tuberculosis in this cohort is consistent with findings

from other cohorts of symptomatic HIV-positive inpatients.[193,194] By contrast, CRP concentration has shown an improved specificity (67%) over W4SS in unselected outpatients not on ART.[188] The low specificities of W4SS and CRP concentration are likely to be due to the high prevalences of other opportunistic diseases in patients without tuberculosis in this patient population. Both W4SS and CRP concentration met the minimum WHO sensitivity threshold of 90%. However, even at this high sensitivity, the W4SS and CRP concentration would miss roughly one in 50 tuberculosis cases. Given the high tuberculosis prevalence and the high mortality associated with missed diagnosis in inpatients in such settings, a small loss in sensitivity might be unacceptable.

Sputum Xpert alone had a low yield because many inpatients had difficulty producing sputum for testing. Urine-based Xpert testing might have an important role in diagnosing inpatients who are unable to produce sputum. We found that 35–54% of participants could not produce sputum for Xpert testing. In one cohort, sputum Xpert combined with urine Xpert diagnosed 83% of all cases, whereas sputum Xpert diagnosed only 41% of cases.[20] In the same cohort, urine Xpert had a higher yield than sputum Xpert,[20] but the opposite was true in another cohort.[18]

There are several reasons to consider rapid diagnostic testing for tuberculosis with Xpert in all HIV-positive inpatients. First, since almost all inpatients with HIV met WHO eligibility requirements for Xpert testing, universal testing might reduce diagnostic complexity. Second, we found that Xpert was positive in 2% of HIV-positive inpatients who did not meet eligibility for testing. Third, in the real world, not all HIV-positive inpatients who qualify for Xpert testing might ultimately receive a test; for example, two of the included studies in this meta-analysis reported a positive W4SS in 90% or more HIV-positive inpatients, but clinicians identified only 38–64% as having possible tuberculosis after clinical assessment.[18,189] Fourth, since the W4SS is also used to assess eligibility for lateral flow urine lipoarabinomannan assay (LF-LAM) in HIV-positive inpatients, both routine Xpert and LF-LAM diagnostic testing might also be considered in this population. For instance, the STAMP trial showed a reduction in all-cause mortality among unselected HIV-positive inpatients when routine LF-LAM and urine Xpert were done in addition to routine sputum Xpert.[18] Combined use of Xpert and LF-LAM has also been shown to improve diagnostic yield over either test alone in tuberculosis bloodstream infection, which predicts mortality.[195] By contrast, obtaining Xpert samples in all HIV-positive inpatients might have a negative

effect on infection control, human resources, laboratory capacity, and cost. However, since almost all inpatients already qualify for Xpert testing using the WHO criteria, Xpert testing in all inpatients would have a small effect on costs.

Although our findings support universal Xpert testing, this strategy would still miss more than 20% of culture-positive cases. Thus, aside from LF-LAM, additional diagnostic approaches that incorporate clinical symptoms and signs, radiological tests (eg, chest x-ray and abdominal ultrasound), and laboratory tests (e.g., haemoglobin concentration) still have an important role in inpatients with a negative Xpert test.[92,121,196,197] Newer technologies might also substantially close this diagnostic gap. For example, Xpert Ultra and Fujifilm SILVAMP TB-LAM (FujiLAM) have shown increased sensitivity compared with Xpert and current LF-LAM tests.[43,48,75] In a recent systematic review, Xpert Ultra increased sensitivity over Xpert by 13% (88% *vs* 75%) in sputum samples from people living with HIV.[43] However, Xpert Ultra's lower specificity might have implications for universal Xpert Ultra testing because inpatients without tuberculosis could be classified as having tuberculosis.

Our study has limitations. First, most data were acquired in sub-Saharan Africa; the generalisability of this study to other geographical regions and low tuberculosis prevalence settings is unclear. Second, although we obtained and included data for all published studies identified by our search, some screening tests had wide 95% CIs because of sparse data. This limitation highlights the need for additional diagnostic accuracy studies among HIV-positive inpatients irrespective of tuberculosis signs and symptoms. Furthermore, no study evaluated Xpert Ultra and we did not assess other molecular WHO-recommended rapid diagnostic tests.[198] Third, some studies only included participants able to produce sputum and did not collect extrapulmonary samples for culture or Xpert; two studies also did not collect culture samples. Since inpatients often present with extrapulmonary or disseminated tuberculosis and produce paucibacillary sputum samples, the reference standard in these studies might have introduced bias, underestimating the specificity and overestimating the sensitivity of existing algorithms. However, our results were consistent across several reference standards: culture, combinations of culture and Xpert, and Xpert (which is the currently recommended confirmatory test). Furthermore, estimates of the proportion of inpatients eligible for Xpert were based on data with higher methodological quality, since these analyses did not require a reference standard. Fourth, the small number of included studies precluded investigation of heterogeneity. Fifth, tuberculosis prevalence estimates in this Article are likely to be

underestimates because of the limitations of our reference standard. Sixth, disseminated disease is more common at low CD4 counts, but we did not do analyses by CD4 count. However, HIV-positive inpatients typically present with advanced immunosuppression, and disseminated disease is not uncommon at higher CD4 counts in this population.[20] Finally, our calculations for a hypothetical cohort should be treated with caution because they were based on diagnostic accuracy results derived from few participants, some of whom had an imperfect reference standard done.

In conclusion, our findings have informed the 2021 WHO recommendation to do molecular rapid diagnostic testing (eg, with Xpert) in all HIV-positive inpatients in high-burden settings (>10% tuberculosis prevalence).[185] More accurate initial screening tests to guide additional diagnostic testing in HIV-positive inpatients need to be developed since current screening tests have suboptimal accuracy and hospitals in resource-limited settings might be unable to do systematic diagnostic testing in all HIV-positive inpatients. Although routine molecular rapid diagnostic testing might reduce the current diagnostic gap, a negative result still does not rule out tuberculosis. Xpert Ultra could additionally bridge the diagnostic gap and requires evaluation in unselected HIV-positive inpatients.

3.6 Contributors

AD, YH, APK, ADK, MXR, TK, AB, CM, DAB, GMe, and GMa designed the study and protocol and interpreted the results. GaM supervised the study. AD and YH did the systematic review. ADK, AG-W, KF, RW, HH, SCMR, CH, DW, SB, ISJ, SST, MMK, and JH contributed data to the meta-analysis. AD analysed the data with assistance from APK, DAB, and YH. AD and GMa wrote the first draft of the manuscript, which was revised on the basis of comments from coauthors. AD, DAB, and YH accessed and verified the data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. All authors approved the final version of the manuscript.

Chapter 4: 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

This study is published in *The Journal of Infection*:

Dhana AV, Hamada Y, Kengne AP, Kerkhoff AD, Broger T, Denkinger CM, Rangaka MX, Gupta-Wright A, Fielding K, Wood R, Huerga H, Rücker SCM, Bjerrum S, Johansen IS, Thit SS, Kyi MM, Hanson J, Barr DA, Meintjes G, Maartens G

"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"
The Journal of Infection. Jul;85(1):40-48.
https://doi: 10.1016/j.jinf.2022.05.010.

4.1 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 followed by AlereLAM if screen positive) with AlereLAM and FujiLAM (a novel LF-LAM test) testing in all HIV-positive inpatients.

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 testing 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), hemoglobin (<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 88% and 96%, respectively. In 2 studies that collected sputum and non-sputum samples for Xpert and/or culture, diagnostic yield of sputum Xpert was 40–41%, 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 61%, 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.

4.2 Introduction

Tuberculosis is the leading cause of hospitalization among people living with HIV (PLHIV) and is responsible for nearly 40% of in-hospital deaths.[12,186] Almost 50% of tuberculosis is undiagnosed at autopsy in PLHIV.[12] 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.[18,19,23,25]

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).[198] Although AlereLAM has only moderate sensitivity,[33] it reduced mortality in inpatients in randomized trials.[18,73] The novel Fujifilm SILVAMP TB-LAM (FujiLAM) test is more sensitive than AlereLAM in inpatients.[74,75] 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%.[185,187] 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.[33,187]

The WHO screening criteria to guide AlereLAM testing may be challenging to implement in busy inpatient settings[142] and its diagnostic accuracy is unknown. The W4SS, which was developed among ambulatory PLHIV,[82] has low specificity for diagnosis of tuberculosis in inpatients.[199] 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.[187] 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.

4.3 Methods

Our findings are reported in accordance with PRISMA-IPD and PRISMA-DTA statements.[148,149] 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;[185,188,199] 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 (appendix Table 8-24).[83] 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 randomized 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,[200] 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.[79] 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; hemoglobin; 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.[190] To assess study quality for diagnostic test accuracy, we used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[150]

We emailed authors of eligible datasets inviting them to contribute data. The appendix (Table 8-25) 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.[151,152] Discrepancies were resolved by contacting the corresponding author.

Statistical analyses

We analyzed IPD using a two-stage approach. First, we analyzed 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.[153] We assessed heterogeneity with Cochran's Q test and I^2 statistic.[154] We used a bivariate generalized linear mixed model to pool sensitivities and specificities.[137] 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.[155] We computed binomial 95% CIs by summing the numbers with disease (or no disease) across studies if all studies had 100% sensitivity or specificity.[201] We used summary receiver-operating characteristic curves to jointly illustrate absolute pooled sensitivity and specificity.[156] 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.

4.4 Results

Characteristics of primary datasets selected and prevalence of tuberculosis

We identified 5 eligible datasets (appendix Figure 8-9), and IPD was obtained for all 5 datasets (n=3504).[18,20,174,189,191] The appendix (Table 8-26) 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 (appendix Table 8-27). 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 (appendix Table 8-28). 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. The appendix (Table 8-29) shows missing data by study. In 3 studies that included participants unable to produce sputum samples,[18,20,189] LF-LAM was missing for ≤3% of inpatients, but sputum Xpert was missing for 35–54% of participants.

Tables 4-1 and the appendix (Table 8-30) 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.

Table 4-1: 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

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=3477) (Table 4-2 and appendix Figure 8-10). 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*=2961). 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*=2961), 4 (*n*=54), and 3 (*n*=3477) percentage points, respectively (appendix Table 31).

Table 4-2: 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

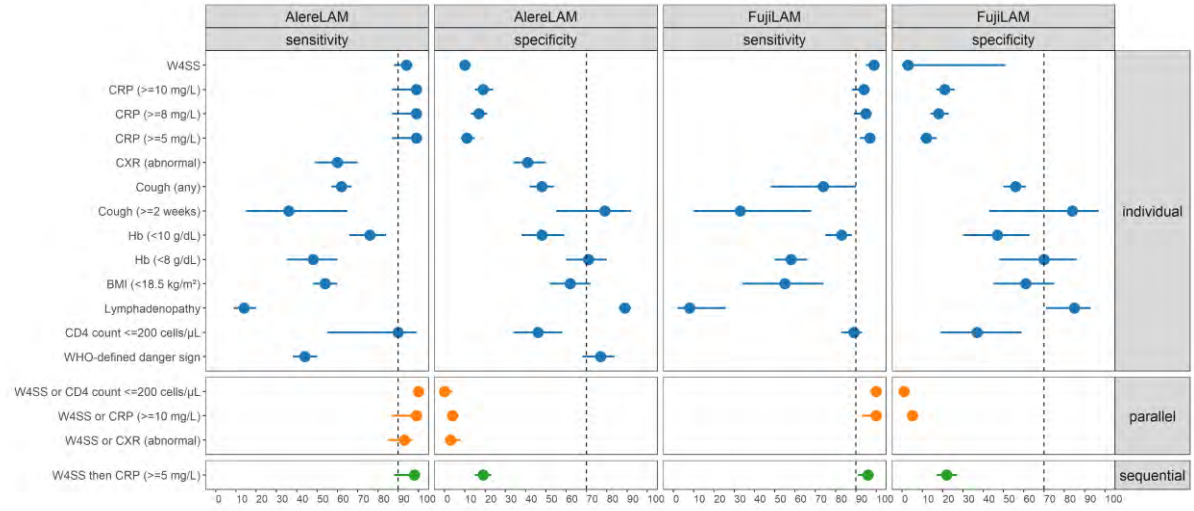
**One study by Bjerrum et al (2015) excluded from analysis as WHO stage 3 or 4 was part of inclusion criteria

Definition of abbreviations: W4SS = WHO four-symptom screen

Diagnostic performance of tuberculosis screening tests/strategies

Figure 4-1 shows plots of sensitivity and specificity of each screening test/strategy for LF-LAM positive tuberculosis, while Table 4-3 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 and 90% and specificities between 37 and 46%. Cough (≥2 weeks), hemoglobin (< 8 g/dL), lymphadenopathy, and WHO-defined danger signs had low sensitivities (14–58%) but high specificities (70–89%).

Figure 4-1: 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

Table 4-3: 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

The appendix (Figure 8-11) shows forest plots and the appendix (Figure 8-12) shows summary receiver operating characteristics curves. The point estimates for the specificities of WHO screening criteria were ≤3% in each individual study (appendix Figure 8-11). The appendix (Figure 8-13) 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=2036) (Table 4-4 and Figure 4-2); 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.

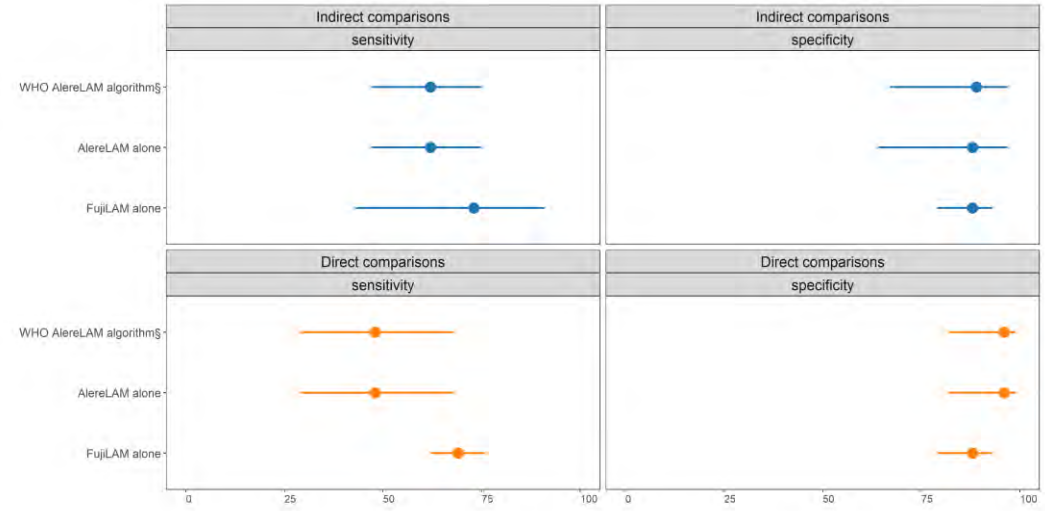
Table 4-4: Pooled sensitivity and specificity along with 95% CIs of WHO AlereLAM algorithm, AlereLAM, and FujiLAM for the detection of tuberculosis§

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 4-2: 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

Diagnostic yield of tuberculosis from different diagnostic tests and sample types

The appendix (Table 8-32) 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,[20] 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 61% 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) (appendix Table 8-33). In univariable and multivariable mixed-effects logistic regression, individual danger signs other than respiratory rate > 30 breaths/min were associated with increased risk of tuberculosis. Multivariable adjusted odds ratios were smaller compared with univariable estimates, reflecting a positive correlation between individual danger signs.

4.5 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 61% 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.[188] Conversely, several other tests (e.g., WHO-defined danger signs and low hemoglobin) 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.[65,187] 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.[79]

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, but the opposite was true in another included study.[18] In a recent study, sensitivity of urine Xpert Ultra was double that of AlereLAM (33% vs 16%).[48] 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.[51] 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.[202] Third, AlereLAM was positive in 5% of HIV-positive inpatients who did not meet WHO criteria for AlereLAM testing. Fourth, two randomized trials have demonstrated a reduction in all-cause mortality among HIV-positive inpatients with the use of AlereLAM in addition to routine diagnostics.[18,73] One trial included HIV-positive inpatients irrespective of tuberculosis signs and symptoms,[18] while the other included inpatients with a positive W4SS (which we found

was present in > 90% of HIV-positive inpatients).[73] Despite these findings, a recent survey of 24 high tuberculosis/HIV burden countries revealed that only 4 (17%) were using AlereLAM in all hospitals.[203] 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.[195]

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 61% 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.[79] 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.[69,76] Nontuberculous mycobacteria could also reduce specificity but were found in only 4% of participants with a false-positive FujiLAM test.[75] 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.[74]

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.[196] 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 extra-pulmonary 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.[79] 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.

4.6 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, 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.

CHAPTER 5: CLINICAL UTILITY OF WHO-RECOMMENDED SCREENING TOOLS, AND DEVELOPMENT AND VALIDATION OF NOVEL CLINICAL PREDICTION MODELS FOR PULMONARY TUBERCULOSIS AMONG OUTPATIENT PEOPLE LIVING WITH HIV: AN INDIVIDUAL PARTICIPANT DATA META-ANALYSIS

This study has been Formatted for submission to *European Respiratory Journal*:

Dhana AV, Gupta RK, Hamada Y, Kengne AP, Kerkhoff AD, Yoon C, Cattamanchi A, Reeve BW, Theron G, Wood R, Drain PK, Calderwood CJ, Noursadeghi M, Boyles T, Meintjes G, Maartens G, Barr DA.

"Clinical utility of WHO-recommended screening tools and development and validation of novel clinical prediction models for tuberculosis screening among outpatient people living with HIV: an individual participant data meta-analysis"

5.1 Abstract

Background: WHO recommends that people living with HIV (PLHIV) undergo tuberculosis screening with the WHO four-symptom screen (W4SS) or C-reactive protein (CRP [5 mg/L cutoff]) followed by confirmatory testing if screen positive. We conducted an individual participant data meta-analysis to determine the performance of WHO-recommended screening tools and 2 newly developed clinical prediction models (CPMs) in outpatient PLHIV.

Methods: Following a systematic review, we identified studies that recruited adult outpatient PLHIV irrespective of tuberculosis signs and symptoms or with a positive W4SS, evaluated CRP, and collected sputum for culture. We used logistic regression to develop an extended CPM (which included CRP and other predictors) and a CRP-only CPM (which only included CRP). We used internal-external cross-validation to evaluate performance.

Results: We pooled data from 8 cohorts (n=4,315 participants). The extended CPM had excellent discrimination (C-statistic 0.81); the CRP-only CPM had similar discrimination (C-statistic 0.79). The C-statistics for CRP at 5mg/L cutoff (0.70) and W4SS (0.57) were lower. For clinical utility, both CPMs had equivalent or higher net benefit compared with WHO-recommended tools. Compared with both CPMs, CRP (5mg/L cutoff) had equivalent net benefit across a clinically useful range of threshold probabilities, while W4SS had lower net benefit. The W4SS would capture 91% of tuberculosis cases and require confirmatory testing for 78% of participants. CRP (5 mg/L cutoff), the extended CPM (4.2% threshold), and the CRP-only CPM (3.6% threshold) would capture similar percentage of cases but reduce confirmatory tests required by 24%, 27%, and 36%, respectively.

Conclusions: CRP and the CPMs show utility for tuberculosis screening among outpatient PLHIV. The WHO-recommended W4SS showed suboptimal performance.

5.2 Introduction

In 2020, there were 214,000 tuberculosis deaths among people living with HIV (PLHIV).[1] Approximately half of HIV-associated tuberculosis deaths go undiagnosed,[12] and appropriate testing and treatment may avert these undiagnosed deaths. Confirmatory testing (e.g., Xpert MTB/RIF Ultra [Xpert Ultra]) for all PLHIV is often unfeasible in low-resource settings, meaning screening strategies are needed to determine who needs further confirmatory testing.

According to WHO, a tuberculosis screening tool should meet optimal (95% sensitivity, 80% specificity) or minimum (90% sensitivity, 70% specificity) performance characteristics.[79] Since 2011, WHO has recommended that outpatient PLHIV be screened for tuberculosis with the WHO four-symptom screen (W4SS) (comprising any one of current cough, fever, night sweats, or weight loss),[82] followed by confirmatory testing (e.g., Xpert Ultra) if the screen is positive. However, the specificity of W4SS is low in some subgroups (e.g., ART-naïve PLHIV),[188] resulting in large numbers of unnecessary, expensive confirmatory testing. Recently, WHO also recommended C-reactive protein (CRP) as a screening tool.[185] CRP (5mg/L cutoff) showed similar sensitivity but higher specificity than W4SS.[188] CRP can be done using a point-of-care assay at ~$2 with results in <3 minutes. The W4SS and CRP were recommended based on sensitivity and specificity, but their clinical utility, using measures such as net benefit,[109] is unknown.

Clinical prediction models (CPMs), which combine multiple predictors, may also be used for screening. Although there are few CPMs for tuberculosis screening in PLHIV,[111-114,116,145,204] these CPMs have limitations. Some have been developed using many predictors relative to number of events,[111,113,114,116] or categorized continuous variables.[111-114] Some have also shown suboptimal performance at external validation,[111,113] not undergone extensive externally validated,[112-114,116] or not been assessed for clinical usefulness.[111,113,114,116]

Using data from 8 cohorts identified following a systematic review,[188] we performed an individual participant data (IPD) meta-analysis to 1) develop and validate CPMs that incorporate CRP for active pulmonary tuberculosis among outpatient PLHIV and 2) compare the performance and clinical utility of WHO-recommended screening tools to the newly developed CPMs.

5.3 Methods

We reported our findings according to the TRIPOD and PRISMA-IPD statements.[149,205] We also adhered to additional guidelines for developing and validating CPMs in an IPD meta-analysis.[139,206,207]

Data sources and study population

We previously conducted a systematic review to compare the accuracy of several tuberculosis screening tools with the W4SS in outpatient PLHIV.[188] From that systematic review, we identified and obtained IPD for prospective cross-sectional studies, observational studies, and randomized trials conducted in facility-based, active case-finding settings that systematically measured CRP and collected sputum for culture from outpatient PLHIV regardless of signs and symptoms of tuberculosis. We only included studies that measured CRP since we aimed to assess if a multivariable modelling approach could improve CRP's performance. Active case-finding involves systematic screening of PLHIV irrespective of symptoms. Passive-case finding involves patients recognizing and seeking care for their symptoms (defined in this study as a positive W4SS). From the systematic review, we identified 4 studies (5 cohorts) from active case-finding settings,[86,88,90,173] comprising outpatient PLHIV not on ART. After contacting experts, we included 2 further studies (3 cohorts) from passive-case finding settings, comprising outpatient PLHIV receiving and not receiving ART.[44,115]

Defining the outcome of prediction models

The primary outcome was active tuberculosis, defined as culture of *Mycobacterium tuberculosis* complex from sputum.

Candidate predictors and sample size

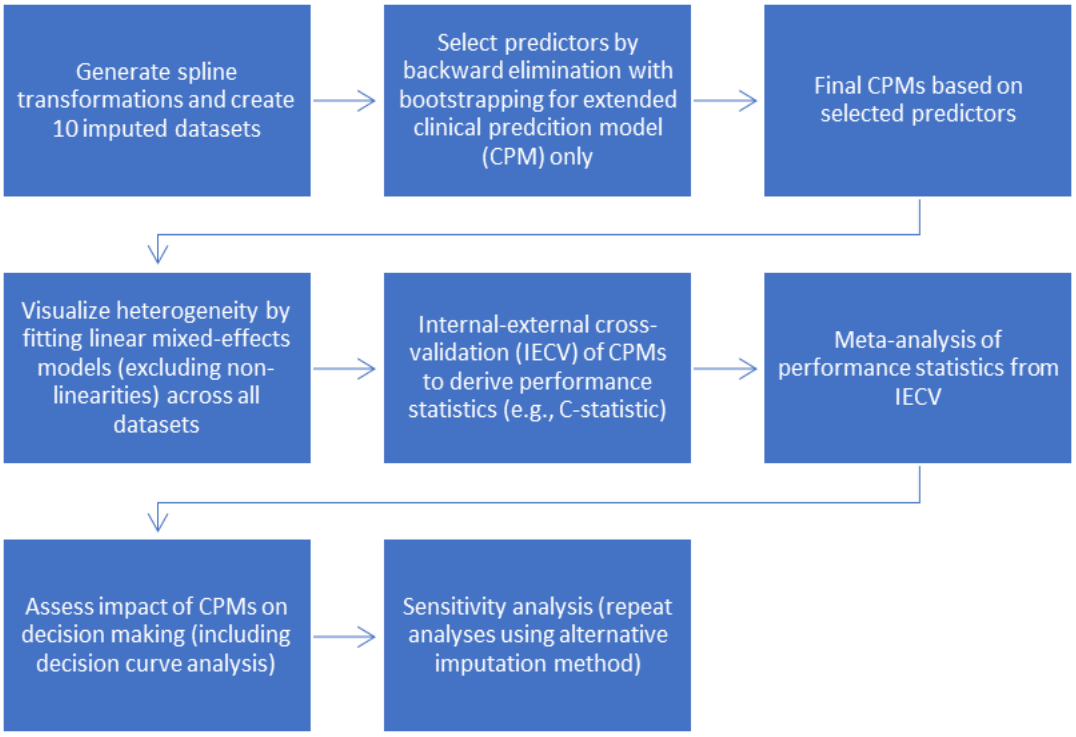
Since all included studies evaluated CRP, we assessed these studies for availability of several variables considered a priori for potential inclusion in the CPMs. We selected variables following expert clinical experience and a systematic review,[188] that are assessable immediately, and that are readily available in resource-limited settings. The clinical variables we evaluated were age, sex, W4SS components (cough, fever, night sweats, and weight loss), and body mass index (BMI). The laboratory variables we evaluated were CD4 count and CRP. The study-level variable we evaluated was case-finding setting (active vs passive case-finding).

The population derived was considered sufficient,[208] and sample size calculations are provided in the appendix.

Statistical analysis

We developed 2 CPMs: an extended CPM, which considered all candidate predictors, and a CRP-only CPM, which only included CRP as a predictor along with spline transformations. Figure 5-1 summarizes steps taken to develop and validate each CPM.

Figure 5-1: Steps taken to develop and validate each clinical prediction model



CPM development

We performed single-level multiple imputation within each cohort to deal with missing data (appendix). We created 10 imputed datasets. All further analyses were performed in each of the imputed datasets and pooled using Rubin's rules.[209]

We used a logistic regression approach for variable selection and CPM development with active tuberculosis as a binary outcome. To model continuous variables, we used restricted cubic splines with a default of 4 knots. For the extended CPM, we performed backward stepwise selection with bootstrapping to select the most predictive variables using the Akaike information criterion.[210] We kept variables retained in ≥70% of bootstrap samples and ≥5 of 10 multiply imputed datasets. This process led to a final CPM based on the selected predictors along with their corresponding estimated β coefficients and the associated intercept term. To visualize predictor heterogeneity, we fitted across all datasets linear mixed-effects models (excluding non-linearities).

Internal external cross-validation (IECV)

To assess CPM generalizability, we used IECV, which involved several steps.[139] First, the CPM is developed in all but one study. Second, the omitted study is used to externally validate the CPM and derive performance statistics. Third, this process is repeated until each study has a chance to be omitted. We performed IECV on each imputed data set and pooled performance statistics using Rubin's rules. In the omitted study, we also compared both CPMs to WHO-recommended screening tools and one other published CPM.[113] Other published CPMs were not evaluated because some predictors were not measured in some or all cohorts.[111,112,114,116,145,204]

We calculated several performance statistics. We assessed discrimination, which quantifies how well a CPM can differentiate between those that have tuberculosis and those that do not, using the C-statistic. We considered a C-statistic of ≥0.7 and ≥0.8 as acceptable and excellent performance, respectively.[108] We then assessed calibration, which refers to agreement between expected and observed outcomes, using calibration-in-the-large (value of 0 indicates perfect calibration), calibration slope (value of 1 indicates perfect calibration), and calibration plots.

We performed a univariate random-effects meta-analysis of performance statistics derived from the IECV.[211,212] To assess heterogeneity, we visually examined forest plots. We also performed a multivariate meta-analysis to pool the c-statistic and calibration slope.[212] We calculated the joint probability that the CPMs would achieve a C-statistic of >0.70, >0.75, or 0.80 and a calibration slope between 0.8 and 1.2 in future patients using bootstrapping.

We assessed clinical utility using 2 approaches. First, during IECV, we performed decision-curve analyses by pooling (stacking) multiply imputed validation datasets.[109,213] Decision curves show net benefit of screening tools over a range of clinically relevant threshold probabilities. A threshold probability is the minimum probability of disease at which further diagnostic workup would be justified.[213] Net benefit is the difference between proportion of true positives and proportion of false positives weighted by the threshold probability. We chose a threshold probability range from 0% to 20% because it is unlikely that more than 20% risk would be required before confirmatory testing is recommended.[214] The CPMs were compared with a confirmatory testing for all strategy (i.e., sputum culture for all), confirmatory testing for none strategy (i.e., sputum culture for none), WHO-recommended screening tests (W4SS and CRP [5mg/L cutoff]), and an existing CPM.[113] Second, we

assessed the trade-off between percentage of tuberculosis cases captured and percentage of participants needing confirmatory testing by applying WHO-recommended screening tests to the stacked validation cohorts during IECV and then comparing them to the newly developed CPMs at a threshold that provides similar sensitivity and 95% sensitivity (WHO optimal sensitivity requirements).[79]

In sensitivity analyses, we used an alternative imputation procedure using the aregImpute function in the rms package in R (using the function's default arguments). We did all analyses using *pmsampsize*, *mice, rms, metamisc, meta, metafor, lme, mada,* and *dcurves* packages in R (version 3.6.1)

5.4 Results

Study population

IPD were provided for all 6 eligible studies (8 cohorts). The cohorts collected data between 2010 and 2020 (Table 8-34). Six cohorts were from South Africa. Five were active case-finding cohorts and 3 were passive-case finding cohorts. Table 5-1 and the appendix (Table 8-35) show participant characteristics overall and by study, respectively. We included 4,315 participants of whom 652 (15%) had tuberculosis. Most participants (85%) were recruited from active case-finding settings. Most (91%) participants were not on ART. The appendix (Table 8-36) shows missing data by study.

Table 5-1: Summary of main characteristics for all participants

Variable	Overall†	N‡
Participants	4315 (100)	
Demographics		
Active case-finding	3667 (85)	4315
Age (years)	33.2 (27-40)	4315
Female	2381 (55)	4315
HIV history		
On ART	380 (9)	4315
CD4 count (cells/μL)	204 (93-319)	4188
CD4 <=200 cells/μL	2056 (49)	4188
Clinical characteristics		
History of tuberculosis	602 (14)	4309
Cough	2256 (52)	4314
Fever	1541 (36)	4283
Weight loss	2638 (62)	4283
Night sweats	1674 (39)	4313
Cough >= 2 weeks	1455 (34)	4306
Tuberculosis diagnostic tests		
Sputum culture +	652 (15)	4209
Laboratory tests		
BMI (kg/m2)	22 (19.4-25.8)	4306
CRP (mg/L)	6.4 (2.5-38.6)	4093
Hb (g/dL)	12.5 (11-13.9)	2453

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

‡Participants with data available for variable

Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein, Hb = haemoglobin

CPM Development

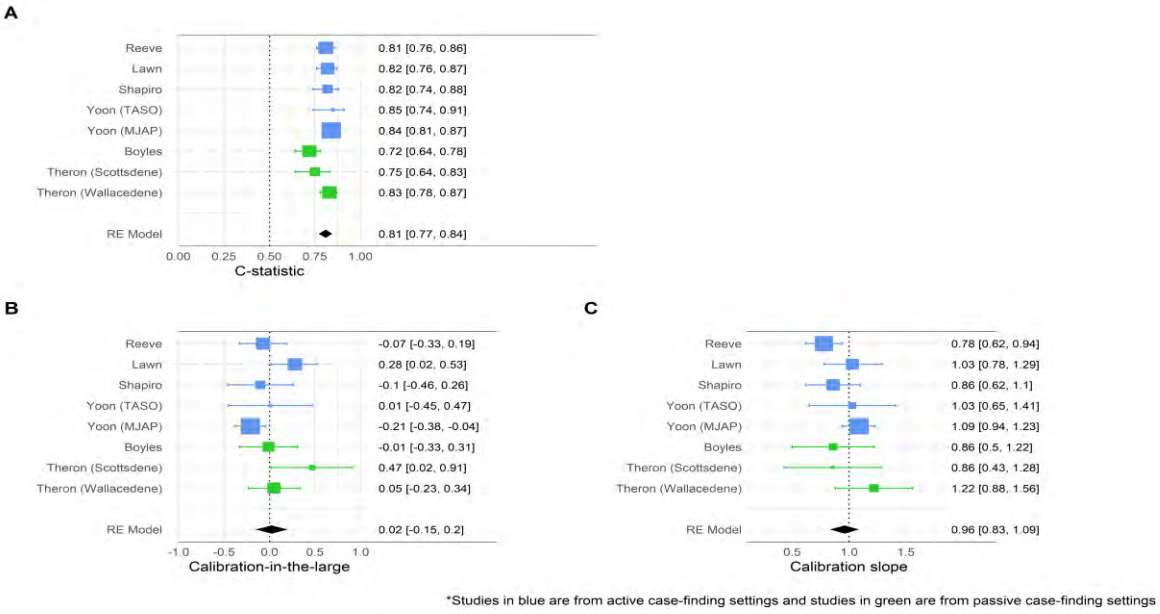
After backward stepwise selection on the full dataset, 5 of 10 candidate predictors were selected in ≥70% of bootstrap samples in ≥5 of 10 multiply imputed datasets (appendix Table 8-37). The predictors selected were age (60% of multiply imputed datasets), BMI (100%), CD4 count (100%), CRP (100%), and cough (50%). The appendix (Tables 8-37 and 8-38) shows all coefficients and knot locations of the CPMs. CRP had the strongest association with the outcome from all predictors (appendix Figure 8-14). The appendix (Figure 8-15) shows the associations (excluding cubic spline transformations) between each retained predictor and the outcome after fitting a linear mixed-effects model across datasets.

CPM validation

For the extended CPM, Figure 5-2 shows forest plots for performance statistics calculated during IECV. Discrimination was excellent. Pooled c-statistic was 0.81 (0.77, 0.84). C-statistics were consistent within active case-finding settings but heterogenous in passive-case finding settings. Calibration was adequate. Pooled calibration-in-the-large was 0.02 (-0.15, 0.20). There was slight underestimation of risk in the Lawn and Scottsdene cohorts. Pooled

calibration slope was 0.96 (0.83, 1.09). The calibration plots suggest reasonable agreement between predicted and observed risk for most cohorts (appendix Figure 8-16) but suboptimal calibration in the Reeve cohort (appendix Figure 8-16A). The joint probability of achieving a C-statistic > 0.80 and a calibration slope between 0.8 and 1.2 was 54% (appendix Table 8-40).

Figure 5-2: Forest plots showing extended CPM discrimination (A), calibration-in-the-large (B), and calibration slope (C)



*Studies in blue are from active case-finding settings and studies in green are from passive case-finding settings

For the CRP-only CPM, Figure 5-3 shows forest plots for performance statistics calculated during IECV. Discrimination was similar to that of the extended CPM with a pooled C-statistic of 0.79 (0.74, 0.83); calibration was also adequate. Pooled calibration-in-the-large was 0.02 (-0.15, 0.20). There was slight underestimation of risk in the Scottsdene cohort and slight overestimation in the Reeve cohort. Pooled calibration slope was 0.98 (0.83, 1.09). The appendix (Figure 8-17) shows that calibration plots were similar to those of the extended CPM with suboptimal calibration in the Reeve and Scottsdene cohorts (appendix Figures 8-17A and 8-17G). The joint probability of achieving a C-statistic > 0.80 and a calibration slope between 0.8 and 1.2 was 38% (appendix Table 8-39).

Figure 5-3: Forest plots showing C-reactive protein only CPM discrimination (A), calibration-in-the-large (B), and calibration slope (C)

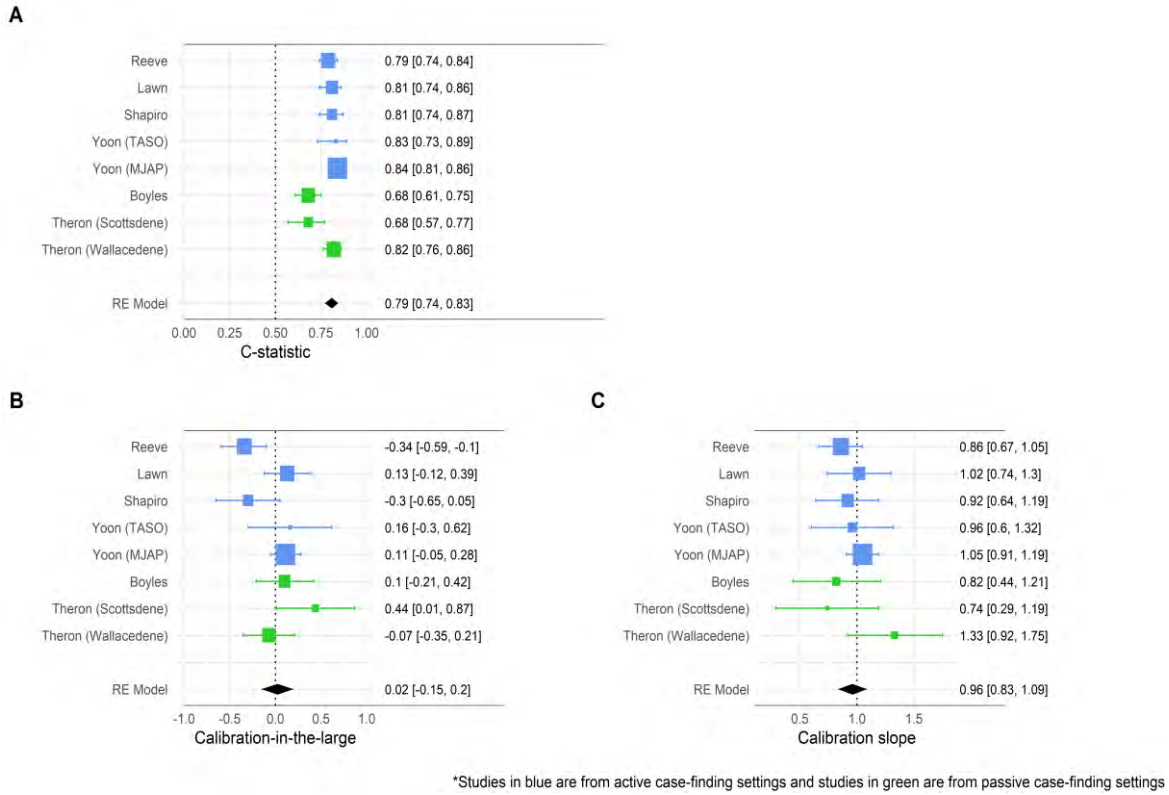


Table 5-2 compares performance statistics for both CPMs with WHO-recommended screening tools and another published CPM. Both CPMs had higher discrimination.

Table 5-2: Performance statistics of extended and C-reactive protein only CPMs, WHO-recommended tools and other CPMs*

Model/Test	Concordance statistic	Calibration-in-the-large	Calibration slope
Extended CPM	0.81 (0.77, 0.84)	0.02 (-0.15, 0.20)	0.96 (0.83, 1.09)
CRP only CPM	0.79 (0.74, 0.83)	0.01 (-0.20, 0.22)	0.97 (0.86, 1.08)
W4SS	0.57 (0.51, 0.63)		
CRP (5mg/dL)	0.70 (0.63, 0.75)		
Hanifa CPM	0.71 (0.68, 0.75)	0.41 (-0.21, 1.03)	1.05 (0.79, 1.32)

*The Extended CPM contained CRP, age, body mass index, CD4 cell count, and cough as predictors along with spline transformations for continuous variables. The CRP-only CPM only included CRP as a predictor along with spline transformations

*W4SS performance statistics calculated only for ACF datasets as all participants in PCF datasets were W4SS positive

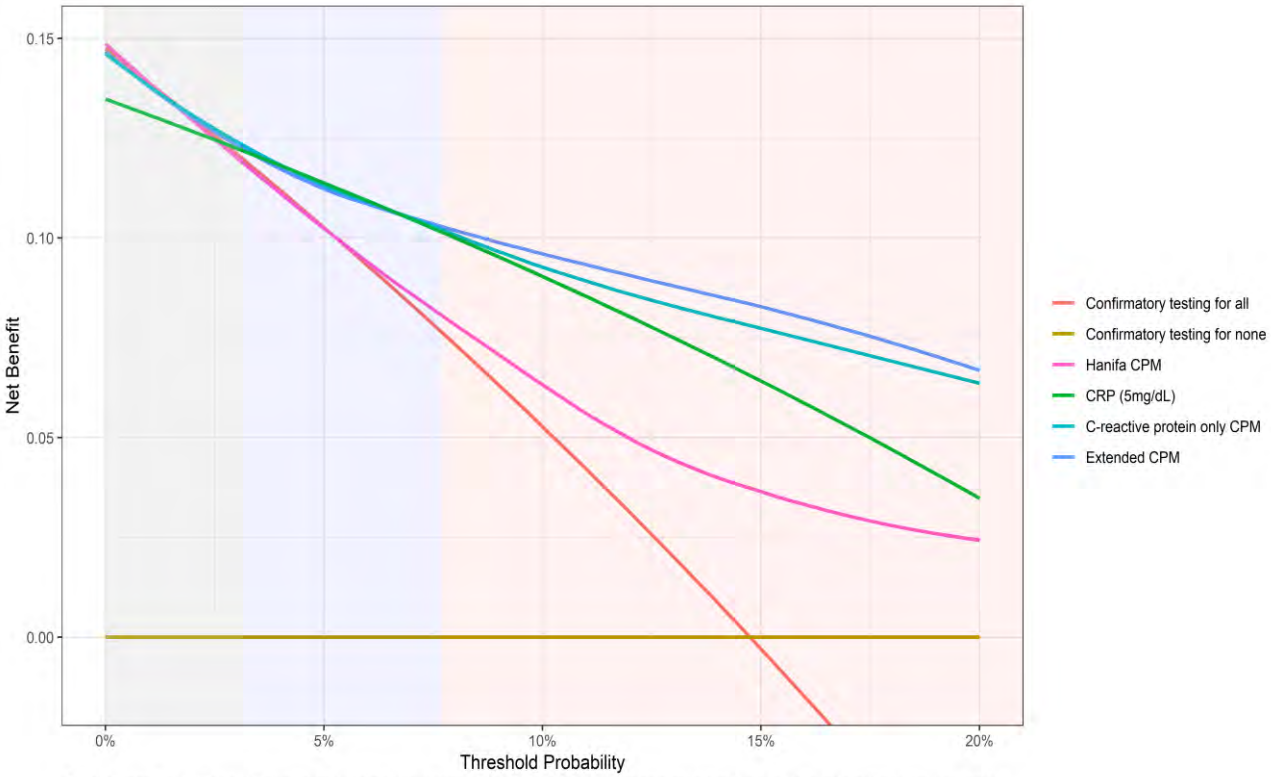
†CRP (5mg/dL cutoff) performance statistics exclude study by Boyles et al as all participants in that study had CRP level >=5mg/L

Definition of abbreviations: CRP = C-reactive protein

Assessment of clinical utility

Figure 5-4 and the appendix (Figure 8-18) show decision-curve analyses from the pooled IECV validation sets. Net benefit of both CPMs was equivalent or higher than that of other strategies across the range of threshold probabilities. Between threshold probabilities of 0% and 3.1%, net benefit of a "confirmatory testing for all" strategy was at least equivalent to that of both CPMs. Between threshold probabilities of 3.1% and 7.7%, net benefit of CRP (5 mg/L cut-off) was at least equivalent to that of both CPMs. At a threshold probability >7.7%, net benefit of both CPMs was higher than that of other strategies. The net benefit of 1 published CPM[113] was generally lower than that of both CPMs except at very low threshold probabilities (<3.1%). Results were similar when excluding passive-case finding cohorts (appendix Figure 8-18). Net benefit of the W4SS was lower than that of both CPMs and CRP (5 mg/L cut-off) across all threshold probabilities.

Figure 5-4: Decision curve analysis comparing the extended and C-reactive protein only CPMs to other tools or strategies other than W4SS among active and passive case-finding cohorts*†



*Grey shaded region: If you're willing to test 32 or more patients per tuberculosis case diagnosed, confirmatory testing for all has at least equivalent net benefit than that of the extended and C-reactive protein only models;
Blue shaded region: If you're willing to test between 32 and 13 patients per tuberculosis case diagnosed, CRP (5mg cut-off) has at least equivalent net benefit than the extended model and C-reactive protein only models;
Red shaded region: If you're willing to test 13 or fewer patients per tuberculosis case diagnosed, the final model has more net benefit than the C-reactive protein only model
†Excludes study by Boyles et al, as all participants in that study had CRP level >=5mg/L
Definition of abbreviations: CRP = C-reactive protein

We applied CRP (5 mg/L cut-off) to the stacked multiply imputed datasets (Table 5-3). CRP (5 mg/L) would have captured 91% of tuberculosis cases and resulted in confirmatory testing for 54% of participants. In comparison, to capture a similar percentage of cases, both CPMs would have resulted in confirmatory testing for a similar percentage of participants. To capture 95% of those with tuberculosis, the extended and CRP-only CPMs would have resulted in confirmatory testing for 74% and 75% of participants, respectively. Excluding passive-case finding cohorts, since all participants in those cohorts were W4SS positive, we applied the W4SS to the stacked multiply imputed datasets (Table 5-3). The W4SS would have captured 91% of tuberculosis cases and resulted in confirmatory testing for 78% of participants. CRP (5 mg/L cutoff), the extended CPM (4.2% threshold), and the CRP-only CPM (3.6% threshold) would have captured a similar percentage of cases but required confirmatory tests for only 50%, 59%, and 57% of participants, respectively.

Table 5-3: Trade-off between percentage of tuberculosis cases captured and percentage of participants needing confirmatory testing for extended CPM, C-reactive protein only CPM, and WHO-recommended tools using the 10 stacked multiply imputed datasets*

CPM or tool	CPM-based tuberculosis threshold	Percentage of tuberculosis cases captured	Percentage of all needing confirmatory testing	Number of confirmatory tests to capture one tuberculosis case
Active and passive case-finding cohorts (n=41,080)*†				
CRP 5 mg/L		91	54	4
Extended CPM†	4.9%	91	56	4.1
C-reactive protein only CPM†	4.2%	91	54	4
Extended CPM††	2.9%	95	74	5.3
C-reactive protein only CPM††	3.0%	95	75	5.3
Active case-finding cohorts (n=36,670)§				
W4SS		91	78	6.5
CRP 5 mg/L		89	50	4.3
Extended CPM#	4.2%	91	59	4.9
C-reactive protein only CPM#	3.6%	91	57	4.8
Extended CPM††	2.7%	95	76	6.1
C-reactive protein only CPM††	2.8%	95	80	6.4

*Excludes study by Boyles et al, as all participants in that study had CRP level >=5mg/L

†For both CPMs, thresholds were selected to capture a similar percentage of tuberculosis cases compared with CRP at 5 mg/L cut-off (91%)

††For both CPMs, thresholds were selected to capture a similar percentage of tuberculosis cases compared with an ideal triage test according to WHO target product profile (95%)

#For both CPMs, thresholds were selected to capture a similar percentage of tuberculosis cases compared with W4SS

§Excludes 3 passive case-finding cohorts, as all participants in that study had a positive W4SS

Definition of abbreviations: CRP = C-reactive protein

Sensitivity analyses

We repeated analyses for the extended CPM using an alternative imputation method (appendix Figure 8-19 and Table 8-41). The results were similar to the main analyses.

5.5 Discussion

We investigated the utility of WHO-recommended screening tools (W4SS and CRP [5 mg/L cut-off]) and developed and validated 2 CPMs for tuberculosis screening in outpatient PLHIV

using 8 cohorts (4,315 participants). At validation, the extended CPM showed excellent discrimination and adequate calibration; the CRP-only CPM showed similar performance. The W4SS and CRP (5 mg/L cut-off) had lower discrimination. Both CPMs had equivalent or higher net benefit across all threshold probabilities compared with other tools or strategies. However, CRP (5 mg/L cut-off) demonstrated similar net benefit to both CPMs over a clinically plausible range of threshold probabilities; CRP (5 mg/L cut-off) also met WHO minimum sensitivity requirements (90% sensitivity). At lower threshold probabilities, or if WHO optimal sensitivity requirements (95% sensitivity) are preferred, both CPMs and a "confirmatory testing for all" strategy had similar net benefit. The W4SS had suboptimal net benefit. CRP (5 mg/L cut-off) had similar sensitivity to the W4SS but required 36% fewer confirmatory tests.

By assessing clinical utility, we provide further evidence of CRP's value for tuberculosis screening in outpatient PLHIV. In a recent meta-analysis, CRP (5mg/L cutoff) showed similar sensitivity but higher specificity compared with W4SS,[188] leading to its inclusion in updated WHO tuberculosis screening guidelines.[185] It is recommended that emerging biomarkers be evaluated against available tools.[215] Our findings suggest that CRP and the newly developed CPMs be used as a benchmark to evaluate emerging biomarkers for tuberculosis screening; CRP may also be combined with other biomarkers to improve predictive performance. The addition of clinical characteristics (i.e., W4SS symptoms) to CRP provided minimal extra information, since both CPMs showed similar performance. The W4SS is a key component of tuberculosis screening guidelines but has suboptimal utility. Variable selection further demonstrated the limited role of symptoms in predicting tuberculosis as only 1 of the 4 W4SS symptoms was retained during backward selection.

Although CRP (5mg/L cutoff) and both CPMs had high net benefit across a wide range of thresholds, a 'confirmatory testing for all' strategy may be considered if a setting has resources to perform many confirmatory tests per case diagnosed. Given the high prior-probability of tuberculosis in this study (tuberculosis prevalence between 25-38% in passive-case finding cohorts and 10-17% in active case-finding cohorts not yet on ART), a 'confirmatory testing for all' strategy may be plausible.

We externally validated a published CPM by Hanifa et al,[113] which showed suboptimal utility and performance compared with our CPMs. Hanifa et al included similar predictors but did not include CRP or account for non-linear associations. Auld et al recently developed a CPM

for tuberculosis in outpatient PLHIV and validated the CPM in 3 cohorts.[111] The CPM included W4SS symptoms, sex, smoking status, temperature, BMI, and hemoglobin as predictors. However, at a cut-off that provides similar sensitivity to W4SS, the score did not improve specificity. The score was also externally validated using a cohort included in this article, showing much lower discrimination than the extended CPM (C-statistic of 0.63 vs 0.82 for the extended CPM).[86] Baik et al recently developed a CPM for tuberculosis in symptomatic outpatients irrespective of HIV status. However, performance was not assessed in PLHIV.[112] Balcha et al developed a relatively complex CPM for tuberculosis among outpatients with a positive W4SS. However, the CPM has not been validated internally or externally.[114] Similarly, the TBscore has been developed but is complex, consisting of 11 symptoms and signs, and has low specificity (36%).[204]

Our study has several strengths. This study is the only one to validate a CPM and other tools for tuberculosis using the recommended IECV framework.[139] We included a large population of outpatient PLHIV from 8 different settings to evaluate generalizability. We also included outpatient PLHIV irrespective of case-finding status to improve generalizability. We used various measures of clinical utility, including net benefit and the trade-off between number of tuberculosis cases captured and unnecessary additional confirmatory testing. For CPM development, we used multiple imputation to handle missing data, selected readily available predictors, avoided categorization of continuous variables, and accounted for non-linear relationships. Finally, we adhered to the TRIPOD statement and additional guidelines.[139,206,207]

Our study has several limitations. First, active case-finding study populations did not include PLHIV on ART and passive case-finding cohorts only comprised 15% of all data. Therefore, results should be extrapolated with caution to these subpopulations. However, PLHIV not on ART - who comprised 91% of participants - currently still represent a third of all PLHIV (~13 million people)[216] and have a high tuberculosis prevalence (~10-15%).[188] Second, all cohorts were drawn from high-burden outpatient settings in South Africa and Uganda, meaning results may not generalize to low-burden settings. Third, we did not include certain well-known predictors of tuberculosis such as hemoglobin,[121] because of missing data. We were also unable to evaluate chest X-ray – another WHO-recommended screening tool – since only 1 study performed chest X-ray. However, chest X-ray has suboptimal diagnostic performance as a screening tool and is only recommended in combination with W4SS.[185] Besides, it is often unavailable in outpatient settings.[54] Fourth, although our results are largely

applicable to pulmonary tuberculosis, extrapulmonary tuberculosis is less likely in outpatient settings. Fifth, we were unable to evaluate several published CPMs with predictors that were not measured in some or all cohorts.[111,112,114,116,145,204] Finally, we did not investigate the cost and resource implications of CRP-based strategies.

In conclusion, our findings define optimal tuberculosis screening strategies in outpatient PLHIV based on currently available data, accounting for the trade-off between the number of tuberculosis cases diagnosed and number of confirmatory tests performed. CRP (5mg/L cutoff) - which has been recently recommended by WHO – showed optimal net benefit across a plausible range of thresholds. Two newly developed CPMs that incorporate CRP as a predictor may add value at more extreme threshold probabilities – where resources allow more or fewer confirmatory tests per diagnosed case. A 'confirmatory testing for all' strategy might also be considered if resources permit. Conversely, the WHO-recommended W4SS showed suboptimal utility. CRP (either alone or as part of a CPM) sets the standard for tuberculosis screening in outpatient PLHIV, and the newly developed CPMs may also be used as a benchmark to evaluate future biomarkers or combined with other biomarkers to improve predictive performance.

5.6 Contributors

AD, RKG, APK, GaM, and DAB designed the study and protocol and interpreted the results. DAB supervised the study. AD and YH did the initial systematic review. RKG, ADK, CY, AC, BWPR, GT, GN, RB, PKD, CJC, MN, TB, and GrM contributed data to the meta-analysis. AD analyzed the data with assistance from RKG, APK, and DAB. AD, RKG, and DAB wrote the first draft of the manuscript, which was revised based on comments from co-authors. AD, RKG, and DAB accessed and verified the data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. All authors approved the final version of the manuscript.

CHAPTER 6: DISCUSSION OF THESIS

6.1 Summary of findings

In this thesis, I evaluated different strategies to improve screening and diagnosis of tuberculosis in PLHIV in resource-limited settings. This section summarizes the key findings of the research papers in chapters 2 to 5, which relate to the four separate objectives.

6.1.1 Objective 1 – To determine the diagnostic accuracy of the W4SS and alternative screening tools and strategies in ambulatory PLHIV, including key subgroups

In chapter 2, I conducted a systematic review and IPDMA using data from 22 studies and 15,666 ambulatory PLHIV to inform an update to WHO tuberculosis screening guidelines in ambulatory PLHIV, including key subgroups.

Among outpatients not on ART, I showed that W4SS sensitivity was 85% (76, 91) but specificity was only 37% (25, 51; n=11,160). CRP (≥10 mg/L) had similar sensitivity (83% [79, 86]) to W4SS, but higher specificity (67% [60, 73]; n=3,187), and a sequential strategy (second screening test offered only if first screening test is positive) of W4SS then CRP (≥5 mg/L) also had a similar sensitivity (84% [75, 90]) to W4SS, but higher specificity (64% [57, 71]; n=3,187) than W4SS; at 10% tuberculosis prevalence, these strategies would require 272 and 244 fewer rapid Xpert confirmatory tests per 1,000 PLHIV than W4SS but miss two and one more tuberculosis cases, respectively.

Among outpatients on ART, W4SS sensitivity was only 53% (35, 71) and specificity was 71% (51, 85; n= 4,309). CRP data was limited, but a parallel strategy (two screening tests offered at the same time) of W4SS with any chest X-ray abnormality had higher sensitivity (89% [70, 97]) than W4SS, but lower specificity (33% [17, 54]; n=2,670); at a tuberculosis prevalence of 5%, this strategy would require 379 more Xpert confirmatory tests per 1,000 PLHIV than W4SS but detect 18 more tuberculosis cases.

Regardless of ART status, chest X-ray had lower sensitivity than W4SS in studies that directly compared both tests, making it unsuitable as a standalone screening test. Cough (≥2 weeks), haemoglobin (<10 g/dL), BMI (<18.5 kg/m^2), and lymphadenopathy had high specificities (80–90%) but low sensitivities (29–43%).

The WHO Xpert algorithm (W4SS followed by Xpert confirmatory testing if W4SS is positive) had a sensitivity of only 58% (50, 66); Xpert confirmatory testing for all (i.e., no screening test) had a slightly higher sensitivity of 68% (57, 76). One study among outpatients not on ART assessed both Xpert and Xpert Ultra; the sensitivity of sputum Xpert Ultra was higher than sputum Xpert (73% [62, 81] vs 57% [47, 67]) and specificities were similar (98% [96, 98] vs 99% [98, 100]).

6.1.2 Objective 2 – To determine the performance of the W4SS and alternative screening tools and strategies in HIV-positive inpatients

In chapter 3, I conducted a systematic review and IPDMA using data from 6 studies and 3,660 HIV-positive inpatients admitted to hospital irrespective of tuberculosis symptoms and signs to inform an update to WHO tuberculosis screening guidelines in this population.

I showed that the pooled proportion of inpatients eligible for an Xpert confirmatory testing using the WHO-recommended W4SS was high (90% [95% CI 89, 91; n=3,658]). Among screening tools to guide Xpert confirmatory testing, W4SS and CRP (≥5 mg/L) had the highest sensitivities (≥96%) but very low specificities (≤12%); cough (≥2 weeks), haemoglobin concentration (<8 g/dL), BMI (<18.5 kg/m^2), and lymphadenopathy had higher specificities (61–90%) but suboptimal sensitivities (12–57%).

The WHO Xpert algorithm (W4SS followed by Xpert confirmatory testing if W4SS is positive) had a sensitivity of 76% (95% CI 67, 84) and specificity of 93% (88, 96; n=637). Xpert for all had similar accuracy to the WHO Xpert algorithm: sensitivity was 78% (95% CI 69, 85) and specificity was 93% (87, 96; n=639). Finally, in two cohorts that had sputum and non-sputum samples collected for culture or Xpert, diagnostic yield of routine sputum Xpert was only 41–70% (mostly because a high proportion of inpatients were unable to produce sputum) and 61–64% for routine urine Xpert.

6.1.3 Objective 3 – To determine the performance of WHO screening criteria and alternative screening tools and strategies to guide LF-LAM testing in HIV-positive inpatients

In chapter 4, I conducted a systematic review and IPDMA using data from 5 studies and 3,504 HIV-positive inpatients admitted to hospital irrespective of tuberculosis symptoms and signs to 1) assess the performance of WHO screening criteria and alternative screening tools/strategies to guide LF-LAM testing and 2) compare diagnostic accuracy of the WHO

AlereLAM algorithm (WHO screening criteria followed by AlereLAM if screen positive) with AlereLAM and FujiLAM (a novel LF-LAM test) testing in all HIV-positive inpatients.

I show that the pooled proportion of inpatients eligible for AlereLAM using WHO screening criteria is high (93% [95%CI 91, 95]). Among screening tools and strategies to guide LF-LAM testing, WHO criteria, CRP (≥5 mg/L), and CD4 cell count (<200 cells/ μL) had high sensitivities but low specificities; cough (≥2 weeks), haemoglobin (<8 g/dL), BMI (<18.5 kg/m^2), lymphadenopathy, and WHO-defined danger signs had higher specificities but suboptimal sensitivities.

AlereLAM in all HIV-positive inpatients had the same sensitivity (62%) and specificity (88%) as that of the WHO AlereLAM algorithm. In 2 studies, sensitivity of FujiLAM was 21 percentage points higher than AlereLAM and specificity was 8 percentage points lower, although confidence intervals overlapped. In 2 studies that collected sputum and non-sputum samples for Xpert and/or culture, diagnostic yield of sputum Xpert was 40–41%, while diagnostic yield of AlereLAM was similar or higher (39–76%), since urine samples were obtained from almost all inpatients, but many were unable to produce sputum. In one study, FujiLAM diagnosed twice as many tuberculosis cases compared with AlereLAM (80% vs 39%), and sputum Xpert combined with AlereLAM, urine Xpert, or FujiLAM diagnosed 61%, 81%, and 92% of all cases, respectively.

6.1.4 Objective 4 – To develop and validate novel CPMs for pulmonary tuberculosis screening in outpatient PLHIV and to determine the clinical utility of these CPMs and WHO-recommended screening tools

In chapter 5, I conducted an IPDMA using data from 8 cohorts and 4,315 outpatient PLHIV (the majority of whom were not on ART) who were enrolled either regardless of signs and symptoms of tuberculosis (ACF) or with a positive W4SS (PCF). I developed and validated 2 novel CPMs in outpatient PLHIV for pulmonary tuberculosis and determined the utility of these CPMs and WHO-recommended screening tools.

I show at validation that the extended CPM (which contained CRP, age, BMI, CD4 cell count, and cough) had excellent discrimination (C-statistic 0.81 [0.76, 0.86]) and the CRP-only CPM (which only included CRP as a predictor along with spline transformations) had similar discrimination (C-statistic 0.79 [0.74, 0.83]). Both CPMs had higher discrimination than WHO-recommended screening tools: CRP (≥5mg/L [C-statistic 0.70 [0.63, 0.75]) and W4SS (C-statistic 0.57 [0.51, 0.63]).

Using decision curve analysis to assess clinical utility, both CPMs showed equivalent or higher net benefit across all threshold probabilities compared with other tools or strategies (including WHO-recommended screening tools). Compared with both CPMs, CRP (≥5mg/L) showed optimal net benefit across a plausible range of thresholds (~13 to 32 confirmatory tests performed to identify one tuberculosis case). The newly developed CRP-based CPMs added value at more extreme thresholds – if resources permit more confirmatory tests per diagnosed case or if resources only allow fewer confirmatory tests per diagnosed case. The W4SS demonstrated lower net benefit compared with other screening tools and both CPMs. The W4SS would capture 91% of tuberculosis cases and result in confirmatory testing for 78% of participants; CRP (≥5 mg/L) would capture a similar number of participants compared with W4SS but reduce confirmatory tests required by 36%.

6.2 Limitations

This section summarizes overall limitations of these analyses. The limitations of each research paper are discussed in chapters 2 to 5.

First, some analyses in certain subgroups (e.g., outpatients on ART, pregnant PLHIV, and inpatients) were based on limited data, leading to inadequate precision and wide 95% CIs. Specifically, data on CRP in PLHIV on ART were scarce, impacting conclusions for chapters 2 and 5. This limitation highlights the need for additional diagnostic accuracy studies in these subgroups irrespective of tuberculosis symptoms and signs. Further diagnostic test accuracy studies are especially needed for PLHIV on ART given the limited data on this group and recent rapid increase in ART coverage. The generalizability of these findings will be further limited as ART coverage increases over the coming years. Another potential limitation is that I did not evaluate diagnostic accuracy by other factors (e.g., age and sex). However, diagnostic accuracy likely varies less for these factors than the factors evaluated (i.e., ART status and CD4 cell count).

Second, most studies included in the IPDMAs were from sub-Saharan Africa, and there were few studies from countries other than South Africa. Almost half of all included studies were done in South Africa, which might also be considered more urbanised than other low-income and middle-income countries. Thus, the generalisability of the findings within these IPDMAs to other geographical regions and to settings with a low tuberculosis prevalence is unclear. However, the WHO African region accounted for 74% of all HIV-associated tuberculosis cases in 2020 and has 23 countries on WHO's global list of 30 countries with a high burden

of HIV-associated tuberculosis.[1] Study findings might not also be generalisable to children with HIV, and test performance might vary in the context of regular screening.

Third, for the studies that contributed to chapters 2, 3, and 4, I only included participants in analyses if they had complete data on both the index test in question and the reference standard. For example, participants who were unable to produce a sputum sample were largely excluded from several analyses because a sputum sample is required for sputum culture, meaning that findings might not generalise to this to those who are unable to produce sputum. In chapter 5, we included participants unable to produce a sputum sample, since multiple imputation was used to deal with missing data prior to prediction modelling.

Fourth, the reference standard might be considered imperfect. Few studies included extra-pulmonary tuberculosis samples for culture or Xpert, meaning our results are more applicable to pulmonary tuberculosis. Since IPDMAs in chapters 3 and 4 were based on inpatient cohorts, an imperfect reference standard might affect the results of these chapters to a greater extent. Inpatients often present with extrapulmonary or disseminated tuberculosis and produce paucibacillary sputum samples.[23] The reference standard in all of the IPDMAs might also be considered imperfect because sputum culture, which was all that was done in most of the included studies, should ideally comprise multiple samples collected in the early morning to maximise sensitivity, but this was not done in any of our included studies. The imperfect reference standard may result in an underestimation of specificity and overestimation of sensitivity of existing algorithms. Tuberculosis prevalence estimates are also likely to be underestimates because of the limitations of our reference standard. A composite reference standard (that includes clinical assessment) would likely have resulted in lower sensitivity and higher specificity. However, a composite reference standard has disadvantages since the individual components are assumed to have the same accuracy and to be independent of one another.[217,218] Other methods, such as latent class analysis, may be useful in the absence of gold standard test.

It is unlikely, however, that this limitation would alter the findings in this thesis. For the IPDMAs described in chapters 2 and 5, most tuberculosis cases in an outpatient screening setting are likely pulmonary tuberculosis cases. Our results were also consistent across several reference standards: culture, combinations of culture and Xpert, and Xpert (which is the currently recommended confirmatory test). Diagnostic yield analyses for Xpert and LF-LAM confirmatory tests in chapters 3 and 4 also did not require a reference standard. For the

IPDMAs in chapters 2, 3, and 4, the alternative reference standards to assess screening tools were the WHO-recommended confirmatory tests Xpert and LF-LAM, which correctly classifies Xpert or LF-LAM positive tuberculosis, respectively. WHO recommends that the diagnostic accuracy of screening tools be assessed against recommended confirmatory tests that follow and not just culture (which is the gold standard).[79] Furthermore, estimates of the proportion of inpatients eligible for Xpert and AlereLAM according to WHO criteria in chapters 2, 3, and 4 were based on data with higher methodological quality, since these analyses did not require a reference standard.

Fifth, although the IPDMAs in chapters 2, 3, and 4 report diagnostic test accuracy using direct comparisons, which minimizes confounding by applying both tests to each individual, these analyses were limited by fewer studies and reduced precision. Furthermore, the limited number of studies included in the IPDMAs for chapters 3, 4, and 5 precluded adequate investigation of heterogeneity, as well as publication bias.

Sixth, I was unable to obtain IPD for 3 out of the 25 studies that contributed to chapter 2, although these 3 studies comprised only 8% of potentially available data. For the IPDMAs described in chapters 3, 4, and 5, all studies identified in the systematic review were obtained and included in analyses.

Seventh, data on some confirmatory tests were limited or not sought. Only 2 studies evaluated FujiLAM. Although we sought data for Xpert Ultra, only 1 study in chapter 1 assessed Xpert Ultra,[173] and no studies in chapters 2 and 3 assessed Xpert Ultra. We also did not attempt to obtain data on other molecular-based tests, such as TB-LAMP and Trunat assays. However, TB-LAMP has suboptimal sensitivity in PLHIV and is not recommended by WHO.[61] Furthermore, WHO recently assessed the diagnostic accuracy of Trunat assays, finding that no study has assessed this assay in PLHIV irrespective of tuberculosis symptoms and signs.[58]

Eighth, data on some screening tools were limited. We used W4SS or CD4 cell count ≤200 cells/μL as WHO eligibility criteria for AlereLAM in chapter 4, given limited data on WHO-defined danger signs and WHO stage. However, if WHO-defined danger signs and WHO stage were included in the definition of WHO eligibility criteria for AlereLAM, the proportion eligible for AlereLAM would be even higher. Thus, this limitation would not alter the conclusions of this chapter. For the IPDMA in chapter 5, several potential predictors of tuberculosis were not included during CPM development. These predictors were missing for

a large proportion of participants or unmeasured in several cohorts. For example, data on haemoglobin, a well-known predictor of tuberculosis,[121] was missing in 43% of individuals overall and systematically missing (i.e., 100% missing) in 2 cohorts. I was also unable to validate several published CPMs in the literature with predictors that were not measured in some or all cohorts.[111,112,114,116,145,204]

Finally, for the IPDMAs in chapters 2 and 3, calculations based on a hypothetical cohort were presented to give insight into consequences of screening and confirmatory testing, but these calculations were often based on heterogenous diagnostic test accuracy measures. Furthermore, in the case of inpatients, these calculations were based on diagnostic accuracy results derived from few participants, some of whom had an imperfect reference standard done. Therefore, these results should be treated with caution given the uncertainty of the estimates that these results were based on.

6.3 Implications of findings

This section discusses the implications of the findings for global tuberculosis control strategies based on the research papers in chapters 2 to 5. The findings in chapters 2 and 3 informed the updated 2021 WHO guidelines on tuberculosis screening in PLHIV, leading to 5 new or updated WHO recommendations (Table 6-1).[185]

Table 6-1: New or updated WHO recommendations for tuberculosis screening among PLHIV[185]

1	Among adults and adolescents living with HIV, systematic screening for tuberculosis disease should be conducted using the WHO-recommended four symptom screen and those who report any one of the symptoms of current cough, fever, weight loss or night sweats may have tuberculosis and should be evaluated for tuberculosis and other diseases (existing recommendation: strong recommendation, moderate certainty of evidence).
2	Among adults and adolescents living with HIV, C-reactive protein with a cut-off of > 5 mg/L may be used to screen for tuberculosis disease. (new recommendation: conditional recommendation, low certainty of evidence for test accuracy).
3	Among adults and adolescents living with HIV, chest X-ray may be used to screen for tuberculosis disease. (new recommendation: conditional recommendation, moderate certainty of evidence for test accuracy).
4	Among adults and adolescents living with HIV, molecular WHO-recommended rapid

	diagnostic tests may be used to screen for tuberculosis disease (new recommendation: conditional recommendation, moderate certainty of evidence for test accuracy).
5	Adult and adolescent inpatients with HIV in medical wards where the tuberculosis prevalence is > 10% should be tested systematically for tuberculosis disease with a molecular WHO-recommended rapid diagnostic test (new recommendation: strong recommendation, moderate certainty of evidence for test accuracy).

6.3.1 Screening for tuberculosis in outpatient PLHIV not on ART

My findings in chapters 2 and 5 suggest that CRP has good diagnostic accuracy and utility in outpatient PLHIV not on ART. CRP (≥10 mg/L) approached WHO minimum thresholds for a screening tool in this subgroup (with 83% sensitivity and 67% specificity vs WHO's thresholds of 90% sensitivity and 70% specificity). Although CRP (≥5 mg/L) has lower specificity (53%), WHO recommended CRP at this cut-off because of higher sensitivity (89%; Table 6-1).[185]

The major advantage of CRP compared with W4SS is that its higher specificity translates into fewer subsequent Xpert tests required. At 10% tuberculosis prevalence, CRP (≥10 mg/L) or a strategy of W4SS then CRP (≥5 mg/L) would reduce the number of Xpert tests required compared with W4SS by 42% and 37%, respectively, while capturing a similar number of tuberculosis cases. Efforts to scale-up Xpert have been slow, especially in decentralised locations in high HIV-tuberculosis burden countries.[54,178] Xpert is also able to provide a result in 2 hours, but in the real world results take several days.[51] By reducing the number of Xpert tests required, CRP may not only allow for broader implementation of Xpert but also reduce the time to a result. Since CRP is a better rule out test than W4SS, CRP would also reduce the time to start IPT in PLHIV.

CRP has the potential for affordable scale-up. Several POC assays are available, ranging from qualitative lateral-flow assays that do not require a power source or refrigeration to quantitative assays that require a small machine.[179] CRP POC assays can cost US$2 per test, provide results in less than 3 min, and be performed with minimal expertise (blood collected by finger prick). In one cohort included in the IPD meta-analysis, the authors evaluated costs of W4SS and CRP strategies in 1,245 outpatient PLHIV.[89] The W4SS followed by Xpert if screen positive cost $12,000, while CRP (≥8 mg/L) followed by Xpert if screen positive cost

$7,968 (a 34% reduction in costs due to a reduction in the number of Xpert tests needed). Both strategies captured a similar number of tuberculosis cases. Given the low sensitivity of the current WHO Xpert algorithm, the reduction in costs could allow for confirmatory testing with both Xpert and culture.[89] Possible barriers to the implementation of POC CRP include negative effects on clinic workflows, human resources, and workload.[219]

The findings in chapter 5 highlight the clinical utility of CRP not only at the new WHO-recommended 5mg cut-off but also when incorporated into CPMs. Both newly developed CPMs outperformed an existing CPM by Hanifa et al and, in 1 cohort, another CPM by Auld et al.[111,113] The CPMs have clinical utility if resources would allow more or fewer confirmatory tests per diagnosed case than that of CRP (≥5 mg/L). The W4SS was found to have suboptimal utility. The addition of clinical characteristics (i.e., W4SS symptoms) to CRP provides minimal extra information, since both CPMs showed similar performance. Variable selection further demonstrated the limited role of symptoms in predicting tuberculosis as only 1 of the 4 W4SS symptoms was retained during backward selection.

Of the other screening tools evaluated, haemoglobin, BMI, and lymphadenopathy had low sensitivities, making them unsuitable as screening tests. Their presence, however, should prompt a thorough search for tuberculosis, given their high specificities and known association with mortality.[181,182]

6.3.2 Screening for tuberculosis in outpatient PLHIV on ART

My findings in chapter 2 suggest that more data and/or new screening strategies are needed to determine the optimal screening approach in outpatients on ART.

The WHO still recommends the W4SS in outpatients on ART, although the W4SS misses approximately 50% of tuberculosis cases. A parallel strategy of W4SS combined with chest x-ray would detect 70% more tuberculosis cases than W4SS alone at 5% tuberculosis prevalence. Therefore, WHO now also recommends this strategy in outpatients on ART (Table 6-1). However, this strategy would require that 40% of outpatients on ART receive a chest X-ray and 80% receive an Xpert test. This strategy might not only pose a substantial cost burden, but also pose other challenges, such as a negative effect on infrastructure and human resources. Furthermore, in a 2016 survey of 14 countries with a high HIV-associated tuberculosis burden, only 14% of those countries had chest X-ray available as a screening tool in primary health centres.[54] However, newer advances (e.g., computer-aided detection software) may facilitate easier implementation of this strategy. A further advantage of chest

X-ray is that it provides additional diagnostic information for diseases other than tuberculosis.

Compared with the parallel strategy of W4SS combined with chest X-ray, Xpert for all (i.e., no screening tool) would likely have similar costs but would detect more tuberculosis cases. CRP holds promise in this population, but limited data precluded any conclusions on its use. In 1 study of 381 outpatients on ART, CRP (≥5mg/L) had higher sensitivity compared with W4SS (40% vs 8%).[91] Finally, of the other screening tools evaluated, haemoglobin, BMI, and lymphadenopathy had high specificities, meaning that those with a positive screen require confirmatory testing for tuberculosis.

6.3.3 Screening for tuberculosis in HIV-positive inpatients

My findings in chapters 2 and 3 suggest that screening tools have suboptimal accuracy in HIV-positive inpatients. Based on the findings, I argue that hospitals should implement confirmatory testing with Xpert and AlereLAM in all HIV-positive medical inpatients. WHO has now recommended Xpert testing in all HIV-positive medical inpatients (Table 6-1).

The specificities of W4SS and CRP were much lower than in outpatient PLHIV. As a result, an estimated 90% and 84% of inpatients had a positive W4SS and CRP (≥10mg/L), respectively, and would require further confirmatory testing with Xpert. The specificity of WHO screening criteria for AlereLAM confirmatory testing was also low. As a result, as an estimated 93% of inpatients had a positive screen and would require further confirmatory testing with AlereLAM. The low specificities of these screening tools are likely a result of the high prevalence of other opportunistic diseases and bacterial infections in inpatients without tuberculosis. Of all potential screening tools to guide AlereLAM testing, CD4 cell count (<200 cells/ μL) provided the best trade-off between sensitivity and specificity. However, CD4 cell count would miss 10% of AlereLAM positive tuberculosis cases.

The diagnostic yield findings in chapters 2 and 3 highlight the value of urine-based confirmatory testing in inpatients. If only universal sputum Xpert is done, an estimated 60% of tuberculosis cases would be missed. The low yield of sputum Xpert is because inpatients have difficulty producing sputum for testing. In 3 included cohorts,[18,20,189] an estimated 35–54% of participants could not produce sputum for Xpert testing. Therefore, AlereLAM or urine Xpert might have an important role in inpatients who are unable to produce sputum. It

is unclear whether urine Xpert or AlereLAM provides higher yield. Urine Xpert had higher yield compared with AlereLAM in one included cohort,[20] but the opposite was true in another included cohort.[18]

Xpert and AlereLAM in all HIV-positive inpatients has several advantages. First, since almost all HIV-positive inpatients met WHO eligibility requirements for both tests, universal testing might reduce diagnostic complexity and accelerate clinical decision making in busy inpatient settings. For example, CD4 cell count is part of the WHO criteria for AlereLAM testing but may not be immediately available to clinicians. Second, Xpert and AlereLAM were positive in 2% and 5% of inpatients who did not meet WHO criteria for testing, respectively. Third, in the real world, all HIV-positive inpatients who qualify for Xpert and AlereLAM testing might not ultimately receive a test. In two included studies, at least 90% of HIV-positive inpatients met WHO criteria for Xpert and AlereLAM, but clinicians identified only 38–64% as having possible tuberculosis after clinical assessment.[18,189] In 1 of the studies, 19% of inpatients without clinically suspected tuberculosis had a positive AlereLAM test.[189] Fourth, routine AlereLAM testing (in addition to routine sputum Xpert) was cost-effective in the STAMP trial.[18] Fifth, two randomised trials have demonstrated a reduction in all-cause mortality among HIV-positive inpatients with the use of urine-based diagnostics (AlereLAM and/or urine Xpert) in addition to routine diagnostics.[18] One trial included HIV-positive inpatients irrespective of tuberculosis signs and symptoms,[18,73] while the other included inpatients with a positive W4SS (which comprise > 90% of HIV-positive inpatients).[73] Despite these findings, a recent survey of 24 high tuberculosis/HIV burden countries revealed that only 4 (17%) were using AlereLAM in all hospitals. [203]

An alternative strategy to tuberculosis screening is empirical tuberculosis treatment. No trials have been conducted in HIV-positive inpatients. However, among outpatient PLHIV with severe immunosuppression, empirical tuberculosis treatment did not reduce mortality compared with treatment following extensive tuberculosis screening.[220] Other randomised controlled trials have also shown that empirical tuberculosis treatment did not reduce mortality compared with IPT/TPT or treatment according to tuberculosis guidelines.[221,222]

6.3.4 The accuracy of the WHO algorithm vs confirmatory testing for all

Based on findings in chapters 2 and 3, WHO now recommends that Xpert for all be considered as an alternative to the WHO Xpert algorithm (Table 6-1), though this strategy is only possible if a setting has resources to perform many confirmatory tests per case

diagnosed. An important implication of the findings in chapters 2, 3, 4, and 5 is that the entire WHO algorithm and confirmatory testing for all are unable to definitively rule out tuberculosis.

In all outpatients, the WHO Xpert algorithm and Xpert for all would miss 40% and 33% of tuberculosis cases, respectively. The low yield is because both the W4SS and Xpert have inadequate sensitivities. However, Xpert Ultra may improve yield. For example, in 1 included cohort of outpatients not on ART, Xpert Ultra improved sensitivity compared with Xpert by 16 percentage points (73% vs 57%).[173]

Similarly, in HIV-positive inpatients, both Xpert and AlereLAM were insufficiently sensitive to identify all tuberculosis cases. In 1 included cohort,[20] sputum Xpert combined with either urine Xpert or AlereLAM missed 19% and 39% of tuberculosis cases, respectively. In those with a negative result on both tests, physicians should consider additional diagnostic approaches that incorporate clinical symptoms and signs, radiological tests (e.g., chest x-ray and abdominal ultrasound), laboratory tests (e.g., haemoglobin concentration), and tuberculosis confirmatory tests on non-sputum samples (e.g., Xpert Ultra).[45,92,121,196,197]

Newer technologies might substantially close the diagnostic gap in inpatient populations. For example, sputum Xpert Ultra and FujiLAM have shown increased sensitivity compared with Xpert and AlereLAM, respectively.[43,75] In chapter 4, sputum Xpert when combined with FujiLAM diagnosed 92% of tuberculosis cases (versus 61% when combined with AlereLAM). In a recent Cochrane systematic review, sputum Xpert Ultra increased sensitivity over sputum Xpert in PLHIV by 13 percentage points (88% vs 75%).[43] Furthermore, a recent study showed that the sensitivity of urine Xpert Ultra was double that of AlereLAM (33% vs 16%).[48]

6.4 Future research

This section discussed suggestions for future research based on gaps identified in the research papers in chapters 2 to 5.

6.4.1 Further studies to assess screening tools and confirmatory tests

For screening tools, future diagnostic test accuracy studies should focus on evaluating CRP for tuberculosis screening in outpatients on ART and pregnant PLHIV, as only 1 study assessed CRP in outpatients on ART and no study assessed CRP in pregnant PLHIV. The

newly developed CRP-based CPMs in chapter 5 also require validation in outpatient PLHIV on ART and in outpatient PLHIV from PCF settings.

Further studies should focus on developing an accurate initial screening tool to guide confirmatory testing in outpatients on ART and HIV-positive inpatients. Possible tools may include biomarkers at different 'omics' levels and CPMs. In one study that assessed several candidate transcriptional signatures for tuberculosis among participants with symptoms and signs of tuberculosis, these signatures approximated or met the minimum WHO TPP for a triage test.[122] However, only 44 participants were PLHIV. In another study that recruited PLHIV irrespective of tuberculosis symptoms and signs from a community setting, the RISK11 blood transcriptional signature approached the minimum WHO TPP for a triage test.[123] The newly developed CPMs in chapter 5 may be used as a benchmark to evaluate emerging biomarkers for tuberculosis screening; CRP may also be combined with other biomarkers to improve predictive performance. Particular attention should be placed on HIV-positive inpatients; in chapter 3, no current screening tool had optimal accuracy in this population, and resource-limited settings might be unable to do systematic confirmatory testing with Xpert in all HIV-positive inpatients.

For confirmatory tests, only 1 study assessed Xpert Ultra. This study enrolled outpatients not on ART irrespective of tuberculosis symptoms and signs, showing that Xpert Ultra improved sensitivity over Xpert by 16 percentage points. No studies assessed Xpert Ultra in outpatients on ART, pregnant PLHIV, and HIV-positive inpatients regardless of tuberculosis symptoms and signs. Since Xpert Ultra improved sensitivity by a substantial margin over Xpert, further studies should assess the accuracy of Xpert Ultra in all PLHIV irrespective of tuberculosis symptoms and signs, including in key subgroups. Other molecular-based tests, such as TB-LAMP and Trunat assays, also require evaluation. In chapter 4, only 2 studies evaluated FujiLAM in 477 HIV-positive inpatients irrespective of tuberculosis symptoms and signs. In these 2 studies, FujiLAM increased sensitivity by 21 percentage points compared with AlereLAM. Another study in outpatient PLHIV not on ART has also showed that FujiLAM increased sensitivity by 23 percentage points compared with AlereLAM.[163]. Thus, future studies should assess the accuracy of FujiLAM in PLHIV irrespective of tuberculosis symptoms and signs.

The diagnostic accuracy of screening tools and confirmatory tests in different settings also warrants attention. For example, future studies should assess test accuracy in countries

outside of Africa, in settings with a low tuberculosis prevalence, and in the context of regular screening.

Finally, future diagnostic test accuracy studies – especially those conducted among inpatients – should be appropriately designed. Investigators should select a robust reference standard that includes several samples for culture and Xpert from both pulmonary and extrapulmonary sites. Furthermore, investigators should look to include participants who are unable to produce sputum for testing.

6.4.2 Further studies to assess utility of certain screening tools and strategies

Diagnostic test accuracy studies should be followed by studies that measure the consequences of a test. In other words, further studies are needed to determine the impact of tests or strategies on health outcomes and costs. In particular, randomised trials are needed to evaluate important health outcomes, such as morbidity, mortality. Other test characteristics (e.g., feasibility, acceptability) also require evaluation.[81]

Since chapters 3 and 4 show that currently available screening tools to guide Xpert and LF-LAM confirmatory testing have low accuracy in inpatients, the analyses in these chapters are likely sufficient to conclude that Xpert and AlereLAM for all inpatients has clinical value.

However, the clinical utility of screening tools in outpatient PLHIV might need to be determined. Chapter 2, for example, showed that the specificity of CRP was higher than W4SS in outpatients not on ART, and a parallel strategy of W4SS and chest X-ray had higher sensitivity but lower specificity in outpatients on ART. A randomized clinical trial might determine the effect of replacing W4SS with these approaches on several outcomes such as mortality, number of confirmatory tests needed, number of participants placed on IPT/TPT, number of bacteriologically confirmed tuberculosis cases, and time to diagnosis of tuberculosis. Furthermore, since decision makers need to consider the resource implications of these alternative approaches, health economic analyses (using randomized trial data or decision models) are also needed. These analyses may include cost-effectiveness, cost minimization, and/or budget impact analyses. Finally, future studies should evaluate the acceptability and feasibility of alternative tools and strategies for both the patient and the health system.

6.5 Conclusion

This thesis identified limitations of current tuberculosis screening strategies in PLHIV and defined alternative strategies using an IPDMA. The WHO-recommended W4SS has suboptimal sensitivity, specificity, and/or clinical utility in different subgroups of PLHIV. In outpatients not ART, CRP (either as a standalone test or combined with W4SS in a sequential strategy) reduces the need for further Xpert confirmatory testing compared with W4SS, without compromising sensitivity. In outpatients on ART, CRP data is scarce, but chest X-ray could be combined in parallel with W4SS, depending on available resources, because this strategy detects more tuberculosis cases than W4SS alone. In all outpatient PLHIV, the WHO tuberculosis screening and diagnostic algorithm (W4SS followed by Xpert if screen positive) has suboptimal sensitivity; Xpert for all would offer slight sensitivity gains but would have major resource implications. In HIV-positive medical inpatients, screening tools (including WHO screening criteria) to guide confirmatory testing have suboptimal accuracy and might complicate the tuberculosis diagnostic algorithm. Therefore, routine Xpert and AlereLAM confirmatory testing should be done in all HIV-positive medical inpatients. Although routine Xpert and AlereLAM testing would improve diagnostic yield, a negative result on both tests does not rule out tuberculosis. The newly developed FujiLAM (and Xpert Ultra) may substantially improve the rapid diagnosis of tuberculosis in this population. The findings of this thesis have informed the updated 2021 WHO guidelines on tuberculosis screening in PLHIV and led to 5 new or updated WHO recommendations.

CHAPTER 7: REFERENCES

1. World Health Organization. Global tuberculosis report 2021. Geneva: World Health Organization, 2021.
2. World Health Organization. What is DOTS?: a guide to understanding the WHO-recommended TB control strategy known as DOTS. Geneva: World Health Organization, 1999.
3. World Health Organisation. Global tuberculosis report 2015. Geneva: World Health Organisation, 2015.
4. Raviglione MC, Uplekar MW. WHO's new Stop TB Strategy. *Lancet (London, England)* 2006; **367**(9514): 952-5.
5. Clark H, Wu H. The sustainable development goals: 17 goals to transform our world. *Furthering the Work of the United Nations; UN: New York, NY, USA* 2016: 36-54.
6. Uplekar M, Weil D, Lonnroth K, et al. WHO's new end TB strategy. *Lancet (London, England)* 2015; **385**(9979): 1799-801.
7. United Nations. Political declaration of the high-level meeting of the General Assembly on the fight against tuberculosis. New York: United Nations, 2018.
8. Jeremiah C, Petersen E, Nantanda R, et al. The WHO Global Tuberculosis 2021 Report - not so good news and turning the tide back to End TB. *Int J Infect Dis* 2022.
9. Getahun H, Gunneberg C, Granich R, Nunn P. HIV infection-associated tuberculosis: the epidemiology and the response. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2010; **50 Suppl 3**: S201-7.
10. Schutz C, Meintjes G, Almajid F, Wilkinson RJ, Pozniak A. Clinical management of tuberculosis and HIV-1 co-infection. *Eur Respir J* 2010; **36**(6): 1460-81.
11. Bell LCK, Noursadeghi M. Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-infection. *Nat Rev Microbiol* 2018; **16**(2): 80-90.
12. 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.
13. World Health Organization. WHO global lists of high burden countries for tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB), 2021–2025: background document. Geneva: World Health Organization, 2021.
14. Loddenkemper R, Lipman M, Zumla A. Clinical Aspects of Adult Tuberculosis. *Cold Spring Harb Perspect Med* 2015; **6**(1): a017848.

15. Lawn SD, Zumla AI. Tuberculosis. *Lancet (London, England)* 2011; **378**(9785): 57-72.
16. Hamada Y, Getahun H, Tadesse BT, Ford N. HIV-associated tuberculosis. *Int J STD AIDS* 2021; **32**(9): 780-90.
17. Naing C, Mak JW, Maung M, Wong SF, Kassim AI. Meta-analysis: the association between HIV infection and extrapulmonary tuberculosis. *Lung* 2013; **191**(1): 27-34.
18. 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.
19. 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.
20. 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.
21. Khan FY. Review of literature on disseminated tuberculosis with emphasis on the focused diagnostic workup. *J Family Community Med* 2019; **26**(2): 83-91.
22. World Health Organization. Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB. Geneva: World Health Organization, 2013.
23. Swaminathan S, Padmapriyadarsini C, Narendran G. HIV-associated tuberculosis: clinical update. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2010; **50**(10): 1377-86.
24. Pai M, Behr MA, Dowdy D, et al. Tuberculosis. *Nat Rev Dis Primers* 2016; **2**: 16076.
25. 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.
26. World Health Organization. Implementing the end TB strategy: the essentials. Geneva: World Health Organization, 2015.

27. World Health Organization. Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings. Geneva: World Health Organization, 2011.
28. World Health Organization. Systematic screening for active tuberculosis: principles and recommendations. Geneva: World Health Organization, 2013.
29. Ho J, Fox GJ, Marais BJ. Passive case finding for tuberculosis is not enough. *Int J Mycobacteriol* 2016; **5**(4): 374-8.
30. World Health Organization. Screening programmes: a short guide. Increase effectiveness, maximize benefits and minimize harm. Geneva: World Health Organization, 2020.
31. Van't Hoog AH, Onozaki I, Lonnroth K. Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden. *BMC Infect Dis* 2014; **14**: 532.
32. Lonnroth K, Corbett E, Golub J, et al. Systematic screening for active tuberculosis: rationale, definitions and key considerations. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2013; **17**(3): 289-98.
33. 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.
34. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Geneva: World Health Organization, 2016.
35. MacLean E, Saravu K, Pai M. Diagnosing active tuberculosis in people living with HIV: an ongoing challenge. *Curr Opin HIV AIDS* 2019; **14**(1): 46-54.
36. Pai M, Nicol MP, Boehme CC. Tuberculosis Diagnostics: State of the Art and Future Directions. *Microbiol Spectr* 2016; **4**(5).
37. Chihota VN, Grant AD, Fielding K, et al. Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2010; **14**(8): 1024-31.
38. World Health Organization. Policy statement: automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF system. Geneva: World Health Organization, 2011.

39. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2014; (1): CD009593.
40. Kohli M, Schiller I, Dendukuri N, et al. Xpert((R)) MTB/RIF assay for extrapulmonary tuberculosis and rifampicin resistance. *Cochrane Database Syst Rev* 2018; **8**: CD012768.
41. Di Tanna GL, Khaki AR, Theron G, et al. Effect of Xpert MTB/RIF on clinical outcomes in routine care settings: individual patient data meta-analysis. *Lancet Glob Health* 2019; **7**(2): e191-e9.
42. World Health Organization. WHO meeting report of a technical expert consultation: non-inferiority analysis of Xpert MTB/RIF Ultra compared to Xpert MTB/RIF. Geneva: World Health Organization; 2017. Geneva: World Health Organization, 2017.
43. Zifodya JS, Kreniske JS, Schiller I, et al. Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. *Cochrane Database Syst Rev* 2021; **2**: CD009593.
44. Mishra H, Reeve BWP, Palmer Z, et al. Xpert MTB/RIF Ultra and Xpert MTB/RIF for diagnosis of tuberculosis in an HIV-endemic setting with a high burden of previous tuberculosis: a two-cohort diagnostic accuracy study. *Lancet Respir Med* 2020; **8**(4): 368-82.
45. Kohli M, Schiller I, Dendukuri N, et al. Xpert MTB/RIF Ultra and Xpert MTB/RIF assays for extrapulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2021; **1**: CD012768.
46. Lawn SD, Kerkhoff AD, Vogt M, Wood R. High diagnostic yield of tuberculosis from screening urine samples from HIV-infected patients with advanced immunodeficiency using the Xpert MTB/RIF assay. *J Acquir Immune Defic Syndr* 2012; **60**(3): 289-94.
47. Peter JG, Theron G, Muchinga TE, Govender U, Dheda K. The diagnostic accuracy of urine-based Xpert MTB/RIF in HIV-infected hospitalized patients who are smear-negative or sputum scarce. *PLoS One* 2012; **7**(7): e39966.
48. 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.
49. Theron G, Peter J, Calligaro G, et al. Determinants of PCR performance (Xpert MTB/RIF), including bacterial load and inhibition, for TB diagnosis using specimens from different body compartments. *Scientific reports* 2014; **4**: 5658.

50. Haraka F, Kakolwa M, Schumacher SG, et al. Impact of the diagnostic test Xpert MTB/RIF on patient outcomes for tuberculosis. *Cochrane Database Syst Rev* 2021; **5**: CD012972.
51. 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.
52. Albert H, Nathavitharana RR, Isaacs C, Pai M, Denkinger CM, Boehme CC. Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J* 2016; **48**(2): 516-25.
53. Chin DP, Hanson CL. Finding the Missing Tuberculosis Patients. *The Journal of infectious diseases* 2017; **216**(suppl_7): S675-S8.
54. Huddart S, MacLean E, Pai M. Location, location, location: tuberculosis services in highest burden countries. *Lancet Glob Health* 2016; **4**(12): e907-e8.
55. Clouse K, Blevins M, Lindegren ML, et al. Low implementation of Xpert MTB/RIF among HIV/TB co-infected adults in the International epidemiologic Databases to Evaluate AIDS (IeDEA) program. *PloS one* 2017; **12**(2): e0171384.
56. MacLean E, Kohli M, Weber SF, et al. Advances in Molecular Diagnosis of Tuberculosis. *J Clin Microbiol* 2020; **58**(10).
57. Penn-Nicholson A, Gomathi SN, Ugarte-Gil C, et al. A prospective multicentre diagnostic accuracy study for the Truenat tuberculosis assays. *Eur Respir J* 2021; **58**(5).
58. World Health Organization. WHO operational handbook on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection. Geneva: World Health Organization, 2020.
59. Cain KP, McCarthy KD, Heilig CM, et al. An algorithm for tuberculosis screening and diagnosis in people with HIV. *N Engl J Med* 2010; **362**(8): 707-16.
60. Notomi T, Okayama H, Masubuchi H, et al. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res* 2000; **28**(12): E63.
61. Shete PB, Farr K, Strnad L, Gray CM, Cattamanchi A. Diagnostic accuracy of TB-LAMP for pulmonary tuberculosis: a systematic review and meta-analysis. *BMC Infect Dis* 2019; **19**(1): 268.
62. World Health Organization. The use of loop-mediated isothermal amplification (TB-LAMP) for the diagnosis of pulmonary tuberculosis: policy guidance. Geneva: World Health Organization, 2016.

63. Bulterys MA, Wagner B, Redard-Jacot M, et al. Point-Of-Care Urine LAM Tests for Tuberculosis Diagnosis: A Status Update. *J Clin Med* 2019; **9**(1).
64. Minion J, Leung E, Talbot E, Dheda K, Pai M, Menzies D. Diagnosing tuberculosis with urine lipoarabinomannan: systematic review and meta-analysis. *Eur Respir J* 2011; **38**(6): 1398-405.
65. World Health Organization. The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV: policy guidance. Geneva: World Health Organization,, 2015.
66. World Health Organization. Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV: policy update (2019): evidence to decision tables. Geneva: World Health Organization, 2019.
67. Shah M, Hanrahan C, Wang ZY, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database Syst Rev* 2016; (5): CD011420.
68. Lawn SD, Gupta-Wright A. Detection of lipoarabinomannan (LAM) in urine is indicative of disseminated TB with renal involvement in patients living with HIV and advanced immunodeficiency: evidence and implications. *Trans R Soc Trop Med Hyg* 2016; **110**(3): 180-5.
69. 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.
70. Gupta-Wright A, Kerkhoff AD, Meintjes G, Corbett EL. Urinary Lipoarabinomannan Detection and Disseminated Nontuberculous Mycobacterial Disease. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018; **66**(1): 158.
71. Nel JS, Lippincott CK, Berhanu R, Spencer DC, Sanne IM, Ive P. Does Disseminated Nontuberculous Mycobacterial Disease Cause False-Positive Determine TB-LAM Lateral Flow Assay Results? A Retrospective Review. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2017; **65**(7): 1226-8.
72. Gupta-Wright A, Peters JA, Flach C, Lawn SD. Detection of lipoarabinomannan (LAM) in urine is an independent predictor of mortality risk in patients receiving treatment for HIV-associated tuberculosis in sub-Saharan Africa: a systematic review and meta-analysis. *BMC Med* 2016; **14**: 53.

73. 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.
74. 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.
75. 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.
76. 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).
77. Huerga H, Bastard M, Lubega AV, et al. Novel FujiLAM assay to detect tuberculosis in HIV-positive ambulatory patients in four African countries: a diagnostic accuracy study. *Lancet Glob Health* 2023; **11**(1): e126-e35.
78. Székely R, Sossen B, Mukoka M, et al. Multicentre accuracy trial of FUJIFILM SILVAMP TB LAM test in people with HIV reveals lot variability. *medRxiv* 2022.
79. 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.
80. Ellis PK, Martin WJ, Dodd PJ. CD4 count and tuberculosis risk in HIV-positive adults not on ART: a systematic review and meta-analysis. *PeerJ* 2017; **5**: e4165.
81. Nathavitharana RR, Yoon C, Macpherson P, et al. Guidance for Studies Evaluating the Accuracy of Tuberculosis Triage Tests. *J Infect Dis* 2019; **220**(220 Suppl 3): S116-S25.
82. 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.
83. 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.
84. Vashist SK, Venkatesh AG, Marion Schneider E, Beaudoin C, Luppa PB, Luong JH. Bioanalytical advances in assays for C-reactive protein. *Biotechnol Adv* 2016; **34**(3): 272-90.

85. Yoon C, Chaisson LH, Patel SM, et al. Diagnostic accuracy of C-reactive protein for active pulmonary tuberculosis: a meta-analysis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2017; **21**(9): 1013-9.
86. Lawn SD, Kerkhoff AD, Vogt M, Wood R. Diagnostic and prognostic value of serum C-reactive protein for screening for HIV-associated tuberculosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2013; **17**(5): 636-43.
87. Yoon C, Davis JL, Huang L, et al. Point-of-care C-reactive protein testing to facilitate implementation of isoniazid preventive therapy for people living with HIV. *J Acquir Immune Defic Syndr* 2014; **65**(5): 551-6.
88. Yoon C, Semitala FC, Atuhumuza E, et al. Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *Lancet Infect Dis* 2017; **17**(12): 1285-92.
89. Yoon C, Semitala FC, Asege L, et al. Yield and Efficiency of Novel Intensified Tuberculosis Case-Finding Algorithms for People Living with HIV. *Am J Respir Crit Care Med* 2019; **199**(5): 643-50.
90. Shapiro AE, Hong T, Govere S, et al. C-reactive protein as a screening test for HIV-associated pulmonary tuberculosis prior to antiretroviral therapy in South Africa. *AIDS (London, England)* 2018; **32**(13): 1811-20.
91. Gersh JK, Barnabas RV, Matemo D, et al. Pulmonary tuberculosis screening in anti-retroviral treated adults living with HIV in Kenya. *BMC Infect Dis* 2021; **21**(1): 218.
92. World Health Organization. Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches. Geneva: World Health Organization, 2016.
93. van't Hoog AH, Langendam M, Mitchell E, et al. Symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative adults and adults with unknown HIV status. *Cochrane Database of Systematic Reviews* 2014; (1).
94. Lawn SD, Edwards DJ, Kranzer K, Vogt M, Bekker LG, Wood R. Urine lipoarabinomannan assay for tuberculosis screening before antiretroviral therapy diagnostic yield and association with immune reconstitution disease. *AIDS (London, England)* 2009; **23**(14): 1875-80.

95. Lewis JJ, Charalambous S, Day JH, et al. HIV infection does not affect active case finding of tuberculosis in South African gold miners. *Am J Respir Crit Care Med* 2009; **180**(12): 1271-8.
96. Shah S, Demissie M, Lambert L, et al. Intensified tuberculosis case finding among HIV-Infected persons from a voluntary counseling and testing center in Addis Ababa, Ethiopia. *J Acquir Immune Defic Syndr* 2009; **50**(5): 537-45.
97. Perlman DC, el-Sadr WM, Nelson ET, et al. Variation of chest radiographic patterns in pulmonary tuberculosis by degree of human immunodeficiency virus-related immunosuppression. The Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA). The AIDS Clinical Trials Group (ACTG). *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 1997; **25**(2): 242-6.
98. Post FA, Wood R, Pillay GP. Pulmonary tuberculosis in HIV infection: radiographic appearance is related to CD4+ T-lymphocyte count. *Tuber Lung Dis* 1995; **76**(6): 518-21.
99. Greenberg SD, Frager D, Suster B, Walker S, Stavropoulos C, Rothpearl A. Active pulmonary tuberculosis in patients with AIDS: spectrum of radiographic findings (including a normal appearance). *Radiology* 1994; **193**(1): 115-9.
100. San KE, Muhamad M. Pulmonary Tuberculosis in HIV Infection : The Relationship of the Radiographic Appearance to CD4 T-Lymphocytes Count. *Malays J Med Sci* 2001; **8**(1): 34-40.
101. Sakurada S, Hang NT, Ishizuka N, et al. Inter-rater agreement in the assessment of abnormal chest X-ray findings for tuberculosis between two Asian countries. *BMC Infect Dis* 2012; **12**: 31.
102. Tavaziva G, Harris M, Abidi SK, et al. Chest X-ray Analysis With Deep Learning-Based Software as a Triage Test for Pulmonary Tuberculosis: An Individual Patient Data Meta-Analysis of Diagnostic Accuracy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2022; **74**(8): 1390-400.
103. van Smeden M, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. *J Clin Epidemiol* 2021; **132**: 142-5.
104. Steyerberg EW. Clinical prediction models: Springer; 2019.
105. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; **35**(29): 1925-31.
106. Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagn Progn Res* 2019; **3**: 16.

107. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162**(1): W1-73.
108. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.
109. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26**(6): 565-74.
110. Zamanipoor Najafabadi AH, Ramspek CL, Dekker FW, et al. TRIPOD statement: a preliminary pre-post analysis of reporting and methods of prediction models. *BMJ Open* 2020; **10**(9): e041537.
111. Auld AF, Kerkhoff AD, Hanifa Y, et al. Derivation and external validation of a risk score for predicting HIV-associated tuberculosis to support case finding and preventive therapy scale-up: A cohort study. *PLoS Med* 2021; **18**(9): e1003739.
112. Baik Y, Rickman HM, Hanrahan CF, et al. A clinical score for identifying active tuberculosis while awaiting microbiological results: Development and validation of a multivariable prediction model in sub-Saharan Africa. *PLoS Med* 2020; **17**(11): e1003420.
113. Hanifa Y, Fielding KL, Chihota VN, et al. A clinical scoring system to prioritise investigation for tuberculosis among adults attending HIV clinics in South Africa. *PLoS One* 2017; **12**(8): e0181519.
114. Balcha TT, Skogmar S, Sturegard E, et al. A Clinical Scoring Algorithm for Determination of the Risk of Tuberculosis in HIV-Infected Adults: A Cohort Study Performed at Ethiopian Health Centers. *Open Forum Infect Dis* 2014; **1**(3): ofu095.
115. Boyles TH, Nduna M, Pitsi T, Scott L, Fox MP, Maartens G. A Clinical Prediction Score Including Trial of Antibiotics and C-Reactive Protein to Improve the Diagnosis of Tuberculosis in Ambulatory People With HIV. *Open Forum Infect Di* 2020; **7**(2).
116. Nanta S, Kantipong P, Pathipvanich P, Ruengorn C, Tawichasri C, Patumanond J. Screening scheme development for active TB prediction among HIV-infected patients. *Southeast Asian J Trop Med Public Health* 2011; **42**(4): 867-75.
117. Cobelens F, Kerkhoff AD. Tuberculosis and anemia-cause or effect? *Environ Health Prev Med* 2021; **26**(1): 93.
118. Kerkhoff AD, Meintjes G, Opie J, et al. Anaemia in patients with HIV-associated TB: relative contributions of anaemia of chronic disease and iron deficiency. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2016; **20**(2): 193-201.

119. World Health Organization. Vitamin and mineral nutrition information system. Geneva: World Health Organization, 2011.
120. Gelaw Y, Getaneh Z, Melku M. Anemia as a risk factor for tuberculosis: a systematic review and meta-analysis. *Environ Health Prev Med* 2021; **26**(1): 13.
121. Kerkhoff AD, Wood R, Vogt M, Lawn SD. Predictive value of anemia for tuberculosis in HIV-infected patients in Sub-Saharan Africa: an indication for routine microbiological investigation using new rapid assays. *J Acquir Immune Defic Syndr* 2014; **66**(1): 33-40.
122. Turner CT, Gupta RK, Tsaliki E, et al. Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. *Lancet Respir Med* 2020; **8**(4): 407-19.
123. Mendelsohn SC, Fiore-Gartland A, Penn-Nicholson A, et al. Validation of a host blood transcriptomic biomarker for pulmonary tuberculosis in people living with HIV: a prospective diagnostic and prognostic accuracy study. *Lancet Glob Health* 2021; **9**(6): e841-e53.
124. Suthar AB, Lawn SD, del Amo J, et al. Antiretroviral therapy for prevention of tuberculosis in adults with HIV: a systematic review and meta-analysis. *PLoS Med* 2012; **9**(7): e1001270.
125. Kufa T, Mabuto T, Muchiri E, et al. Incidence of HIV-associated tuberculosis among individuals taking combination antiretroviral therapy: a systematic review and meta-analysis. *PLoS One* 2014; **9**(11): e111209.
126. Gupta A, Wood R, Kaplan R, Bekker LG, Lawn SD. Tuberculosis incidence rates during 8 years of follow-up of an antiretroviral treatment cohort in South Africa: comparison with rates in the community. *PLoS One* 2012; **7**(3): e34156.
127. Cegielski JP, McMurray DN. The relationship between malnutrition and tuberculosis: evidence from studies in humans and experimental animals. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2004; **8**(3): 286-98.
128. Sinha P, Davis J, Saag L, et al. Undernutrition and Tuberculosis: Public Health Implications. *J Infect Dis* 2019; **219**(9): 1356-63.
129. Carwile ME, Hochberg NS, Sinha P. Undernutrition is feeding the tuberculosis pandemic: A perspective. *J Clin Tuberc Other Mycobact Dis* 2022; **27**: 100311.
130. Purnell JQ. Definitions, classification, and epidemiology of obesity. *Endotext [Internet]* 2018.

131. Alebel A, Demant D, Petrucka P, Sibbritt D. Effects of undernutrition on mortality and morbidity among adults living with HIV in sub-Saharan Africa: a systematic review and meta-analysis. *BMC Infect Dis* 2021; **21**(1): 1.
132. Suresh PK, Poojary S, Basavaiah SH, Kini JR, Lobo FD, Sahu KK. Utility of fine-needle aspiration cytology in the diagnosis of HIV lymphadenopathy. *Diagn Cytopathol* 2019; **47**(10): 1011-7.
133. Swindells S, Komarow L, Tripathy S, et al. Screening for pulmonary tuberculosis in HIV-infected individuals: AIDS Clinical Trials Group Protocol A5253. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2013; **17**(4): 532-9.
134. World Health Organization. Improving the diagnosis and treatment of smear-negative pulmonary and extrapulmonary tuberculosis among adults and adolescents: recommendations for HIV-prevalent and resource-constrained settings. Geneva: World Health Organization, 2007.
135. Floridia M, Ciccacci F, Andreotti M, et al. Tuberculosis Case Finding With Combined Rapid Point-of-Care Assays (Xpert MTB/RIF and Determine TB LAM) in HIV-Positive Individuals Starting Antiretroviral Therapy in Mozambique. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2017; **65**(11): 1878-83.
136. Lawn SD, Brooks SV, Kranzer K, et al. Screening for HIV-associated tuberculosis and rifampicin resistance before antiretroviral therapy using the Xpert MTB/RIF assay: a prospective study. *PLoS Med* 2011; **8**(7): e1001067.
137. 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.
138. Riley RD, Stewart LA, Tierney JF. Individual Participant Data Meta-Analysis for Healthcare Research. *Individual Participant Data Meta-Analysis: A Handbook for Healthcare Research* 2021: 1-6.
139. Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPDM-aMg. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med* 2015; **12**(10): e1001886.
140. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008; **61**(1): 76-86.

141. Ford N, Shubber Z, Meintjes G, et al. Causes of hospital admission among people living with HIV worldwide: a systematic review and meta-analysis. *The lancet HIV* 2015; **2**(10): e438-44.
142. Mwaura M, Engel N. Constructing confidence: User perspectives on AlereLAM testing for tuberculosis. *Int J Infect Dis* 2021; **112**: 237-42.
143. World Health Organization. Global tuberculosis report 2020. Geneva: World Health Organization, 2020.
144. World Health Organization. WHO consolidated guidelines on tuberculosis: tuberculosis preventive treatment: module 1: prevention: tuberculosis preventive treatment. Geneva: World Health Organization, 2020.
145. Rangaka MX, Wilkinson RJ, Glynn JR, et al. Effect of antiretroviral therapy on the diagnostic accuracy of symptom screening for intensified tuberculosis case finding in a South African HIV clinic. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012; **55**(12): 1698-706.
146. Shine B, de Beer FC, Pepys MB. Solid phase radioimmunoassays for human C-reactive protein. *Clin Chim Acta* 1981; **117**(1): 13-23.
147. Claus DR, Osmand AP, Gewurz H. Radioimmunoassay of human C-reactive protein and levels in normal sera. *J Lab Clin Med* 1976; **87**(1): 120-8.
148. 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.
149. 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.
150. 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.
151. Higgins JP. Cochrane handbook for systematic reviews of interventions version 5.0.: The Cochrane Collaboration; 2008.
152. 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.
153. Lin L, Chu H. Meta-analysis of Proportions Using Generalized Linear Mixed Models. *Epidemiology* 2020; **31**(5): 713-7.
154. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**(7414): 557-60.

155. 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.
156. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20**(19): 2865-84.
157. Chu H, Nie L, Cole SR, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Stat Med* 2009; **28**(18): 2384-99.
158. Schunemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J Clin Epidemiol* 2020; **122**: 142-52.
159. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; **56**(2): 455-63.
160. Affolabi D, Wachinou AP, Bekou W, et al. Screening tuberculosis in HIV infected patients: which algorithms work best? A multicountry survey in Benin, Guinea and Senegal (RAFAscreen project). 49th World Conference On Lung Health Of The International Union Against Tuberculosis And Lung Disease. The Hague, The Netherlands; 24–27 October, 2018.
161. Ahmad Khan F, Verkuijl S, Parrish A, et al. Performance of symptom-based tuberculosis screening among people living with HIV: not as great as hoped. *AIDS (London, England)* 2014; **28**(10): 1463-72.
162. Al-Darraji HA, Abd Razak H, Ng KP, Altice FL, Kamarulzaman A. The diagnostic performance of a single GeneXpert MTB/RIF assay in an intensified tuberculosis case finding survey among HIV-infected prisoners in Malaysia. *PLoS One* 2013; **8**(9): e73717.
163. Bjerrum S, Broger T, Szekely R, et al. Diagnostic Accuracy of a Novel and Rapid Lipoarabinomannan Test for Diagnosing Tuberculosis Among People With Human Immunodeficiency Virus. *Open Forum Infect Dis* 2020; **7**(1): ofz530.
164. Hanifa Y, Fielding KL, Charalambous S, et al. Tuberculosis among adults starting antiretroviral therapy in South Africa: the need for routine case finding. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2012; **16**(9): 1252-9.
165. Hoffmann CJ, Variava E, Rakgokong M, et al. High prevalence of pulmonary tuberculosis but low sensitivity of symptom screening among HIV-infected pregnant women in South Africa. *PLoS One* 2013; **8**(4): e62211.

166. Kempker RR, Chkhartishvili N, Kinkladze I, et al. High Yield of Active Tuberculosis Case Finding Among HIV-Infected Patients Using Xpert MTB/RIF Testing. *Open Forum Infect Dis* 2019; **6**(6): ofz233.
167. Kerkhoff AD, Wood R, Lowe DM, Vogt M, Lawn SD. Blood neutrophil counts in HIV-infected patients with pulmonary tuberculosis: association with sputum mycobacterial load. *PLoS One* 2013; **8**(7): e67956.
168. Kufa T, Mngomezulu V, Charalambous S, et al. Undiagnosed tuberculosis among HIV clinic attendees: association with antiretroviral therapy and implications for intensified case finding, isoniazid preventive therapy, and infection control. *J Acquir Immune Defic Syndr* 2012; **60**(2): e22-8.
169. LaCourse SM, Cranmer LM, Matemo D, et al. Tuberculosis Case Finding in HIV-Infected Pregnant Women in Kenya Reveals Poor Performance of Symptom Screening and Rapid Diagnostic Tests. *J Acquir Immune Defic Syndr* 2016; **71**(2): 219-27.
170. Mbu ET, Sauter F, Zoufaly A, et al. Tuberculosis in people newly diagnosed with HIV at a large HIV care and treatment center in Northwest Cameroon: Burden, comparative screening and diagnostic yields, and patient outcomes. *PLoS One* 2018; **13**(6): e0199634.
171. Modi S, Cavanaugh JS, Shiraishi RW, et al. Performance of Clinical Screening Algorithms for Tuberculosis Intensified Case Finding among People Living with HIV in Western Kenya. *PLoS One* 2016; **11**(12): e0167685.
172. Nguyen DT, Bang ND, Hung NQ, Beasley RP, Hwang LY, Graviss EA. Yield of chest radiograph in tuberculosis screening for HIV-infected persons at a district-level HIV clinic. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2016; **20**(2): 211-7.
173. Reeve B, Ndlangalavu G, Palmer Z, et al. Accuracy of Xpert Ultra and Xpert MTB/RIF in people living with HIV initiating antiretroviral treatment who have minimal TB symptoms. 50th World Conference On Lung Health Of The International Union Against Tuberculosis And Lung Disease. Hyderabad, India; Oct 30–Nov 2, 2019. p. S115.
174. 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.
175. Calnan M. Developing strategies for TB screening among HIV-infected and HIV-uninfected pregnant and postpartum women in Swaziland. 47th World Conference on Lung Health of the International Union Against Tuberculosis and Lung Disease; 2016; 2016.

176. Drain PK, Losina E, Coleman SM, et al. Rapid urine lipoarabinomannan assay as a clinic-based screening test for active tuberculosis at HIV diagnosis. *BMC Pulm Med* 2016; **16**(1): 147.
177. Telisinghe L, Fielding KL, Malden JL, et al. High tuberculosis prevalence in a South African prison: the need for routine tuberculosis screening. *PLoS One* 2014; **9**(1): e87262.
178. Médecins Sans Frontières. Out of Step: TB policies in 29 countries. Geneva: MSF Access Campaign; 2017.
179. van Griensven J, Cnops L, De Weggheleire A, Declercq S, Bottieau E. Point-of-Care Biomarkers to Guide Antibiotic Prescription for Acute Febrile Illness in Sub-Saharan Africa: Promises and Caveats. *Open Forum Infect Dis* 2020; **7**(8): ofaa260.
180. Wasserman S, Meintjes G. The diagnosis, management and prevention of HIV-associated tuberculosis. *S Afr Med J* 2014; **104**(12): 886-93.
181. Hanrahan CF, Golub JE, Mohapi L, et al. Body mass index and risk of tuberculosis and death. *AIDS (London, England)* 2010; **24**(10): 1501-8.
182. Kerkhoff AD, Wood R, Cobelens FG, Gupta-Wright A, Bekker LG, Lawn SD. The predictive value of current haemoglobin levels for incident tuberculosis and/or mortality during long-term antiretroviral therapy in South Africa: a cohort study. *BMC Med* 2015; **13**: 70.
183. Lawn SD, Kerkhoff AD, Wood R. Progression of subclinical culture-positive tuberculosis to symptomatic disease in HIV-infected individuals. *AIDS (London, England)* 2011; **25**(17): 2190-1.
184. Oni T, Burke R, Tsekela R, et al. High prevalence of subclinical tuberculosis in HIV-1-infected persons without advanced immunodeficiency: implications for TB screening. *Thorax* 2011; **66**(8): 669-73.
185. World Health Organization. WHO operational handbook on tuberculosis: module 2: screening: systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021.
186. 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.
187. 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.

188. 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.
189. 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.
190. 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.
191. 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.
192. 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.
193. Mendelson F, Griesel R, Tiffin N, et al. C-reactive protein and procalcitonin to discriminate between tuberculosis, Pneumocystis jirovecii pneumonia, and bacterial pneumonia in HIV-infected inpatients meeting WHO criteria for seriously ill: a prospective cohort study. *BMC Infect Dis* 2018; **18**(1): 399.
194. Santos VS, Goletti D, Kontogianni K, et al. Acute phase proteins and IP-10 as triage tests for the diagnosis of tuberculosis: systematic review and meta-analysis. *Clin Microbiol Infect* 2019; **25**(2): 169-77.
195. 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.
196. 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.
197. Van Hoving DJ, Griesel R, Meintjes G, Takwoingi Y, Maartens G, Ochodo EA. Abdominal ultrasound for diagnosing abdominal tuberculosis or disseminated tuberculosis with abdominal involvement in HIV-positive individuals. *Cochrane Database Syst Rev* 2019; **9**: CD012777.

198. World Health Organization. WHO operational handbook on tuberculosis. Module 3: diagnosis - rapid diagnostics for tuberculosis detection, 2021 update. Geneva: World Health Organization, 2021.
199. 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. *The lancet HIV* 2021(In press).
200. 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.
201. 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.
202. 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.
203. 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.
204. Aunsborg JW, Honge BL, Jespersen S, et al. A clinical score has utility in tuberculosis case-finding among patients with HIV: A feasibility study from Bissau. *Int J Infect Dis* 2020; **92S**: S78-S84.
205. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Ann Intern Med* 2015; **162**(10): 735-6.
206. Ahmed I, Debray TP, Moons KG, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. *BMC medical research methodology* 2014; **14**: 3.
207. Debray TP, Moons KG, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Statistics in medicine* 2013; **32**(18): 3158-80.
208. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; **368**: m441.
209. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons; 2004.
210. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. *Biom J* 2018; **60**(3): 431-49.

211. Debray TP, Damen JA, Snell KI, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017; **356**: i6460.
212. Snell KI, Hua H, Debray TP, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol* 2016; **69**: 40-50.
213. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019; **3**: 18.
214. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; **352**: i6.
215. Kerkhoff AD, Cattamanchi A, Muyoyeta M, Denkinger CM, Dowdy DW. Validating novel diagnostic assays for tuberculosis in the context of existing tools. *Lancet Glob Health* 2021; **9**(9): e1209.
216. Joint United Nations Programme on HIV/AIDS (UNAIDS). 2020 Global AIDS Update - Seizing the Moment - Tackling entrenched inequalities to end epidemics, 2020, 2020.
217. Dendukuri N, Schiller I, de Groot J, et al. Concerns about composite reference standards in diagnostic research. *BMJ* 2018; **360**: j5779.
218. MacLean EL, Kohli M, Koppel L, et al. Bayesian latent class analysis produced diagnostic accuracy estimates that were more interpretable than composite reference standards for extrapulmonary tuberculosis tests. *Diagn Progn Res* 2022; **6**(1): 11.
219. Huddy JR, Ni MZ, Barlow J, Majeed A, Hanna GB. Point-of-care C reactive protein for the diagnosis of lower respiratory tract infection in NHS primary care: a qualitative study of barriers and facilitators to adoption. *BMJ Open* 2016; **6**(3): e009959.
220. Blanc FX, Badje AD, Bonnet M, et al. Systematic or Test-Guided Treatment for Tuberculosis in HIV-Infected Adults. *N Engl J Med* 2020; **382**(25): 2397-410.
221. Grant AD, Charalambous S, Tlali M, et al. Algorithm-guided empirical tuberculosis treatment for people with advanced HIV (TB Fast Track): an open-label, cluster-randomised trial. *The lancet HIV* 2020; **7**(1): e27-e37.
222. Hosseinipour MC, Bisson GP, Miyahara S, et al. Empirical tuberculosis therapy versus isoniazid in adult outpatients with advanced HIV initiating antiretroviral therapy (REMEMBER): a multicountry open-label randomised controlled trial. *Lancet (London, England)* 2016; **387**(10024): 1198-209.

CHAPTER 8: APPENDICES

8.1 Appendix for Chapter 2

Table 8-1: 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 8-2: 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}
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}
sputum xpert result	{positive, negative, NA}, indeterminate = negative
sputum xpert ultra result	{positive, negative, NA}, indeterminate = negative
sputum culture result	{positive, negative, NA}, contaminated culture = negative
non-sputum xpert result	{positive, negative, NA}, indeterminate = negative
non-sputum xpert ultra 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 8-3: Study-level characteristics

Author, year	Country	Study period	Study population	Study setting	Exclusion criteria	Sputum culture	Sputum Xpert	Non-sputum culture/Xpert	Liquid or solid culture
Abed Al-Darraji, 2013[162]	Malaysia	2012-2012	PLHIV who are inmates	1 prison in Malaysia	Receiving ATT, anticipated release within 24 hours	2 spot samples	1 morning sample	No	Liquid
Affolabi, 2018[160]	Multi-country*	2015-2018	PLHIV of any age	3 HIV clinics	On ATT for TB	1 spot and 1 early morning samples	1 spot and 1 early morning samples	No	Solid
Ahmad, 2014[161]	South Africa	2011-2012	PLHIV aged ≥18 years	2 HIV clinics	On ATT currently or awaiting results of TB investigations	1 morning sample	No	No	Liquid
Balcha, 2014[114]	Ethiopia	2011-2013	ART-naive PLHIV aged ≥18 years with WHO stage 4 or CD4 cell count <350 per µL	5 public health centers	Unable to submit ≥ 1 pair of sputum samples, previous ART, ATT >2 weeks	1 morning sample	1 morning sample	FNA sample of LN >1cm for culture and Xpert	Liquid
Bjerrum, 2015[191]	Ghana	2013-2014	ART-naive PLHIV aged ≥18 years with WHO stage 3/4 or CD4 cell count ≤350 per µL or pregnant	1 public hospital (out- and inpatient departments)	On ATT > 2 days <3 months before or unable to produce sputum or urine samples	1 spot and 1 early morning samples	1 or 2 samples	No	Both
Gersh, 2018[91]	Kenya	2017-2018	PLHIV aged ≥18 and ≤70 years	2 HIV clinics	Currently on treatment for TB or LTBI, pregnant, incarcerated	1 spot sample, induced or early morning if necessary	1 spot sample, induced or early morning if necessary	No	Liquid
Hanifa, 2012[164]	South Africa	2007-2008	ART-naive PLHIV aged >17 years with WHO stage 4 or CD4 cell count ≤200 per µL	1 public HIV clinic	On ATT currently or <3 months before	2 spot samples	No	No	Liquid
Heidebrecht, 2016[192]	South Africa	2013-2013	Inpatient PLHIV admitted to medical wards	1 hospital (5 medical wards)	≥3 doses of ATT	1 spot sample	1 spot sample	No	Both
Hoffman, 2013[165]	South Africa	2010-2011	Pregnant PLHIV aged ≥18 years	16 PHC centre prenatal clinics and 1 regional hospital	-	1 spot sample	No	No	Liquid
Kempker, 2019[166]	Georgia	2014-2015	Newly diagnosed ART-naive PLHIV aged ≥18 years	1 national reference center for HIV diagnosis and treatment	ATT <90 days before	1 spot and 1 early morning samples	1 spot and 1 early morning samples for 1 test	No	Solid
Kerkhoff, 2013[167]	South Africa	2010-2011	ART-naive PLHIV aged ≥18 years	1 community-based ART clinic	No current TB diagnosis	2 spot samples with ≥ 1 induced	2 spot samples with ≥ 1 induced	No	Liquid
Kufa, 2012[168]	South Africa	2009-2010	PLHIV aged ≥18 years	1 HIV clinic in PHC centre	On TB treatment, completed TB treatment <3 months before, dialysis, prisoners	2 spot samples, induced if necessary	No	Blood culture	Liquid
LaCourse, 2016[169]	Kenya	2013-2014	Pregnant PLHIV aged ≥16 years	2 antenatal care clinics	On ATT for TB or LTBI, or were treated for TB or LTBI <1 year before	1 spot and 1 early morning samples	1 spot sample, on 2nd sample if no spot sample or if 2nd sample culture positive for TB	No	Liquid

Author, year	Country	Study period	Study population	Study setting	Exclusion criteria	Sputum culture	Sputum Xpert	Non-sputum culture/Xpert	Liquid or solid culture
Mbu, 2018[170]	Cameroon	2012-2013	ART-naive PLHIV aged ≥18 years	1 regional hospital (HIV testing and ART treatment center)	First diagnosis of HIV <1 month before, currently on ATT	1 spot and 1 early morning samples	No	No	Both
Modi, 2016[171]	Kenya	2011-2012	ART-naïve PLHIV aged ≥7 years	15 public HIV care and treatment clinics	Receipt of any HIV-related care <2 years before, ATT <1 year before	1 spot and 1 early morning samples	1 spot and 1 early morning samples	No	Liquid
Nguyen, 2016[172]	Vietnam	2009-2010	PLHIV aged ≥15 years	1 outpatient urban HIV clinic	Screened for TB in past month, received TB treatment <1 year before	1 spot and 1 early morning samples	No	No	Solid
Rangaka, 2012[145]	South Africa	2007-2009	PLHIV aged ≥18 years	1 HIV clinic	-	1 spot sample, induced if necessary	No	No	Both
Reeve, 2019[173]	South Africa	2017-2020	ART-naive PLHIV aged ≥18 years	1 outpatient clinic	On ATT <60 days before or has unknown treatment status	2 spot samples, majority induced	1 spot sample and Xpert Ultra on 1 spot sample	Urine Xpert Ultra	Liquid
Shapiro, 2018[90]	South Africa	2014-2015	ART-naive PLHIV aged ≥18 years	1 urban HIV clinic	-	2 samples, induced if necessary	No	No	Liquid
Swindells, 2013[133]	Multi-country**	2010-2010	ART-naive PLHIV aged ≥13 years	11 outpatient clinics	ART or diagnosis of TB <90 days before and current or recent receipt of ATT	3 spot samples, induced if necessary	No	No	Both
Thit, 2017[174]	Myanmar	2015-2015	Inpatient or outpatient PLHIV	1 tertiary referral hospital	-	1 spot sample	1 spot sample	No	Solid
Yoon, 2018[88]	Uganda	2013-2016	ART-naive PLHIV aged ≥18 years with CD4 cell count ≤350 per μL	2 HIV clinics in Kampala	Diagnosis of active tuberculosis, taking ATT (anti-TB or TB preventive therapy, fluoroquinolones) ≤3 days before	2 spot samples, 2nd induced if necessary	1 spot sample, induced if necessary	No	Both

*Benin, Guinea, and Senegal

**Botswana, Malawi, South Africa, Zimbabwe, India, Brazil and Peru

Definition of abbreviations: ART = antiretroviral therapy, ATT = anti-tuberculosis treatment, FNA = fine needle aspiration, LN = Lymph node, LTBI = latent tuberculosis infection, PHC = primary health care, PLHIV = people living with HIV, TB = tuberculosis

Table 8-4: Percentage of missing data for each variable by study†§

Variable	Affolabi	Ahmad	Al_Darraji	Balcha	Bjerrum	Gersh	Hanifa	Hoffmann	Kemper	Kerkhoff	Kufa	LaCourse	Mbu	Modi	Nguyen	Rangaka	Reeve	Shapiro	Swindells	Thit	Yoon
Clinical setting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Age	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ART status	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
History of tuberculosis	100	0	0	1	1	1	0	0	0	0	0	6	100	100	0	0	0	0	0	0	0
Currently smoking	0	100	0	0	24	0	0	0	0	0	100	100	100	100	1	100	0	0	100	0	8
Pregnancy*	100	100	100	100	0	0	100	0	100	1	0	0	100	4	0	100	100	100	0	100	0
Currently on IPT	1	0	100	0	100	0	100	100	100	100	100	0	100	100	100	100	100	100	0	0	0
W4SS**	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Cough	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
Fever	0	0	0	0	2	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
Weight loss	0	0	0	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	4	0	0
Night sweats	0	0	0	0	1	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
Cough >=2 weeks	100	100	0	100	0	0	0	0	0	0	5	0	0	4	0	0	0	0	0	100	0
Body mass index	1	6	0	1	1	1	1	100	2	0	3	0	100	13	100	0	0	0	1	0	0
Lymphadenopathy	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100
CD4 cell count	3	0	0	1	2	18	1	2	0	0	0	17	0	7	0	0	10	3	0	0	0
CRP	100	100	100	100	100	1	100	100	100	4	100	100	100	100	100	100	8	0	100	100	0
Hb	100	100	100	6	9	5	1	8	100	6	0	100	100	100	100	100	100	100	0	21	100
CXR (any abnormality)	3	3	100	100	100	100	1	100	100	9	7	100	100	23	100	100	100	100	4	7	100
CXR (suggests tuberculosis)	3	3	100	100	100	100	1	100	100	9	8	100	100	23	0	100	100	100	4	100	100
Total Xpert***	1	100	0	1	65	1	100	100	1	0	100	1	100	0	100	100	6	100	100	0	0
Total culture	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Total (culture±Xpert)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	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

*Pregnancy could be ascertained by testing or interview; missing percentages based on available data for females in the study

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

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

Definition of abbreviations: ART = antiretroviral therapy, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, IPT = Isoniazid preventive therapy, W4SS = WHO four-symptom screen

Table 8-5: Summary of main characteristics for participants overall and by each study

Variable	All	Ahmad	Al Darraji	Balcha	Bjerrum	Gersh	Hanifa	Hoffmann	Kemper	Kerkhoff	Kufa
Participants	15666 (100)	611 (3.9)	125 (0.8)	812 (5.2)	395 (2.5)	387 (2.5)	351 (2.2)	1404 (9)	103 (0.7)	523 (3.3)	415 (2.6)
Clinical setting											
Outpatient	15541 (99.2)	611 (100)	0 (0)	812 (100)	395 (100)	387 (100)	351 (100)	1404 (100)	103 (100)	523 (100)	415 (100)
Other setting*	125 (0.8)	0 (0)	125 (100)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
N	15666	611	125	812	395	387	351	1404	103	523	415
Age (years)	34 (28-42)	37 (31-45)	37 (33-42)	32 (28-40)	38 (31-45)	37 (31-45)	38 (32-46)	27 (23-32)	42 (35-49)	34 (28-41)	37 (31-44)
N	15666	611	125	812	395	387	351	1404	103	523	415
Female	10388 (66.3)	461 (75.5)	12 (9.6)	476 (58.6)	264 (66.8)	225 (58.1)	232 (66.1)	1404 (100.0)	26 (25.2)	335 (64.1)	274 (66.0)
N	15666	611	125	812	395	387	351	1404	103	523	415
On ART	4347 (27.8)	423 (69.2)	19 (15.2)	0 (0.0)	0 (0.0)	383 (99.7)	0 (0.0)	2 (0.1)	0 (0.0)	0 (0.0)	210 (50.6)
N	15663	611	125	812	395	384	351	1404	103	523	415
CD4 count (cells/μL)	269 (142-439)	313 (178-472)	338 (150-492)	208 (117-320)	140 (45-268)	404 (276-558)	119 (72-168)	395 (271-533)	120 (26-272)	169 (96-232)	216 (108-350)
N	15281	611	125	807	387	317	349	1377	103	521	415
History of tuberculosis	1955 (17.5)	235 (38.5)	36 (28.8)	51 (6.4)	23 (5.9)	58 (15.1)	99 (28.2)	110 (7.9)	3 (2.9)	141 (27.0)	109 (26.3)
N	11148	611	125	800	392	384	351	1401	103	523	415
Current Smoker	1191 (11.6)	-	117 (93.6)	35 (4.3)	11 (3.7)	11 (2.8)	40 (11.4)	40 (2.8)	64 (62.1)	122 (23.4)	-
N	10301		125	812	299	387	351	1404	103	522	
Pregnant	1938 (35.1)	-	-	-	24 (6.1)	0 (0.0)	-	1404 (100.0)	-	20 (3.8)	11 (2.7)
N	5519				395	387		1404		521	415
On IPT	41 (0.5)	3 (0.5)	-	19 (2.3)	-	0 (0.0)	-	-	-	-	-
N	7593	609		810		387					
W4SS	8028 (51.3)	331 (54.2)	85 (68.0)	651 (80.2)	359 (91.1)	47 (12.1)	331 (94.3)	227 (16.2)	62 (60.2)	452 (86.4)	355 (85.5)
N	15652	611	125	812	394	387	351	1404	103	523	415
Cough	4629 (29.6)	223 (36.5)	62 (49.6)	326 (40.2)	172 (43.7)	36 (9.3)	161 (45.9)	110 (7.8)	29 (28.2)	260 (49.7)	201 (48.4)
N	15623	611	125	811	394	387	351	1404	103	523	415
Fever	3391 (21.7)	57 (9.3)	37 (29.6)	389 (48.0)	189 (48.6)	5 (1.3)	100 (28.5)	52 (3.7)	55 (53.4)	152 (29.1)	163 (39.3)
N	15631	611	125	811	389	387	351	1404	103	523	415
Weight loss	5575 (35.7)	170 (27.8)	19 (15.2)	514 (63.6)	323 (82.2)	9 (2.3)	316 (90.3)	103 (7.3)	43 (41.7)	355 (68.0)	254 (61.2)
N	15602	611	125	808	393	387	350	1404	103	522	415
Night sweats	3270 (20.9)	144 (23.6)	30 (24.0)	396 (48.8)	131 (33.5)	18 (4.7)	147 (41.9)	49 (3.5)	26 (25.2)	210 (40.2)	190 (45.8)
N	15630	611	125	811	391	387	351	1404	103	523	415
Cough >= 2 weeks	2205 (20.2)	-	14 (11.2)	-	130 (33.0)	8 (2.1)	103 (29.3)	54 (3.8)	28 (27.2)	107 (20.5)	151 (38.1)
N	10919		125		394	387	351	1404	103	523	396
Lymphadenopathy	374 (15.6)	-	-	28 (3.5)	76 (19.2)	-	-	-	-	-	-
N	2394			810	395						
CXR (suggests tuberculosis)	1296 (21.0)	265 (44.8)	-	-	-	-	177 (50.9)	-	-	196 (41.3)	72 (18.8)
N	6177	591					348			475	383
CXR (any abnormality)	2158 (34.7)	354 (59.9)	-	-	-	-	222 (63.8)	-	-	239 (50.1)	145 (37.8)

Variable	All	Ahmad	Al Darraji	Balcha	Bjerrum	Gersh	Hanifa	Hoffmann	Kemper	Kerkhoff	Kufa
N	6222	591					348			477	384
Total Xpert positive**	616 (7.1)	-	8 (6.4)	96 (11.9)	21 (15.1)	4 (1.0)	-	-	12 (11.8)	70 (13.4)	-
N	8625		125	804	139	384			102	523	
Total culture positive***	1347 (8.6)	57 (9.3)	15 (12.0)	124 (15.3)	37 (9.4)	5 (1.3)	64 (18.2)	35 (2.5)	12 (11.8)	89 (17.0)	24 (5.8)
N	15611	611	125	809	395	387	351	1404	102	522	415
Total Xpert & culture positive	1453 (9.3)	57 (9.3)	15 (12.0)	137 (16.9)	40 (10.1)	8 (2.1)	64 (18.2)	35 (2.5)	15 (14.6)	95 (18.2)	24 (5.8)
N	15666	611	125	812	395	387	351	1404	103	523	415
BMI (kg/m²)	22 (19-26)	24 (21-28)	22 (21-25)	19 (17-21)	20 (18-23)	22 (20-25)	20 (18-24)	-	22 (21-24)	23 (21-27)	24 (21-28)
N	12704	575	125	801	390	383	349	-	101	522	403
CRP (mg/L)	4 (2-21)	-	-	-	-	1 (1-4)	-	-	-	10 (2-32)	-
N	3582	-	-	-	-	385	-	-	-	502	-
CRP (>=10 mg/L)	1259 (35.1)	-	-	-	-	39 (10.1)	-	-	-	251 (50.0)	-
N	3582					385				502	
Hb (g/dL)	12 (10-13)	-	-	12 (10-13)	10 (9-11)	13 (11-14)	11 (10-13)	11 (10-12)	-	12 (11-13)	12 (11-14)
N	5118	-	-	762	360	366	348	1287	-	490	415
Hb (<10 g/dL)	1093 (21.4)	-	-	165 (21.7)	184 (51.1)	53 (14.5)	91 (26.1)	204 (15.9)	-	86 (17.6)	64 (15.4)
N	5118			762	360	366	348	1287		490	415

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

*One study among a prison population

**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, IPT = Isoniazid preventive therapy, W4SS = WHO four-symptom screen

Variable	LaCourse	Mbu	Affolabi	Modi	Nguyen	Rangaka	Reeve	Shapiro	Swindells	Thit	Yoon
Participants	292 (1.9)	940 (6)	2805 (17.9)	731 (4.7)	397 (2.5)	1429 (9.1)	807 (5.2)	425 (2.7)	726 (4.6)	463 (3)	1525 (9.7)
Clinical setting											
Outpatient	292 (100)	940 (100)	2805 (100)	731 (100)	397 (100)	1429 (100)	807 (100)	425 (100)	726 (100)	463 (100)	1525 (100)
Other setting*	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
N	292	940	2805	731	397	1429	807	425	726	463	1525
Age (years)	25 (22-30)	35 (28-42)	40 (33-49)	30 (24-39)	30 (27-34)	34 (30-40)	32 (26-39)	32 (27-39)	33 (28-39)	34 (30-40)	33 (27-40)
N	292	940	2805	731	397	1429	807	425	726	463	1525
Female	292 (100.0)	641 (68.2)	1878 (67.0)	481 (65.8)	113 (28.5)	1053 (73.7)	478 (59.2)	248 (58.4)	458 (63.1)	231 (49.9)	806 (52.9)
N	292	940	2805	731	397	1429	807	425	726	463	1525
On ART	159 (54.5)	0 (0.0)	1808 (64.5)	0 (0.0)	230 (57.9)	775 (54.2)	0 (0.0)	0 (0.0)	0 (0.0)	338 (73.0)	0 (0.0)
N	292	940	2805	731	397	1429	807	425	726	463	1525
CD4 count (cells/μL)	437 (345-560)	291 (116-496)	376 (197-551)	342 (168-512)	336 (217-489)	209 (145-331)	306 (174-489)	306 (176-468)	273 (167-435)	286 (145-466)	160 (70-265)
N	242	940	2731	683	397	1429	726	411	723	462	1525
History of tuberculosis	25 (9.1)	-	-	-	178 (44.8)	552 (38.6)	115 (14.3)	19 (4.5)	53 (7.3)	94 (20.3)	54 (3.5)
N	274				397	1429	806	425	725	463	1524
Current Smoker	-	-	102 (3.6)	-	42 (10.7)	-	295 (36.6)	105 (24.8)	-	137 (29.7)	70 (5.0)
N			2805		393		807	424		462	1407
Pregnant	292 (100.0)	-	-	134 (29.1)	0 (0.0)	-	-	-	10 (1.4)	-	43 (5.3)
N	292			461	113				726		805
On IPT	0 (0.0)	-	7 (0.3)	-	-	-	-	-	0 (0.0)	12 (2.6)	0 (0.0)
N	292		2781						726	463	1525
W4SS	57 (19.5)	812 (86.4)	780 (27.8)	366 (50.9)	147 (37.0)	206 (14.4)	442 (54.8)	279 (65.6)	481 (66.3)	223 (48.2)	1335 (87.5)
N	292	940	2805	719	397	1429	807	425	725	463	1525
Cough	44 (15.1)	448 (47.7)	567 (20.2)	171 (24.6)	105 (26.4)	125 (8.7)	249 (30.9)	191 (44.9)	311 (43.0)	74 (16.0)	764 (50.1)
N	292	940	2805	694	397	1429	807	425	723	462	1525
Fever	14 (4.8)	220 (23.4)	444 (15.8)	236 (33.5)	26 (6.5)	9 (0.6)	55 (6.8)	129 (30.4)	222 (30.7)	61 (13.2)	776 (50.9)
N	292	940	2805	705	397	1429	807	425	724	463	1525
Weight loss	3 (1.0)	592 (63.0)	395 (14.1)	241 (34.0)	79 (19.9)	112 (7.8)	334 (41.4)	146 (34.4)	265 (38.2)	193 (41.7)	1109 (72.7)
N	292	940	2805	708	397	1429	806	425	694	463	1525
Night sweats	20 (6.8)	395 (42.0)	140 (5.0)	182 (26.0)	11 (2.8)	71 (5.0)	191 (23.7)	139 (32.7)	210 (29.0)	40 (8.6)	530 (34.8)
N	292	940	2805	701	397	1429	807	425	725	463	1525
Cough >= 2 weeks	14 (4.8)	228 (24.3)	-	138 (19.7)	105 (26.4)	82 (5.7)	166 (20.6)	129 (30.5)	229 (31.6)	-	519 (34.0)
N	292	940		699	397	1429	807	423	724		1525
Lymphadenopathy	-	-	-	-	-	-	-	-	238 (32.8)	32 (6.9)	-
N									726	463	
CXR (suggests tuberculosis)	-	-	251 (9.2)	99 (17.6)	127 (32.0)	-	-	-	109 (15.6)	-	-
N			2724	561	397				698		
CXR (any abnormality)	-	-	635 (23.3)	248 (43.8)	-	-	-	-	159 (22.8)	156 (36.1)	-
N			2726	566					698	432	
Total Xpert positive**	4 (1.4)	-	111 (4.0)	61 (8.4)	-	-	61 (8.0)	-	-	25 (5.4)	143 (9.4)

Variable	LaCourse	Mbu	Affolabi	Modi	Nguyen	Rangaka	Reeve	Shapiro	Swindells	Thit	Yoon
N	289		2781	728			762			463	1525
Total culture positive* **	7 (2.4)	131 (13.9)	85 (3.1)	75 (10.3)	28 (7.1)	126 (8.8)	93 (11.7)	42 (9.9)	56 (7.7)	10 (2.2)	232 (15.2)
N	292	940	2766	731	397	1429	796	425	726	463	1525
Total Xpert & culture positive	8 (2.7)	131 (13.9)	118 (4.2)	82 (11.2)	28 (7.1)	126 (8.8)	100 (12.4)	42 (9.9)	56 (7.7)	30 (6.5)	242 (15.9)
N	292	940	2805	731	397	1429	807	425	726	463	1525
BMI (kg/m²)	24 (22-26)	-	22 (19-25)	21 (19-22)	-	25 (22-29)	24 (21-29)	24 (21-29)	22 (20-26)	21 (19-23)	21 (19-24)
N	292	-	2769	634	-	1429	803	425	719	461	1523
CRP (mg/L)	-	-	-	-	-	-	6 (2-28)	4 (1-17)	-	-	4 (2-23)
N	-	-	-	-	-	-	745	425	-	-	1525
CRP (>=10 mg/L)	-	-	-	-	-	-	303 (40.7)	140 (32.9)	-	-	526 (34.5)
N							745	425			1525
Hb (g/dL)	-	-	-	-	-	-	-	-	12 (11-14)	11 (9-12)	-
N	-	-	-	-	-	-	-	-	724	366	-
Hb (<10 g/dL)	-	-	-	-	-	-	-	-	113 (15.6)	133 (36.3)	-
N									724	366	

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

*One study among a prison population

**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, IPT = Isoniazid preventive therapy, W4SS = WHO four-symptom screen

Table 8-6: Prevalence of tuberculosis in all participants and by different subgroups (using culture or Xpert as reference standard)

					Heterogeneity			
Subgroup§	**No studies**	**N**	**No tuberculosis**	**Prevalence % (95% CI)†**	**I² (95% CI)**	**P-value**	**Egger's test (p-value)**	**Subgroup analysis (p-value)††**
All	**21**	**15,666**	**1,453**	**8.7 (6.7-11.3)**	**95 (94-96)**	**<0.0001**	**0.06**	**-**
All (setting and ART status)	**21**	**15,663**	**1,453**	**8.7 (6.7-11.3)**	**95 (94-96)**	**<0.0001**	**0.06**	**-**
Outpatients (on ART)*	9	4,328	169	4.2 (3.1-5.6)	75 (52-87)	<0.0001	0.78	<0.0001
Outpatients (not on ART)	20	11,210	1,269	10.2 (8.0-12.9)	92 (89-94)	<0.0001	0.06	-
Other setting**	1	125	15	12.0 (7.4-19.0)	- (-)	-	-	-
All (CD4 count)	**21**	**15,281**	**1,423**	**8.7 (6.7-11.2)**	**95 (93-96)**	**<0.0001**	**0.06**	**-**
CD4 count <=200 cells/µL	21	5,641	922	15.1 (12.8-17.7)	82 (73-88)	<0.0001	0.05	<0.0001
CD4 count >200 cells/µL	21	9,640	501	5.5 (4.2-7.1)	88 (84-92)	<0.0001	0.42	-
All (pregnancy status)***	**21**	**10,388**	**746**	**7.0 (5.3-9.2)**	**91 (88-94)**	**<0.0001**	**0.22**	**-**
Pregnant	8	1,938	56	2.9 (2.2-3.7)	16 (0-59)	<0.0001	0.09	<0.0001
Non-pregnant	19	8,450	690	8.0 (6.2-10.3)	89 (85-92)	<0.0001	0.32	-

§Subgroup in bold is the overall comparator. For example, all (setting and ART status) contains combined subgroups outpatients (on ART), outpatients (not on ART), and other setting

†Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

††P-value of between subgroups heterogeneity statistic Q (based on random effects model)

*P(subgroup) compares outpatients (on ART) with outpatients (not on ART)

**One study among a prison population

***Pregnancy status unavailable for some studies, female participants in those studies categorized as non-pregnant

Definition of abbreviations: ART = antiretroviral therapy

Table 8-7: Indirect comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in subgroups

Table 8-7A - Indirect comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in outpatients (on ART)†

					Difference from W4SS††	
Test	**No studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Sensitivity (p-value)**	**Specificity (p-value)**
W4SS	9	4309	53 (35-71)	71 (51-85)	-	-
CRP (>=10 mg/L)	1	381	20 (3-69)	90 (93-87)	0.296	0.29
CRP (>=8 mg/L)	1	381	40 (10-80)	89 (92-86)	0.689	0.342
CRP (>=5 mg/L)	1	381	40 (10-80)	80 (84-75)	0.689	0.697
CXR (abnormal)	4	2670	73 (60-83)	63 (50-75)	0.142	0.59
CXR (suggests tuberculosis)	4	2581	70 (55-82)	78 (62-89)	0.188	0.571
Cough (any)	9	4309	40 (24-58)	83 (73-90)	0.286	0.206
Cough (>=2 weeks)	6	1746	19 (5-52)	93 (79-98)	0.054	0.031
Hb (<10 g/dL)	4	844	55 (34-75)	82 (70-91)	0.996	0.315
Hb (<8 g/dL)	4	844	18 (5-49)	94 (91-96)	0.057	0.013
BMI (<18.5 kg/m²)	7	4036	16 (8-30)	93 (88-96)	0.004	0.007
Lymphadenopathy	1	338	22 (6-58)	93 (95-89)	0.278	0.204
W4SS with CRP (>=10 mg/L)¶	1	381	20 (3-69)	80 (83-75)	0.296	0.706
W4SS with CXR (abnormal)¶	4	2670	89 (70-97)	33 (17-54)	0.014	0.036
W4SS then CRP (>=5 mg/L)¶	1	381	8 (1-62)	96 (97-93)	0.04	0.09
W4SS then Xpert*	4	2645	37 (25-52)	100 (96-100)	-	-
Xpert alone*	4	2645	53 (22-83)	99 (97-99)	0.219	0.306

†Using culture as a reference standard. Indirect comparisons are based on all studies that evaluated at least one of the W4SS or relevant screening tests

††For Xpert alone, the comparator is W4SS then Xpert

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result). Alternative algorithms are W4SS then single sputum Xpert (4 studies; 2645 participants; sensitivity 37 (25-52), specificity 100 (96-100) and single sputum Xpert alone (4 studies; 2645 participants; sensitivity 53 (22-83), specificity 99 (97-99).

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

Table 8-7B - Indirect comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in outpatients (not on ART)†

					Difference from W4SS††	
Test	**No studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Sensitivity (p-value)**	**Specificity (p-value)**
W4SS	20	11160	85 (76-91)	37 (25-51)	-	-
CRP (>=10 mg/L)	5	3187	83 (79-86)	67 (60-73)	0.953	0.085
CRP (>=8 mg/L)	5	3187	85 (81-88)	63 (56-70)	0.8	0.135
CRP (>=5 mg/L)	5	3187	89 (86-92)	53 (46-61)	0.318	0.375
CXR (abnormal)	8	3525	71 (64-78)	62 (51-72)	0.102	0.034
CXR (suggests tuberculosis)	8	3569	62 (55-68)	79 (67-87)	0.01	0.001
Cough (any)	20	11131	58 (50-65)	70 (63-76)	<0.0001	<0.0001
Cough (>=2 weeks)	16	9032	43 (34-52)	82 (75-87)	<0.0001	<0.0001
Hb (<10 g/dL)	9	4269	44 (32-56)	78 (70-85)	<0.0001	<0.0001
Hb (<8 g/dL)	9	4269	11 (8-15)	96 (93-98)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	17	8486	31 (23-41)	87 (82-91)	<0.0001	<0.0001
Lymphadenopathy	4	2053	30 (13-55)	91 (76-97)	0.001	<0.0001
W4SS with CRP (>=10 mg/L)¶	5	3187	94 (87-97)	20 (12-33)	0.113	0.155
W4SS with CXR (abnormal)¶	8	3516	95 (91-97)	18 (9-33)	0.007	0.064
W4SS then CRP (>=5 mg/L)¶	5	3187	84 (75-90)	64 (57-71)	0.793	0.056
W4SS then Xpert*§	11	5784	64 (57-71)	99 (98-99)	-	-
Xpert alone*§	11	5797	74 (64-82)	99 (98-99)	0.199	0.59

†Using culture as a reference standard. Indirect comparisons are based on all studies that evaluated at least one of the W4SS or relevant screening tests

††For Xpert alone, the comparator is W4SS then Xpert

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result). Alternative algorithms are W4SS then single sputum Xpert (11 studies; 5783 participants; sensitivity 61 (53-69), specificity 99 (98-99) and single sputum Xpert alone (11 studies; 5796 participants; sensitivity 70 (58-80), specificity 99 (98-99).

§One study assessed Xpert and Xpert Ultra among 733 participants. The accuracy of sputum Xpert was: sensitivity 57 (47-67), specificity 99 (98-100); sputum Xpert Ultra: sensitivity 73 (62-81), specificity 98 (96-98); urine Xpert Ultra: sensitivity 27 (19-38), specificity 98 (96-99); sputum and urine Xpert Ultra: sensitivity 75 (65-83), specificity 95 (94-97)

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

Table 8-7C - Indirect comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in participants with CD4 cell count <=200 cells/μL†

					Difference from W4SS††	
Test	**No studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Sensitivity (p-value)**	**Specificity (p-value)**
W4SS	21	5617	87 (77-93)	33 (20-48)	-	-
CRP (>=10 mg/L)	5	1593	88 (82-92)	62 (50-73)	0.923	0.07
CRP (>=8 mg/L)	5	1593	90 (86-93)	58 (45-70)	0.62	0.113
CRP (>=5 mg/L)	5	1593	92 (88-95)	48 (34-62)	0.385	0.319
CXR (abnormal)	8	2201	73 (65-80)	58 (48-67)	0.127	0.063
CXR (suggests tuberculosis)	8	2134	64 (57-71)	74 (64-82)	0.026	0.003
Cough (any)	21	5607	59 (51-66)	66 (58-73)	<0.0001	<0.0001
Cough (>=2 weeks)	17	4197	42 (33-51)	78 (70-85)	<0.0001	<0.0001
Hb (<10 g/dL)	9	1970	54 (41-65)	74 (64-81)	0.004	0.002
Hb (<8 g/dL)	9	1970	14 (9-19)	95 (91-97)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	18	4950	35 (27-44)	83 (77-88)	<0.0001	<0.0001
Lymphadenopathy	4	1001	33 (15-58)	89 (73-96)	0.004	0.002
W4SS with CRP (>=10 mg/L)¶	5	1593	93 (57-99)	23 (9-47)	0.211	0.548
W4SS with CXR (abnormal)¶	8	2199	96 (91-98)	14 (7-25)	0.022	0.06
W4SS then CRP (>=5 mg/L)¶	5	1593	80 (49-94)	66 (37-86)	0.579	0.066
W4SS then Xpert*§	12	3110	69 (61-76)	98 (97-99)	-	-
Xpert alone*§	12	3115	77 (68-84)	98 (97-99)	0.167	0.879

†Using culture as a reference standard. Indirect comparisons are based on all studies that evaluated at least one of the W4SS or relevant screening tests

††For Xpert alone, the comparator is W4SS then Xpert

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result). Alternative algorithms are W4SS then single sputum Xpert (12 studies; 3110 participants; sensitivity 65 (57-72), specificity 98 (97-99) and single sputum Xpert alone (12 studies; 3115 participants; sensitivity 72 (62-81), specificity 98 (97-99).

§One study assessed Xpert and Xpert Ultra among 191 participants. The accuracy of sputum Xpert was: sensitivity 69 (54-81), specificity 98 (94-99); sputum Xpert Ultra: sensitivity 83 (69-92), specificity 95 (90-97); urine Xpert Ultra: sensitivity 40 (27-56), specificity 95 (90-98); sputum and urine Xpert Ultra: sensitivity 88 (74-95), specificity 91 (86-95)

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

Table 8-7D - Indirect comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in participants with CD4 cell count >200 cells/μL†

					Difference from W4SS††	
Test	**No studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Sensitivity (p-value)**	**Specificity (p-value)**
W4SS	21	9598	71 (59-81)	49 (36-63)	-	-
CRP (>=10 mg/L)	5	1822	65 (52-76)	79 (68-87)	0.543	0.03
CRP (>=8 mg/L)	5	1822	67 (53-79)	76 (64-85)	0.693	0.058
CRP (>=5 mg/L)	5	1822	78 (65-87)	65 (54-75)	0.469	0.274
CXR (abnormal)	8	3892	66 (57-74)	64 (53-74)	0.777	0.23
CXR (suggests tuberculosis)	8	3915	59 (49-69)	81 (70-89)	0.341	0.006
Cough (any)	21	9581	51 (41-61)	76 (68-82)	0.019	0.002
Cough (>=2 weeks)	17	6409	34 (23-48)	87 (80-91)	<0.0001	<0.0001
Hb (<10 g/dL)	9	3047	25 (17-37)	84 (78-89)	<0.0001	0.001
Hb (<8 g/dL)	9	3047	6 (3-15)	96 (94-98)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	18	7362	20 (13-29)	92 (88-95)	<0.0001	<0.0001
Lymphadenopathy	4	1373	30 (10-61)	90 (73-97)	0.017	0.004
W4SS with CRP (>=10 mg/L)¶	5	1822	78 (53-92)	35 (15-61)	0.496	0.349
W4SS with CXR (abnormal)¶	8	3886	91 (81-96)	28 (16-44)	0.008	0.073
W4SS then CRP (>=5 mg/L)¶	5	1822	57 (15-91)	80 (62-91)	0.639	0.03
W4SS then Xpert*§	12	5128	46 (39-52)	99 (99-100)	-	-
Xpert alone*§	12	5135	57 (47-67)	99 (99-99)	0.052	0.218

†Using culture as a reference standard. Indirect comparisons are based on all studies that evaluated at least one of the W4SS or relevant screening tests

††For Xpert alone, the comparator is W4SS then Xpert

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result). Alternative algorithms are W4SS then single sputum Xpert (12 studies; 5127 participants; sensitivity 44 (38-51), specificity 100 (99-100) and single sputum Xpert alone (12 studies; 5134 participants; sensitivity 54 (44-64), specificity 99 (99-99).

§One study assessed Xpert and Xpert Ultra among 472 participants. The accuracy of sputum Xpert was: sensitivity 42 (28-58), specificity 99 (98-100); sputum Xpert Ultra: sensitivity 62 (47-76), specificity 98 (96-99); urine Xpert Ultra: sensitivity 8 (2-21), specificity 99 (97-99); sputum and urine Xpert Ultra: sensitivity 62 (47-76), specificity 97 (95-98)

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

Table 8-7E - Indirect comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in pregnant participants†§

					Difference from W4SS††	
Test	**No studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Sensitivity (p-value)**	**Specificity (p-value)**
W4SS	8	1935	84 (24-99)	58 (39-75)	-	-
CRP (>=10 mg/L)#	-	-	-	-	-	-
CRP (>=8 mg/L)#	-	-	-	-	-	-
CRP (>=5 mg/L)#	-	-	-	-	-	-
CXR (abnormal)	1	8	75 (11-99)	69 (91-33)	0.538	0.619
CXR (suggests tuberculosis)	1	7	75 (11-99)	93 (100-42)	0.456	0.041
Cough (any)	8	1933	67 (26-92)	81 (71-88)	0.445	0.038
Cough (>=2 weeks)	8	1933	47 (18-78)	92 (86-95)	0.203	0.001
Hb (<10 g/dL)	5	1350	20 (10-36)	75 (61-85)	0.132	0.227
Hb (<8 g/dL)	5	1350	0 (0-100)	98 (97-99)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	7	472	0 (0-100)	96 (94-98)	0.003	<0.0001
Lymphadenopathy#	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶#	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	1	8	75 (11-99)	56 (84-24)	0.541	0.987
W4SS then CRP (>=5 mg/L)¶#	-	-	-	-	-	-
W4SS then Xpert*	5	489	36 (16-62)	100 (0-100)	-	-
Xpert alone*	5	492	53 (29-76)	99 (98-100)	0.339	0.05

†Using culture as a reference standard. Indirect comparisons are based on all studies that evaluated at least one of the W4SS or relevant screening tests

§For some analyses, all studies had 0% or 100% sensitivity/specificity; therefore, models may have given unreliable estimates such as 95% CIs that range from 0 to 100

††For Xpert alone, the comparator is W4SS then Xpert

¶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

#Insufficient data to perform meta-analysis

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result). Alternative algorithms are W4SS then single sputum Xpert (5 studies; 489 participants; sensitivity 36 (16-62), specificity 100 (0-100) and single sputum Xpert alone (5 studies; 492 participants; sensitivity 47 (24-71), specificity 99 (98-100).

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

Table 8-8: Direct comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in all participants and subgroups

Table 8-8A - Direct comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in all participants†

	Index Test				W4SS				Difference from W4SS	
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (p-value)	Specificity (p-value)
CRP (>=10 mg/L)	5	3571	77 (50-92)	74 (52-88)	5	3571	82 (58-94)	39 (20-63)	0.7	0.052
CRP (>=8 mg/L)	5	3571	79 (54-93)	70 (48-86)	5	3571	82 (58-94)	39 (20-63)	0.833	0.079
CRP (>=5 mg/L)	5	3571	86 (67-95)	60 (37-79)	5	3571	82 (59-94)	39 (20-62)	0.706	0.231
CXR (abnormal)	8	6186	72 (62-80)	62 (46-75)	8	6186	86 (80-91)	32 (20-47)	0.011	0.017
CXR (suggests tuberculosis)	8	6141	64 (53-73)	78 (63-88)	8	6141	84 (76-89)	33 (19-51)	0.004	0.001
Cough (any)	21	15568	56 (44-66)	72 (62-81)	21	15568	82 (74-88)	42 (31-54)	<0.0001	0.001
Cough (>=2 weeks)	17	10906	38 (25-52)	85 (75-91)	17	10906	81 (70-88)	41 (28-56)	<0.0001	<0.0001
Hb (<10 g/dL)	9	5115	44 (25-65)	80 (65-89)	9	5115	85 (70-93)	33 (18-51)	0.006	0.001
Hb (<8 g/dL)	9	5115	12 (5-25)	96 (91-98)	9	5115	85 (69-93)	33 (18-52)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	18	12639	29 (20-40)	89 (82-93)	18	12639	84 (76-90)	41 (29-54)	<0.0001	<0.0001
Lymphadenopathy	4	2389	31 (16-52)	90 (77-96)	4	2389	91 (79-96)	27 (12-49)	0.001	0.001
W4SS with CRP (>=10 mg/L)¶	5	3571	89 (63-97)	31 (12-58)	5	3571	82 (50-95)	39 (17-67)	0.617	0.65
W4SS with CXR (abnormal)¶	8	6186	94 (89-97)	20 (10-37)	8	6186	87 (78-92)	32 (17-52)	0.05	0.32
W4SS then CRP (>=5 mg/L)¶	5	3571	72 (36-92)	75 (50-90)	5	3571	81 (49-95)	39 (18-66)	0.62	0.074

†Using culture as a reference standard. Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

¶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

Table 8-8B - Direct comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in outpatients (on ART)†

	Index Test				W4SS				Difference from W4SS	
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (p-value)	Specificity (p-value)
CRP (>=10 mg/L)	1	381	20 (3-69)	90 (87-93)	1	381	8 (1-62)	88 (84-91)	0.458	0.246
CRP (>=8 mg/L)	1	381	40 (10-80)	89 (86-92)	1	381	8 (1-62)	88 (84-91)	0.223	0.571
CRP (>=5 mg/L)	1	381	40 (10-80)	80 (75-84)	1	381	8 (1-62)	88 (84-91)	0.223	0.003
CXR (abnormal)	4	2670	74 (55-87)	63 (42-80)	4	2670	70 (50-85)	52 (31-72)	0.786	0.469
CXR (suggests tuberculosis)	4	2581	71 (52-84)	78 (57-91)	4	2581	64 (44-80)	54 (31-76)	0.587	0.144
Cough (any)	9	4309	40 (23-58)	83 (70-91)	9	4309	54 (36-72)	71 (54-84)	0.286	0.206
Cough (>=2 weeks)	6	1746	20 (6-50)	93 (79-98)	6	1746	52 (23-80)	74 (45-91)	0.136	0.09
Hb (<10 g/dL)	4	844	52 (15-87)	84 (60-94)	4	844	65 (21-93)	54 (26-79)	0.689	0.107
Hb (<8 g/dL)	4	844	15 (1-78)	95 (84-98)	4	844	67 (9-98)	54 (27-78)	0.285	0.008
BMI (<18.5 kg/m²)	7	4036	16 (8-32)	93 (85-97)	7	4036	53 (33-72)	73 (54-86)	0.009	0.017
Lymphadenopathy	1	338	22 (6-58)	93 (89-95)	1	338	78 (42-94)	55 (50-60)	0.027	<0.0001
W4SS with CRP (>=10 mg/L)¶	1	381	20 (3-69)	80 (75-83)	1	381	8 (1-62)	88 (84-91)	0.458	0.003
W4SS with CXR (abnormal)¶	4	2670	89 (71-96)	33 (15-57)	4	2670	73 (47-89)	52 (28-75)	0.187	0.291
W4SS then CRP (>=5 mg/L)¶	1	381	8 (1-62)	96 (93-97)	1	381	8 (1-62)	88 (84-91)	1	<0.0001

†Using culture as a reference standard. Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

¶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

Table 8-8C - Direct comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in outpatients (not on ART)†

	Index Test				W4SS				Difference from W4SS	
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (p-value)	Specificity (p-value)
CRP (>=10 mg/L)	5	3187	84 (74-90)	66 (53-78)	5	3187	90 (83-94)	27 (17-40)	0.236	0.005
CRP (>=8 mg/L)	5	3187	85 (75-92)	63 (49-75)	5	3187	90 (82-95)	27 (17-40)	0.349	0.008
CRP (>=5 mg/L)	5	3187	90 (83-95)	53 (39-66)	5	3187	90 (83-94)	27 (18-40)	0.952	0.031
CXR (abnormal)	8	3516	71 (62-79)	62 (48-74)	8	3516	89 (83-93)	27 (17-40)	0.001	0.002
CXR (suggests tuberculosis)	8	3560	62 (52-71)	79 (65-88)	8	3560	87 (81-92)	28 (17-44)	<0.0001	<0.0001
Cough (any)	20	11131	58 (47-68)	70 (60-78)	20	11131	85 (78-90)	37 (27-48)	<0.0001	<0.0001
Cough (>=2 weeks)	16	9032	42 (29-55)	82 (72-89)	16	9032	83 (74-90)	37 (25-50)	<0.0001	<0.0001
Hb (<10 g/dL)	9	4268	44 (27-63)	79 (65-88)	9	4268	89 (78-95)	26 (15-41)	0.001	<0.0001
Hb (<8 g/dL)	9	4268	11 (5-21)	96 (92-98)	9	4268	88 (78-94)	26 (15-42)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	17	8475	31 (22-42)	87 (81-92)	17	8475	87 (81-92)	34 (24-46)	<0.0001	<0.0001
Lymphadenopathy	4	2051	30 (15-51)	91 (78-96)	4	2051	91 (80-96)	25 (11-48)	0.001	0.001
W4SS with CRP (>=10 mg/L)¶	5	3187	94 (87-97)	20 (11-34)	5	3187	90 (79-95)	28 (16-43)	0.351	0.407
W4SS with CXR (abnormal)¶	8	3516	95 (91-97)	18 (9-32)	8	3516	89 (82-93)	27 (15-44)	0.045	0.364
W4SS then CRP (>=5 mg/L)¶	5	3187	83 (72-91)	65 (51-76)	5	3187	90 (82-95)	27 (18-40)	0.254	0.004

†Using culture as a reference standard. Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

¶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

Table 8-8D - Direct comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in participants with CD4 cell count <=200 cells/µL†

	Index Test				W4SS				Difference from W4SS	
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (p-value)	Specificity (p-value)
CRP (>=10 mg/L)	5	1593	87 (59-97)	64 (39-83)	5	1593	86 (58-97)	32 (15-56)	0.992	0.094
CRP (>=8 mg/L)	5	1593	89 (69-97)	60 (35-80)	5	1593	88 (64-96)	32 (15-57)	0.851	0.145
CRP (>=5 mg/L)	5	1593	92 (74-98)	49 (26-73)	5	1593	87 (63-97)	32 (14-57)	0.655	0.351
CXR (abnormal)	8	2199	74 (64-81)	58 (44-70)	8	2199	90 (84-94)	22 (14-32)	0.002	0.001
CXR (suggests tuberculosis)	8	2132	65 (55-73)	74 (60-84)	8	2132	88 (82-92)	24 (14-37)	<0.0001	<0.0001
Cough (any)	21	5607	57 (45-69)	67 (55-77)	21	5607	87 (79-92)	32 (22-44)	<0.0001	<0.0001
Cough (>=2 weeks)	17	4197	39 (25-55)	80 (67-88)	17	4197	85 (75-92)	31 (19-47)	<0.0001	<0.0001
Hb (<10 g/dL)	9	1969	54 (33-73)	74 (56-86)	9	1969	88 (75-95)	26 (13-43)	0.013	0.002
Hb (<8 g/dL)	9	1969	15 (7-30)	95 (89-98)	9	1969	89 (76-95)	25 (13-44)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	18	4947	34 (23-47)	84 (75-90)	18	4947	88 (80-93)	31 (21-44)	<0.0001	<0.0001
Lymphadenopathy	4	1000	33 (16-56)	89 (75-95)	4	1000	94 (84-98)	18 (8-37)	0.001	0.001
W4SS with CRP (>=10 mg/L)¶	5	1593	94 (62-99)	23 (8-50)	5	1593	83 (36-98)	32 (12-61)	0.489	0.614
W4SS with CXR (abnormal)¶	8	2199	96 (92-98)	14 (7-24)	8	2199	90 (83-94)	22 (12-36)	0.042	0.267
W4SS then CRP (>=5 mg/L)¶	5	1593	77 (35-95)	66 (36-87)	5	1593	86 (50-98)	32 (12-63)	0.631	0.144

†Using culture as a reference standard. Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

¶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

Table 8-8E - Direct comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in participants with CD4 cell count >200 cells/µL†

	Index Test				W4SS				Difference from W4SS	
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (p-value)	Specificity (p-value)
CRP (>=10 mg/L)	5	1822	62 (35-83)	79 (60-91)	5	1822	72 (45-89)	43 (23-66)	0.581	0.035
CRP (>=8 mg/L)	5	1822	65 (37-86)	76 (55-89)	5	1822	72 (44-90)	43 (23-66)	0.694	0.055
CRP (>=5 mg/L)	5	1822	78 (53-92)	65 (43-82)	5	1822	70 (43-88)	43 (23-65)	0.626	0.182
CXR (abnormal)	8	3886	67 (54-78)	64 (49-77)	8	3886	79 (68-87)	41 (27-57)	0.111	0.048
CXR (suggests tuberculosis)	8	3909	60 (47-72)	81 (68-90)	8	3909	77 (65-86)	41 (25-59)	0.058	0.002
Cough (any)	21	9581	52 (40-63)	75 (66-83)	21	9581	71 (60-80)	50 (38-61)	0.018	0.002
Cough (>=2 weeks)	17	6409	34 (22-49)	87 (78-92)	17	6409	70 (56-81)	48 (34-62)	0.001	<0.0001
Hb (<10 g/dL)	9	3047	26 (12-48)	84 (72-91)	9	3047	79 (58-91)	41 (25-58)	0.004	0.001
Hb (<8 g/dL)	9	3047	6 (2-18)	97 (93-98)	9	3047	79 (56-92)	41 (24-59)	<0.0001	<0.0001
BMI (<18.5 kg/m²)	18	7356	20 (13-30)	92 (87-95)	18	7356	74 (64-83)	49 (36-62)	<0.0001	<0.0001
Lymphadenopathy	4	1373	30 (12-56)	90 (77-96)	4	1373	85 (65-95)	34 (16-58)	0.008	0.003
W4SS with CRP (>=10 mg/L)¶	5	1822	77 (48-93)	35 (15-62)	5	1822	71 (39-90)	43 (20-70)	0.707	0.665
W4SS with CXR (abnormal)¶	8	3886	91 (81-96)	28 (16-45)	8	3886	82 (67-91)	41 (25-59)	0.159	0.293
W4SS then CRP (>=5 mg/L)¶	5	1822	61 (21-90)	80 (58-92)	5	1822	71 (29-93)	43 (21-69)	0.731	0.049

†Using culture as a reference standard. Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

¶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

Table 8-8F - Direct comparisons between each test and WHO four-symptom screen for the detection of tuberculosis in pregnant participants†§

	Index Test				W4SS				Difference from W4SS	
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (p-value)	Specificity (p-value)
CRP (>=10 mg/L)#	-	-	-	-	-	-	-	-	-	-
CRP (>=8 mg/L)#	-	-	-	-	-	-	-	-	-	-
CRP (>=5 mg/L)#	-	-	-	-	-	-	-	-	-	-
CXR (abnormal)	1	8	75 (11-99)	69 (33-91)	1	8	75 (11-99)	81 (42-96)	1	0.567
CXR (suggests tuberculosis)	1	7	75 (11-99)	93 (42-100)	1	7	75 (11-99)	79 (38-96)	1	0.465
Cough (any)	8	1933	69 (28-93)	80 (68-89)	8	1933	83 (39-97)	59 (42-74)	0.448	0.04
Cough (>=2 weeks)	8	1933	53 (20-83)	92 (84-96)	8	1933	78 (36-95)	59 (43-74)	0.215	0.001
Hb (<10 g/dL)	5	1350	53 (6-95)	74 (58-86)	5	1350	92 (7-100)	59 (41-75)	0.125	0.184
Hb (<8 g/dL)	5	1350	0 (0-100)	98 (94-99)	5	1350	95 (1-100)	59 (40-75)	0.001	<0.0001
BMI (<18.5 kg/m²)	7	471	0 (0-98)	97 (93-99)	7	471	100 (0-100)	54 (38-69)	<0.0001	<0.0001
Lymphadenopathy#	-	-	-	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶#	-	-	-	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	1	8	75 (11-99)	56 (24-84)	1	8	75 (11-99)	81 (42-96)	1	0.292
W4SS then CRP (>=5 mg/L)¶#	-	-	-	-	-	-	-	-	-	-

†Using culture as a reference standard. Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

§For some analyses, all studies had 0% or 100% sensitivity/specificity; therefore, models may have given unreliable estimates such as 95% CIs that range from 0 to 100

¶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

#Insufficient data to perform meta-analysis

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

Table 8-9: Additional diagnostic accuracy estimates

Table 8-9A - Additional diagnostic accuracy estimates in all participants

									Metaregression†	
	Univariate diagnostic odds ratio				Positive screen (%)††		Trivariate GLMM†††		Reference standard	Prevalence
Test	Estimate (95% CI)	I² (95% CI)	Trim-and-fill estimate (95% CI)	Egger's test (p-value)	Estimate (95% CI)	I² (95% CI)	PPV (95% CI)	NPV (95% CI)	P-value	P-value
W4SS	3.21 (2.61-3.95)	27 (0-57)	3.39 (2.74-4.20)	0.16	60 (45-73)	100 (99-100)	11.6 (9.3-14.3)	96.2 (94.9-97.2)	0.63	0.19
CRP (>=10 mg/L)	9.06 (5.55-14.78)	63 (3-86)	13.83 (8.18-23.36)	0.21	32 (20-47)	97 (95-98)	20.7 (10.4-37.1)	96.8 (94.7-98.1)	-	0.23
CRP (>=8 mg/L)	8.93 (5.11-15.63)	70 (25-88)	14.15 (7.88-25.38)	0.38	35 (22-50)	97 (96-98)	20.5 (11.5-33.9)	97.1 (94.9-98.3)	-	0.35
CRP (>=5 mg/L)	8.79 (5.06-15.26)	60 (0-85)	12.33 (6.80-22.33)	0.3	45 (32-59)	98 (96-98)	16.7 (8.4-30.6)	97.5 (95.6-98.6)	-	0.28
CXR (abnormal)	4.07 (2.97-5.58)	45 (0-76)	4.07 (2.97-5.58)	0.84	41 (32-52)	99 (98-99)	14.0 (9.5-20.2)	96.1 (93.4-97.7)	0.08	0.27
CXR (suggests tuberculosis)	6.37 (3.89-10.43)	81 (63-90)	6.37 (3.92-10.35)	0.95	26 (17-38)	99 (99-99)	22.6 (17.6-28.5)	95.4 (93.1-96.9)	0.12	0.11
Cough (any)	3.23 (2.76-3.77)	28 (0-58)	3.23 (2.76-3.77)	0.98	30 (23-38)	99 (98-99)	15.1 (12.4-18.2)	95.3 (93.5-96.5)	0.84	<0.0001
Cough (>=2 weeks)	3.32 (2.62-4.20)	55 (22-74)	2.95 (2.31-3.76)	0.41	18 (12-25)	98 (97-98)	19.2 (15.9-22.9)	94.1 (91.7-95.8)	0.4	<0.0001
Hb (<10 g/dL)	3.03 (2.31-3.97)	22 (0-63)	2.96 (2.25-3.89)	0.96	22 (16-30)	97 (96-98)	12.5 (6.6-22.4)	95.2 (91.3-97.4)	0.4	0.93
Hb (<8 g/dL)	3.25 (1.73-6.09)	60 (18-81)	4.33 (2.22-8.44)	0.54	5 (3-8)	94 (90-96)	15.5 (7.2-30.3)	93.9 (89.1-96.7)	0.12	0.63
BMI (<18.5 kg/m²)	3.11 (2.65-3.64)	9 (0-45)	3.45 (2.83-4.20)	0.72	13 (9-18)	98 (97-98)	18.3 (13.4-24.5)	93.6 (91.1-95.5)	0.65	0.26
Lymphadenopathy	3.96 (2.72-5.77)	0 (0-83)	4.81 (3.24-7.15)	0.93	12 (5-26)	99 (98-99)	18.3 (8.1-36.3)	94.5 (89.0-97.3)	0.21	0.32
W4SS with CRP (>=10 mg/L)	3.61 (2.41-5.40)	0 (0-71)	3.61 (2.41-5.40)	0.9	71 (45-88)	99 (99-100)	12.6 (8.1-19.0)	96.5 (93.7-98.1)	-	0.28
W4SS with CXR (abnormal)	5.22 (3.61-7.54)	0 (0-62)	5.44 (3.79-7.83)	0.38	81 (65-91)	99 (99-99)	10.0 (6.9-14.4)	97.8 (94.6-99.2)	0.91	0.04
W4SS then CRP (>=5 mg/L)	8.49 (4.69-15.37)	72 (30-89)	8.95 (5.04-15.90)	0.66	30 (13-54)	98 (96-99)	21.9 (13.9-33.0)	96.7 (94.6-98.0)	-	0.61
W4SS then Xpert*	142.47 (95.22-213.16)	32 (0-65)	142.49 (95.27-213.13)	0.37	5 (3-9)	93 (90-95)	87.9 (75.4-94.5)	96.6 (94.6-97.9)	0.66	0.54
Xpert alone*	165.89 (95.97-286.77)	62 (29-80)	273.62 (157.19-476.31)	0.56	7 (4-10)	92 (88-95)	81.1 (66.8-90.1)	97.5 (95.5-98.6)	0.65	0.78

†For the meta-regressions, the outcome variable was the diagnostic odds ratio

††Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

†††Pooled using a trivariate generalized linear mixed model that jointly models diagnostic test prevalence and predictive values

*Estimates for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, GLMM = generalized linear mixed model, Hb = haemoglobin, NPV = negative predictive value, PPV = positive predictive value, W4SS = WHO four-symptom screen

Table 8-9B - Additional diagnostic accuracy estimates in outpatients (On ART)

									Metaregression†	
	Univariate diagnostic odds ratio				Positive screen (%)††		Trivariate GLMM†††		Reference standard	Prevalence
Test	Estimate (95% CI)	I² (95% CI)	Trim-and-fill estimate (95% CI)	Egger's test (p-value)	Estimate (95% CI)	I² (95% CI)	PPV (95% CI)	NPV (95% CI)	P-value	P-value
W4SS	3.07 (2.09-4.52)	0 (0-63)	3.80 (2.50-5.79)	0.38	30 (16-50)	99 (98-99)	6.0 (3.9-9.1)	97.9 (96.7-98.7)	0.93	0.41
CRP (>=10 mg/L)	2.36 (0.26-21.70)	-	-	-	10 (7-13)	-	2.7 (0.1-13.8)	98.8 (97.0-99.5)	-	-
CRP (>=8 mg/L)	5.45 (0.88-33.56)	-	-	-	11 (8-15)	-	4.7 (1.3-15.5)	99.1 (97.4-99.7)	-	-
CRP (>=5 mg/L)	2.63 (0.43-16.03)	-	-	-	20 (17-25)	-	2.6 (0.7-8.9)	99.0 (97.1-99.7)	-	-
CXR (abnormal)	4.31 (2.21-8.39)	22 (0-88)	4.31 (2.22-8.35)	0.64	38 (26-52)	98 (97-99)	5.9 (4.3-8.0)	98.7 (97.4-99.3)	0.15	0.44
CXR (suggests tuberculosis)	8.84 (4.39-17.80)	37 (0-78)	8.84 (4.44-17.58)	0.8	24 (12-41)	99 (98-99)	11.3 (8.7-14.5)	98.5 (97.3-99.2)	0.16	0.12
Cough (any)	3.16 (2.11-4.74)	0 (0-65)	3.16 (2.11-4.74)	0.74	18 (11-28)	97 (96-98)	7.1 (4.9-10.1)	97.8 (96.4-98.6)	0.21	0.21
Cough (>=2 weeks)	3.23 (1.53-6.82)	0 (0-72)	2.27 (1.16-4.46)	0.12	7 (2-22)	98 (96-98)	9.6 (5.9-15.2)	97.4 (95.0-98.6)	0.19	0.2
Hb (<10 g/dL)	5.42 (2.19-13.45)	0 (0-87)	5.42 (2.19-13.45)	0.76	19 (10-31)	94 (88-97)	7.2 (3.4-14.7)	98.7 (97.3-99.3)	0.59	0.81
Hb (<8 g/dL)	5.23 (1.62-16.91)	0 (0-83)	8.22 (3.08-21.89)	0.15	6 (4-9)	54 (0-85)	6.7 (1.7-22.7)	98.0 (96.3-98.9)	0.96	0.46
BMI (<18.5 kg/m²)	2.61 (1.56-4.37)	0 (0-0)	2.61 (1.56-4.37)	0.44	7 (4-13)	95 (91-97)	6.6 (3.5-12.1)	97.4 (96.1-98.2)	0.35	0.65
Lymphadenopathy	3.63 (0.71-18.45)	-	-	-	8 (5-11)	-	7.7 (2.1-24.1)	97.8 (95.4-98.9)	-	-
W4SS with CRP (>=10 mg/L)	0.97 (0.11-8.81)	-	-	-	20 (17-25)	-	1.3 (0.1-6.9)	98.7 (96.7-99.5)	-	-
W4SS with CXR (abnormal)	3.69 (1.96-6.93)	0 (0-61)	3.69 (1.96-6.93)	0.96	68 (46-84)	99 (99-99)	4.7 (3.4-6.4)	98.8 (96.9-99.5)	0.79	0.41
W4SS then CRP (>=5 mg/L)	2.12 (0.11-40.08)	-	-	-	4 (2-6)	-	0.0 (0.0-20.4)	98.6 (96.8-99.4)	-	-
W4SS then Xpert*	82.01 (14.71-457.12)	65 (0-88)	230.45 (35.98-1476.17)	0.82	1 (0-4)	88 (71-95)	78.3 (22.7-97.8)	98.9 (98.4-99.2)	-	0.34
Xpert alone*	88.54 (10.76-728.44)	81 (51-93)	442.60 (45.58-4298.07)	0.46	2 (1-4)	82 (54-93)	42.7 (20.7-68.2)	99.2 (98.2-99.7)	-	0.85

†For the meta-regressions, the outcome variable was the diagnostic odds ratio

††Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

†††Pooled using a trivariate generalized linear mixed model that jointly models diagnostic test prevalence and predictive values

*Estimates for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, GLMM = generalized linear mixed model, Hb = haemoglobin, NPV = negative predictive value, PPV = positive predictive value, W4SS = WHO four-symptom screen

Table 8-9C - Additional diagnostic accuracy estimates in outpatients (Not on ART)

									Metaregression†	
	Univariate diagnostic odds ratio				Positive screen (%)††		Trivariate GLMM†††		Reference standard	Prevalence
Test	Estimate (95% CI)	I² (95% CI)	Trim-and-fill estimate (95% CI)	Egger's test (p-value)	Estimate (95% CI)	I² (95% CI)	PPV (95% CI)	NPV (95% CI)	P-value	P-value
W4SS	3.14 (2.61-3.78)	0 (0-46)	3.11 (2.59-3.75)	0.72	65 (52-76)	99 (99-99)	12.8 (10.5-15.6)	95.8 (94.4-96.8)	0.63	0.78
CRP (>=10 mg/L)	9.66 (5.97-15.65)	67 (3-89)	13.79 (8.17-23.28)	0.35	40 (33-46)	91 (82-96)	28.4 (22.9-34.6)	95.9 (94.6-96.9)	-	0.66
CRP (>=8 mg/L)	9.30 (5.10-16.95)	77 (37-92)	14.05 (7.45-26.52)	0.39	43 (36-50)	92 (84-96)	26.6 (21.1-32.9)	96.1 (94.8-97.1)	-	0.47
CRP (>=5 mg/L)	9.72 (5.65-16.70)	62 (0-87)	14.97 (8.02-27.93)	0.61	52 (46-59)	92 (86-96)	23.3 (18.8-28.4)	96.8 (95.3-97.8)	-	0.95
CXR (abnormal)	4.08 (3.02-5.51)	28 (0-68)	4.08 (3.02-5.50)	0.93	42 (31-52)	98 (97-98)	19.0 (15.3-23.2)	95.0 (91.8-96.9)	0.44	0.01
CXR (suggests tuberculosis)	5.98 (3.42-10.46)	81 (64-90)	5.49 (3.23-9.32)	0.99	26 (17-38)	98 (97-99)	28.3 (24.6-32.4)	93.7 (91.2-95.6)	0.29	<0.0001
Cough (any)	3.12 (2.74-3.56)	0 (0-41)	2.95 (2.57-3.39)	0.29	33 (27-40)	98 (97-98)	16.9 (14.0-20.1)	94.4 (92.5-95.9)	0.46	0.06
Cough (>=2 weeks)	3.26 (2.61-4.06)	48 (5-71)	2.92 (2.34-3.65)	0.17	21 (15-28)	97 (96-98)	20.6 (17.6-23.9)	93.3 (90.9-95.1)	0.72	0.01
Hb (<10 g/dL)	2.88 (2.19-3.80)	24 (0-66)	2.88 (2.19-3.80)	0.41	24 (17-32)	96 (95-98)	13.3 (6.2-26.5)	94.2 (89.5-96.9)	0.55	0.44
Hb (<8 g/dL)	3.11 (1.48-6.52)	69 (31-86)	3.11 (1.50-6.43)	0.74	5 (3-8)	93 (90-96)	16.6 (6.3-37.0)	92.6 (86.9-95.9)	0.08	0.91
BMI (<18.5 kg/m²)	3.05 (2.54-3.66)	17 (0-53)	2.86 (2.37-3.46)	0.7	15 (10-21)	97 (97-98)	21.8 (16.7-28.0)	92.1 (89.5-94.1)	0.59	0.34
Lymphadenopathy	4.01 (2.73-5.89)	0 (0-83)	3.43 (2.37-4.98)	0.85	11 (4-26)	98 (97-99)	19.5 (7.1-43.6)	94.8 (87.7-97.9)	0.22	0.35
W4SS with CRP (>=10 mg/L)	3.78 (2.51-5.70)	0 (0-67)	3.44 (2.36-5.00)	0.17	82 (70-90)	98 (97-99)	16.3 (14.9-17.7)	95.5 (93.3-97.0)	-	0.79
W4SS with CXR (abnormal)	4.85 (3.07-7.67)	0 (0-66)	5.73 (3.69-8.90)	0.22	83 (69-92)	98 (97-98)	12.5 (9.1-16.8)	97.2 (92.8-99.0)	0.82	0.1
W4SS then CRP (>=5 mg/L)	8.94 (4.86-16.45)	78 (39-92)	8.94 (4.88-16.38)	0.9	42 (35-50)	93 (86-96)	27.0 (23.3-31.0)	96.0 (94.3-97.2)	-	0.88
W4SS then Xpert*	154.61 (111.63-214.14)	0 (0-44)	154.61 (111.63-214.14)	0.21	8 (6-10)	82 (68-89)	83.7 (71.7-91.2)	96.6 (94.2-98.0)	0.32	0.85
Xpert alone*	184.61 (120.87-281.97)	30 (0-66)	184.60 (120.90-281.86)	0.99	9 (7-11)	74 (53-86)	84.0 (74.7-90.3)	97.7 (95.1-99.0)	0.29	0.36

†For the meta-regressions, the outcome variable was the diagnostic odds ratio

††Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

†††Pooled using a trivariate generalized linear mixed model that jointly models diagnostic test prevalence and predictive values

*Estimates for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, GLMM = generalized linear mixed model, Hb = haemoglobin, NPV = negative predictive value, PPV = positive predictive value, W4SS = WHO four-symptom screen

Table 8-9D - Additional diagnostic accuracy estimates in participants with a CD4 cell count <=200 cells/µL

									Metaregression†	
	Univariate diagnostic odds ratio				Positive screen (%)††		Trivariate GLMM†††		Reference standard	Prevalence
Test	Estimate (95% CI)	I² (95% CI)	Trim-and-fill estimate (95% CI)	Egger's test (p-value)	Estimate (95% CI)	I² (95% CI)	PPV (95% CI)	NPV (95% CI)	P-value	P-value
W4SS	2.98 (2.33-3.81)	0 (0-42)	3.28 (2.54-4.24)	0.29	70 (55-82)	98 (98-99)	17.5 (14.7-20.7)	94.0 (92.0-95.5)	0.29	0.46
CRP (>=10 mg/L)	12.06 (7.99-18.21)	9 (0-81)	14.72 (9.32-23.24)	0.02	46 (32-60)	91 (82-96)	37.5 (31.7-43.7)	95.4 (93.7-96.6)	-	0.1
CRP (>=8 mg/L)	13.34 (8.96-19.87)	1 (0-80)	15.50 (10.22-23.51)	0.04	50 (35-64)	92 (84-96)	36.0 (29.8-42.8)	96.0 (94.0-97.3)	-	0.59
CRP (>=5 mg/L)	9.69 (5.15-18.26)	35 (0-76)	12.41 (6.74-22.87)	0.26	58 (43-72)	94 (88-97)	31.7 (26.4-37.5)	96.1 (93.6-97.6)	-	0.55
CXR (abnormal)	3.78 (2.83-5.05)	0 (0-66)	4.12 (3.08-5.50)	0.44	46 (37-56)	96 (94-98)	20.4 (15.7-26.2)	93.7 (90.6-95.9)	0.23	0.4
CXR (suggests tuberculosis)	4.95 (3.15-7.78)	61 (16-82)	5.52 (3.53-8.64)	0.7	32 (23-42)	96 (95-98)	29.7 (25.8-33.9)	92.3 (89.7-94.3)	0.24	0.1
Cough (any)	2.63 (2.24-3.09)	0 (0-40)	2.65 (2.26-3.11)	0.98	37 (30-45)	96 (95-97)	21.8 (18.6-25.3)	91.4 (88.9-93.5)	0.77	0.01
Cough (>=2 weeks)	2.46 (1.93-3.13)	34 (0-64)	2.29 (1.78-2.96)	0.41	25 (18-33)	94 (92-96)	27.1 (23.1-31.4)	89.0 (85.9-91.4)	0.1	<0.0001
Hb (<10 g/dL)	3.25 (2.21-4.79)	37 (0-71)	3.13 (2.09-4.69)	0.71	30 (22-39)	94 (90-96)	20.4 (13.1-30.3)	92.3 (88.1-95.0)	0.97	0.64
Hb (<8 g/dL)	3.49 (1.55-7.87)	64 (27-82)	4.65 (1.93-11.16)	0.7	7 (5-10)	83 (70-91)	26.6 (14.0-44.7)	89.9 (85.1-93.3)	0.3	0.9
BMI (<18.5 kg/m²)	2.57 (2.15-3.09)	0 (0-43)	2.41 (2.03-2.87)	0.61	19 (14-26)	96 (95-97)	24.6 (19.1-31.0)	89.1 (86.3-91.4)	0.99	0.91
Lymphadenopathy#	3.97 (2.36-6.68)	0 (0-80)	3.81 (2.31-6.28)	0.59	14 (6-30)	97 (95-98)	-	-	0.32	0.52
W4SS with CRP (>=10 mg/L)	5.14 (2.11-12.54)	14 (0-82)	5.15 (2.22-11.96)	0.63	80 (56-92)	96 (93-98)	21.3 (17.0-26.4)	95.8 (91.6-98.0)	-	0.15
W4SS with CXR (abnormal)	3.90 (2.25-6.79)	0 (0-36)	5.11 (3.14-8.31)	0.33	88 (77-94)	97 (96-98)	14.4 (11.0-18.7)	96.7 (90.5-98.9)	0.82	0.29
W4SS then CRP (>=5 mg/L)	7.50 (3.88-14.48)	57 (0-84)	12.80 (6.37-25.74)	0.39	41 (17-70)	85 (67-93)	32.5 (27.0-38.6)	94.7 (91.5-96.8)	-	0.88
W4SS then Xpert*	127.84 (86.56-188.80)	0 (0-52)	142.03 (90.33-223.30)	0.11	12 (9-14)	68 (42-82)	85.4 (73.1-92.6)	95.4 (92.6-97.1)	0.47	0.29
Xpert alone*	167.10 (100.59-277.58)	25 (0-63)	218.91 (123.79-387.11)	0.41	13 (10-16)	63 (31-80)	86.0 (74.3-92.9)	96.6 (94.0-98.1)	0.73	0.63

†For the meta-regressions, the outcome variable was the diagnostic odds ratio

††Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

†††Pooled using a trivariate generalized linear mixed model that jointly models diagnostic test prevalence and predictive values

#Trivariate random-effects model did not converge

*Estimates for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, GLMM = generalized linear mixed model, Hb = haemoglobin, NPV = negative predictive value, PPV = positive predictive value, W4SS = WHO four-symptom screen

Table 8-9E - Additional diagnostic accuracy estimates in participants with a CD4 cell count >200 cells/µL

									Metaregression†	
	Univariate diagnostic odds ratio				Positive screen (%)††		Trivariate GLMM†††		Reference standard	Prevalence
Test	Estimate (95% CI)	I² (95% CI)	Trim-and-fill estimate (95% CI)	Egger's test (p-value)	Estimate (95% CI)	I² (95% CI)	PPV (95% CI)	NPV (95% CI)	P-value	P-value
W4SS	2.33 (1.73-3.12)	26 (0-57)	2.26 (1.69-3.02)	0.8	52 (38-65)	99 (99-99)	7.3 (5.8-9.1)	96.8 (95.5-97.8)	0.94	0.06
CRP (>=10 mg/L)	6.23 (3.64-10.65)	36 (0-76)	6.23 (3.64-10.65)	0.97	23 (14-37)	95 (92-97)	16.8 (11.2-24.3)	97.2 (94.7-98.5)	-	0.81
CRP (>=8 mg/L)	5.81 (3.27-10.31)	41 (0-78)	5.81 (3.27-10.31)	0.99	27 (16-40)	96 (92-98)	15.1 (9.7-22.8)	97.2 (94.9-98.5)	-	0.93
CRP (>=5 mg/L)	6.56 (4.14-10.42)	0 (0-71)	6.56 (4.14-10.42)	0.8	37 (26-50)	95 (92-97)	12.8 (8.0-20.0)	97.8 (96.2-98.7)	-	0.59
CXR (abnormal)	3.60 (2.33-5.56)	23 (0-65)	3.60 (2.33-5.56)	0.91	37 (28-48)	98 (96-98)	8.3 (5.0-13.5)	97.5 (95.0-98.7)	0.25	0.11
CXR (suggests tuberculosis)	6.51 (3.79-11.16)	52 (0-78)	6.50 (3.82-11.06)	0.82	22 (13-33)	98 (98-99)	15.3 (10.6-21.7)	97.1 (95.2-98.3)	0.13	0.12
Cough (any)	3.16 (2.51-3.99)	13 (0-47)	3.16 (2.51-3.99)	0.94	26 (20-34)	98 (97-98)	10.1 (8.1-12.4)	96.9 (95.7-97.8)	0.52	0.02
Cough (>=2 weeks)	3.45 (2.48-4.81)	29 (0-60)	3.45 (2.48-4.81)	0.98	15 (10-22)	96 (94-97)	13.0 (10.2-16.4)	96.1 (94.6-97.3)	0.57	0.01
Hb (<10 g/dL)	1.99 (1.29-3.07)	0 (0-64)	1.99 (1.29-3.07)	0.99	16 (12-22)	92 (86-95)	6.7 (3.2-13.5)	96.1 (92.7-98.0)	0.31	0.74
Hb (<8 g/dL)	2.87 (1.40-5.89)	0 (0-30)	3.48 (1.76-6.85)	0.51	4 (2-6)	86 (75-92)	7.4 (2.7-18.6)	95.9 (92.2-97.9)	0.45	0.59
BMI (<18.5 kg/m²)	2.86 (2.17-3.75)	0 (0-0)	2.90 (2.21-3.81)	0.55	9 (6-13)	95 (93-96)	11.3 (7.4-16.8)	95.7 (94.0-96.9)	0.47	0.72
Lymphadenopathy	4.02 (2.26-7.16)	0 (0-0)	4.02 (2.26-7.16)	0.73	11 (4-29)	97 (96-98)	12.5 (4.3-31.3)	96.2 (92.1-98.2)	0.69	0.55
W4SS with CRP (>=10 mg/L)	1.66 (1.01-2.72)	0 (0-74)	1.42 (0.87-2.31)	0.17	66 (40-85)	99 (98-99)	7.3 (4.4-11.9)	96.0 (91.8-98.1)	-	0.98
W4SS with CXR (abnormal)	3.56 (1.94-6.54)	19 (0-62)	4.02 (1.85-8.71)	0.33	73 (58-84)	99 (98-99)	6.2 (3.9-9.9)	98.4 (96.2-99.4)	0.62	0.13
W4SS then CRP (>=5 mg/L)	7.56 (3.48-16.40)	57 (0-84)	5.24 (2.61-10.50)	0.52	22 (10-43)	96 (93-98)	15.0 (8.9-24.2)	97.7 (95.7-98.8)	-	0.81
W4SS then Xpert*	128.84 (74.31-223.38)	10 (0-50)	149.81 (83.48-268.82)	0.55	3 (2-5)	83 (72-90)	82.5 (67.5-91.5)	97.1 (95.3-98.2)	0.05	0.09
Xpert alone*	145.18 (78.32-269.13)	36 (0-68)	151.58 (83.62-274.78)	0.78	4 (3-6)	78 (62-87)	77.0 (66.0-85.2)	97.9 (96.0-98.9)	0.02	0.03

†For the meta-regressions, the outcome variable was the diagnostic odds ratio

††Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

†††Pooled using a trivariate generalized linear mixed model that jointly models diagnostic test prevalence and predictive values

*Estimates for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, GLMM = generalized linear mixed model, Hb = haemoglobin, NPV = negative predictive value, PPV = positive predictive value, W4SS = WHO four-symptom screen

Table 8-9F - Additional diagnostic accuracy estimates in pregnant participants

									Metaregression†	
	Univariate diagnostic odds ratio				Positive screen (%)††		Trivariate GLMM†††		Reference standard	Prevalence
Test	Estimate (95% CI)	I² (95% CI)	Trim-and-fill estimate (95% CI)	Egger's test (p-value)	Estimate (95% CI)	I² (95% CI)	PPV (95% CI)	NPV (95% CI)	P-value	P-value
W4SS	2.22 (1.23-4.00)	0 (0-0)	2.27 (1.27-4.04)	0.89	43 (26-62)	94 (90-96)	5.8 (3.4-9.8)	98.7 (92.8-99.8)	0.86	0.88
CRP (>=10 mg/L)#	-	-	-	-	-	-	-	-	-	-
CRP (>=8 mg/L)#	-	-	-	-	-	-	-	-	-	-
CRP (>=5 mg/L)#	-	-	-	-	-	-	-	-	-	-
CXR (abnormal)	6.60 (0.19-225.79)	-	-	-	38 (12-72)	-	33.3 (1.7-79.2)	100.0 (56.6-100.0)	-	-
CXR (suggests tuberculosis)	39.00 (0.53-2883.60)	-	-	-	14 (2-58)	-	100.0 (5.1-100.0)	100.0 (61.0-100.0)	-	-
Cough (any)	4.05 (2.15-7.61)	0 (0-0)	3.73 (2.03-6.85)	0.18	20 (13-32)	90 (82-94)	10.0 (5.5-17.5)	98.6 (96.1-99.5)	0.79	0.57
Cough (>=2 weeks)	7.25 (3.68-14.27)	0 (0-0)	6.89 (3.55-13.39)	0.64	9 (5-16)	82 (67-91)	17.0 (9.0-29.8)	98.0 (95.6-99.1)	0.9	0.96
Hb (<10 g/dL)	1.33 (0.57-3.09)	0 (0-85)	1.33 (0.57-3.09)	0.57	26 (16-40)	78 (47-91)	3.1 (1.5-6.3)	98.2 (89.9-99.7)	0.24	0.51
Hb (<8 g/dL)##	1.98 (0.25-15.58)	0 (0-83)	0.69 (0.12-4.11)	0.02	2 (1-3)	6 (0-80)	-	-	0.51	0.29
BMI (<18.5 kg/m²)	5.68 (1.39-23.17)	0 (0-69)	5.68 (1.39-23.17)	0.85	4 (2-6)	32 (0-71)	1.3 (0.0-93.6)	97.5 (95.3-98.7)	0.96	0.76
Lymphadenopathy#	-	-	-	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶#	-	-	-	-	-	-	-	-	-	-
W4SS with CXR (abnormal)	3.86 (0.12-126.73)	-	-	-	50 (20-80)	-	25.0 (1.3-69.9)	100.0 (51.0-100.0)	-	-
W4SS then CRP (>=5 mg/L)¶#	-	-	-	-	-	-	-	-	-	-
W4SS then Xpert*	165.04 (24.50-1111.61)	0 (0-6)	165.04 (24.50-1111.61)	0.12	1 (0-2)	0 (0-29)	100.0 (0.0-100.0)	98.1 (96.5-99.0)	-	0.67
Xpert alone*	141.49 (31.79-629.82)	0 (0-0)	158.18 (39.84-627.97)	0.64	2 (1-4)	0 (0-76)	73.4 (40.5-91.8)	98.5 (97.0-99.3)	-	0.66

†For the meta-regressions, the outcome variable was the diagnostic odds ratio

††Calculated using meta-analysis of proportions. We used a generalized linear mixed model with logit transformation in preference to protocol specified DerSimonian and Laird random effects model for proportions with variance stabilization by applying the Freeman-Tukey double arcsine transformation

†††Pooled using a trivariate generalized linear mixed model that jointly models diagnostic test prevalence and predictive values

#Insufficient data to perform meta-analysis

##Trivariate random-effects model did not converge

*Estimates for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, GLMM = generalized linear mixed model, Hb = haemoglobin, NPV = negative predictive value, PPV = positive predictive value, W4SS = WHO four-symptom screen

Table 8-10: Yield of different screening and diagnostic algorithms when screening a population of 1000 persons

Table 8-10A - Yield of different screening and diagnostic algorithms at different prevalences when screening a population of 1000 participants

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
	1% prevalence														
W4SS	10	582	8	574	416	2	1.4	99.6	-	-	-	-	-	-	-
CRP (>=10 mg/L)	10	265	8	257	733	2	2.9	99.7	-	-	-	-	-	-	-
CRP (>=8 mg/L)	10	305	8	297	693	2	2.7	99.7	-	-	-	-	-	-	-
CRP (>=5 mg/L)	10	405	9	396	594	1	2.1	99.8	-	-	-	-	-	-	-
CXR (abnormal)	10	383	7	376	614	3	1.9	99.5	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	10	224	6	218	772	4	2.8	99.5	-	-	-	-	-	-	-
Cough (any)	10	283	6	277	713	4	2	99.4	-	-	-	-	-	-	-
Cough (>=2 weeks)	10	162	4	158	832	6	2.3	99.3	-	-	-	-	-	-	-
Hb (<10 g/dL)	10	202	4	198	792	6	2.1	99.3	-	-	-	-	-	-	-
Hb (<8 g/dL)	10	41	1	40	950	9	2.9	99.1	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	10	112	3	109	881	7	2.6	99.2	-	-	-	-	-	-	-
Lymphadenopathy	10	102	3	99	891	7	3	99.2	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	10	692	9	683	307	1	1.3	99.6	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	10	801	9	792	198	1	1.2	99.7	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	10	255	7	248	742	3	2.8	99.6	-	-	-	-	-	-	-
Xpert alone*†	10	-	-	-	-	-	-	-	7	980	10	3	40.7	99.7	143
WHO screen then Xpert*†	10	-	-	-	-	-	-	-	6	980	10	4	36.9	99.6	167
	5% prevalence														
W4SS	50	592	41	551	399	9	6.9	97.8	-	-	-	-	-	-	-
CRP (>=10 mg/L)	50	285	38	247	703	12	13.5	98.4	-	-	-	-	-	-	-
CRP (>=8 mg/L)	50	325	40	285	665	10	12.4	98.6	-	-	-	-	-	-	-
CRP (>=5 mg/L)	50	424	44	380	570	6	10.3	98.9	-	-	-	-	-	-	-
CXR (abnormal)	50	397	36	361	589	14	9.1	97.7	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	50	241	32	209	741	18	13.1	97.6	-	-	-	-	-	-	-
Cough (any)	50	294	28	266	684	22	9.5	96.9	-	-	-	-	-	-	-
Cough (>=2 weeks)	50	171	19	152	798	31	11.1	96.3	-	-	-	-	-	-	-
Hb (<10 g/dL)	50	212	22	190	760	28	10.2	96.4	-	-	-	-	-	-	-
Hb (<8 g/dL)	50	44	6	38	912	44	13.6	95.4	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	50	118	14	104	846	36	12.2	96	-	-	-	-	-	-	-
Lymphadenopathy	50	111	16	95	855	34	14	96.1	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	50	700	44	656	294	6	6.3	98	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	50	807	47	760	190	3	5.8	98.4	-	-	-	-	-	-	-

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
W4SS then CRP (>=5 mg/L)¶	50	273	35	238	712	15	12.8	97.9	-	-	-	-	-	-	-
Xpert alone*†	50	-	-	-	-	-	-	-	34	940	10	16	78.2	98.3	29
WHO screen then Xpert*†	50	-	-	-	-	-	-	-	29	940	10	21	75.3	97.8	34
	10% prevalence														
W4SS	100	604	82	522	378	18	13.6	95.5	-	-	-	-	-	-	-
CRP (>=10 mg/L)	100	311	77	234	666	23	24.8	96.7	-	-	-	-	-	-	-
CRP (>=8 mg/L)	100	351	81	270	630	19	23.1	97.1	-	-	-	-	-	-	-
CRP (>=5 mg/L)	100	447	87	360	540	13	19.5	97.6	-	-	-	-	-	-	-
CXR (abnormal)	100	414	72	342	558	28	17.4	95.2	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	100	261	63	198	702	37	24.1	95	-	-	-	-	-	-	-
Cough (any)	100	308	56	252	648	44	18.2	93.6	-	-	-	-	-	-	-
Cough (>=2 weeks)	100	182	38	144	756	62	20.9	92.4	-	-	-	-	-	-	-
Hb (<10 g/dL)	100	223	43	180	720	57	19.3	92.7	-	-	-	-	-	-	-
Hb (<8 g/dL)	100	48	12	36	864	88	25	90.8	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	100	128	29	99	801	71	22.7	91.9	-	-	-	-	-	-	-
Lymphadenopathy	100	121	31	90	810	69	25.6	92.2	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	100	709	88	621	279	12	12.4	95.9	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	100	814	94	720	180	6	11.5	96.8	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	100	295	70	225	675	30	23.7	95.7	-	-	-	-	-	-	-
Xpert alone*†	100	-	-	-	-	-	-	-	68	891	9	32	88.3	96.5	15
WHO screen then Xpert*†	100	-	-	-	-	-	-	-	58	891	9	42	86.6	95.5	17
	20% prevalence														
W4SS	200	628	164	464	336	36	26.1	90.3	-	-	-	-	-	-	-
CRP (>=10 mg/L)	200	362	154	208	592	46	42.5	92.8	-	-	-	-	-	-	-
CRP (>=8 mg/L)	200	402	162	240	560	38	40.3	93.6	-	-	-	-	-	-	-
CRP (>=5 mg/L)	200	494	174	320	480	26	35.2	94.9	-	-	-	-	-	-	-
CXR (abnormal)	200	448	144	304	496	56	32.1	89.9	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	200	302	126	176	624	74	41.7	89.4	-	-	-	-	-	-	-
Cough (any)	200	336	112	224	576	88	33.3	86.7	-	-	-	-	-	-	-
Cough (>=2 weeks)	200	204	76	128	672	124	37.3	84.4	-	-	-	-	-	-	-
Hb (<10 g/dL)	200	246	86	160	640	114	35	84.9	-	-	-	-	-	-	-
Hb (<8 g/dL)	200	56	24	32	768	176	42.9	81.4	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	200	146	58	88	712	142	39.7	83.4	-	-	-	-	-	-	-
Lymphadenopathy	200	142	62	80	720	138	43.7	83.9	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	200	728	176	552	248	24	24.2	91.2	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	200	828	188	640	160	12	22.7	93	-	-	-	-	-	-	-

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
W4SS then CRP (>=5 mg/L)¶	200	340	140	200	600	60	41.2	90.9	-	-	-	-	-	-	-
Xpert alone*†	200	-	-	-	-	-	-	-	136	792	8	64	94.4	92.5	7
WHO screen then Xpert*†	200	-	-	-	-	-	-	-	116	792	8	84	93.5	90.4	9

§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e. the number who need subsequent Xpert testing)

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

†The test accuracy of Xpert in those who were W4SS positive was: 12 studies; 4558 participants; sensitivity 0.74 (0.65-0.81), specificity 0.98 (0.97-0.99).

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-10B - Yield of different screening and diagnostic algorithms at different prevalences when screening a population of 1000 outpatients (on ART)

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
	1% prevalence														
W4SS	10	292	5	287	703	5	1.8	99.3	-	-	-	-	-	-	-
CRP (>=10 mg/L)	10	101	2	99	891	8	2	99.1	-	-	-	-	-	-	-
CRP (>=8 mg/L)	10	113	4	109	881	6	3.5	99.3	-	-	-	-	-	-	-
CRP (>=5 mg/L)	10	202	4	198	792	6	2	99.2	-	-	-	-	-	-	-
CXR (abnormal)	10	373	7	366	624	3	2	99.6	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	10	225	7	218	772	3	3.1	99.6	-	-	-	-	-	-	-
Cough (any)	10	172	4	168	822	6	2.3	99.3	-	-	-	-	-	-	-
Cough (>=2 weeks)	10	71	2	69	921	8	2.7	99.1	-	-	-	-	-	-	-
Hb (<10 g/dL)	10	184	6	178	812	4	3	99.4	-	-	-	-	-	-	-
Hb (<8 g/dL)	10	61	2	59	931	8	2.9	99.1	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	10	71	2	69	921	8	2.3	99.1	-	-	-	-	-	-	-
Lymphadenopathy	10	71	2	69	921	8	3.1	99.2	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	10	200	2	198	792	8	1	99	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	10	672	9	663	327	1	1.3	99.7	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	10	41	1	40	950	9	2	99	-	-	-	-	-	-	-
Xpert alone*†	10	-	-	-	-	-	-	-	5	980	10	5	34.9	99.5	200
WHO screen then Xpert*†	10	-	-	-	-	-	-	-	4	990	0	6	100	99.4	250
	5% prevalence														
W4SS	50	302	26	276	674	24	8.8	96.6	-	-	-	-	-	-	-
CRP (>=10 mg/L)	50	105	10	95	855	40	9.5	95.5	-	-	-	-	-	-	-
CRP (>=8 mg/L)	50	124	20	104	846	30	16.1	96.6	-	-	-	-	-	-	-
CRP (>=5 mg/L)	50	210	20	190	760	30	9.5	96.2	-	-	-	-	-	-	-
CXR (abnormal)	50	388	36	352	598	14	9.4	97.8	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	50	244	35	209	741	15	14.3	98	-	-	-	-	-	-	-
Cough (any)	50	181	20	161	789	30	11	96.3	-	-	-	-	-	-	-
Cough (>=2 weeks)	50	76	10	66	884	40	12.5	95.6	-	-	-	-	-	-	-
Hb (<10 g/dL)	50	199	28	171	779	22	13.9	97.2	-	-	-	-	-	-	-
Hb (<8 g/dL)	50	66	9	57	893	41	13.6	95.6	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	50	74	8	66	884	42	10.7	95.5	-	-	-	-	-	-	-
Lymphadenopathy	50	77	11	66	884	39	14.2	95.8	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	50	200	10	190	760	40	5	95	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	50	680	44	636	314	6	6.5	98.3	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	50	42	4	38	912	46	9.5	95.2	-	-	-	-	-	-	-
Xpert alone*†	50	-	-	-	-	-	-	-	26	940	10	24	73.6	97.6	38

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	50	-	-	-	-	-	-	-	18	950	0	32	100	96.8	56
	10% prevalence														
W4SS	100	314	53	261	639	47	16.9	93.1	-	-	-	-	-	-	-
CRP (>=10 mg/L)	100	110	20	90	810	80	18.2	91	-	-	-	-	-	-	-
CRP (>=8 mg/L)	100	139	40	99	801	60	28.8	93	-	-	-	-	-	-	-
CRP (>=5 mg/L)	100	220	40	180	720	60	18.2	92.3	-	-	-	-	-	-	-
CXR (abnormal)	100	406	73	333	567	27	18	95.5	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	100	268	70	198	702	30	26.1	95.9	-	-	-	-	-	-	-
Cough (any)	100	193	40	153	747	60	20.7	92.6	-	-	-	-	-	-	-
Cough (>=2 weeks)	100	82	19	63	837	81	23.2	91.2	-	-	-	-	-	-	-
Hb (<10 g/dL)	100	217	55	162	738	45	25.3	94.3	-	-	-	-	-	-	-
Hb (<8 g/dL)	100	72	18	54	846	82	25	91.2	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	100	79	16	63	837	84	20.3	90.9	-	-	-	-	-	-	-
Lymphadenopathy	100	85	22	63	837	78	25.9	91.5	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	100	200	20	180	720	80	10	90	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	100	692	89	603	297	11	12.9	96.4	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	100	44	8	36	864	92	18.2	90.4	-	-	-	-	-	-	-
Xpert alone*†	100	-	-	-	-	-	-	-	53	891	9	47	85.5	95	19
WHO screen then Xpert*†	100	-	-	-	-	-	-	-	37	900	0	63	100	93.5	27
	20% prevalence														
W4SS	200	338	106	232	568	94	31.4	85.8	-	-	-	-	-	-	-
CRP (>=10 mg/L)	200	120	40	80	720	160	33.3	81.8	-	-	-	-	-	-	-
CRP (>=8 mg/L)	200	168	80	88	712	120	47.6	85.6	-	-	-	-	-	-	-
CRP (>=5 mg/L)	200	240	80	160	640	120	33.3	84.2	-	-	-	-	-	-	-
CXR (abnormal)	200	442	146	296	504	54	33	90.3	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	200	316	140	176	624	60	44.3	91.2	-	-	-	-	-	-	-
Cough (any)	200	216	80	136	664	120	37	84.7	-	-	-	-	-	-	-
Cough (>=2 weeks)	200	94	38	56	744	162	40.4	82.1	-	-	-	-	-	-	-
Hb (<10 g/dL)	200	254	110	144	656	90	43.3	87.9	-	-	-	-	-	-	-
Hb (<8 g/dL)	200	84	36	48	752	164	42.9	82.1	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	200	88	32	56	744	168	36.4	81.6	-	-	-	-	-	-	-
Lymphadenopathy	200	100	44	56	744	156	44	82.7	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	200	200	40	160	640	160	20	80	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	200	714	178	536	264	22	24.9	92.3	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	200	48	16	32	768	184	33.3	80.7	-	-	-	-	-	-	-
Xpert alone*†	200	-	-	-	-	-	-	-	106	792	8	94	93	89.4	9

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	200	-	-	-	-	-	-	-	74	800	0	126	100	86.4	14

§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e. the number who need subsequent Xpert testing)

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

†The test accuracy of Xpert in those who were W4SS positive was: 4 studies; 564 participants; sensitivity 0.91 (0.42-0.99), specificity 0.98 (0.91-1).

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-10C - Yield of different screening and diagnostic algorithms at different prevalences when screening a population of 1000 outpatients (not on ART)

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
	1% prevalence														
W4SS	10	632	8	624	366	2	1.3	99.6	-	-	-	-	-	-	-
CRP (>=10 mg/L)	10	335	8	327	663	2	2.5	99.7	-	-	-	-	-	-	-
CRP (>=8 mg/L)	10	374	8	366	624	2	2.3	99.8	-	-	-	-	-	-	-
CRP (>=5 mg/L)	10	474	9	465	525	1	1.9	99.8	-	-	-	-	-	-	-
CXR (abnormal)	10	383	7	376	614	3	1.9	99.5	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	10	214	6	208	782	4	2.9	99.5	-	-	-	-	-	-	-
Cough (any)	10	303	6	297	693	4	1.9	99.4	-	-	-	-	-	-	-
Cough (>=2 weeks)	10	182	4	178	812	6	2.4	99.3	-	-	-	-	-	-	-
Hb (<10 g/dL)	10	222	4	218	772	6	2	99.3	-	-	-	-	-	-	-
Hb (<8 g/dL)	10	41	1	40	950	9	2.7	99.1	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	10	132	3	129	861	7	2.4	99.2	-	-	-	-	-	-	-
Lymphadenopathy	10	92	3	89	901	7	3.3	99.2	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	10	801	9	792	198	1	1.2	99.7	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	10	822	10	812	178	0	1.2	99.7	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	10	364	8	356	634	2	2.3	99.7	-	-	-	-	-	-	-
Xpert alone*†	10	-	-	-	-	-	-	-	7	980	10	3	42.8	99.7	143
WHO screen then Xpert*†	10	-	-	-	-	-	-	-	6	980	10	4	39.3	99.6	167
	5% prevalence														
W4SS	50	640	42	598	352	8	6.6	97.9	-	-	-	-	-	-	-
CRP (>=10 mg/L)	50	355	42	313	637	8	11.7	98.7	-	-	-	-	-	-	-
CRP (>=8 mg/L)	50	394	42	352	598	8	10.8	98.8	-	-	-	-	-	-	-
CRP (>=5 mg/L)	50	490	44	446	504	6	9.1	98.9	-	-	-	-	-	-	-
CXR (abnormal)	50	397	36	361	589	14	9	97.6	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	50	230	31	199	751	19	13.4	97.5	-	-	-	-	-	-	-
Cough (any)	50	314	29	285	665	21	9.2	96.9	-	-	-	-	-	-	-
Cough (>=2 weeks)	50	193	22	171	779	28	11.2	96.5	-	-	-	-	-	-	-
Hb (<10 g/dL)	50	231	22	209	741	28	9.5	96.4	-	-	-	-	-	-	-
Hb (<8 g/dL)	50	44	6	38	912	44	12.6	95.3	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	50	140	16	124	826	34	11.2	96	-	-	-	-	-	-	-
Lymphadenopathy	50	100	15	85	865	35	14.9	96.1	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	50	807	47	760	190	3	5.8	98.4	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	50	827	48	779	171	2	5.7	98.6	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	50	384	42	342	608	8	10.9	98.7	-	-	-	-	-	-	-
Xpert alone*†	50	-	-	-	-	-	-	-	37	940	10	13	79.6	98.6	27

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	50	-	-	-	-	-	-	-	32	940	10	18	77.1	98.1	31
	10% prevalence														
W4SS	100	652	85	567	333	15	13	95.7	-	-	-	-	-	-	-
CRP (>=10 mg/L)	100	380	83	297	603	17	21.8	97.3	-	-	-	-	-	-	-
CRP (>=8 mg/L)	100	418	85	333	567	15	20.3	97.4	-	-	-	-	-	-	-
CRP (>=5 mg/L)	100	512	89	423	477	11	17.4	97.7	-	-	-	-	-	-	-
CXR (abnormal)	100	413	71	342	558	29	17.2	95.1	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	100	251	62	189	711	38	24.7	94.9	-	-	-	-	-	-	-
Cough (any)	100	328	58	270	630	42	17.7	93.8	-	-	-	-	-	-	-
Cough (>=2 weeks)	100	205	43	162	738	57	21	92.8	-	-	-	-	-	-	-
Hb (<10 g/dL)	100	242	44	198	702	56	18.2	92.6	-	-	-	-	-	-	-
Hb (<8 g/dL)	100	47	11	36	864	89	23.4	90.7	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	100	148	31	117	783	69	20.9	91.9	-	-	-	-	-	-	-
Lymphadenopathy	100	111	30	81	819	70	27	92.1	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	100	814	94	720	180	6	11.5	96.8	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	100	833	95	738	162	5	11.4	97	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	100	408	84	324	576	16	20.6	97.3	-	-	-	-	-	-	-
Xpert alone*†	100	-	-	-	-	-	-	-	74	891	9	26	89.2	97.2	14
WHO screen then Xpert*†	100	-	-	-	-	-	-	-	64	891	9	36	87.7	96.1	16
	20% prevalence														
W4SS	200	674	170	504	296	30	25.2	90.8	-	-	-	-	-	-	-
CRP (>=10 mg/L)	200	430	166	264	536	34	38.6	94	-	-	-	-	-	-	-
CRP (>=8 mg/L)	200	466	170	296	504	30	36.5	94.4	-	-	-	-	-	-	-
CRP (>=5 mg/L)	200	554	178	376	424	22	32.1	95.1	-	-	-	-	-	-	-
CXR (abnormal)	200	446	142	304	496	58	31.8	89.5	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	200	292	124	168	632	76	42.5	89.3	-	-	-	-	-	-	-
Cough (any)	200	356	116	240	560	84	32.6	87	-	-	-	-	-	-	-
Cough (>=2 weeks)	200	230	86	144	656	114	37.4	85.2	-	-	-	-	-	-	-
Hb (<10 g/dL)	200	264	88	176	624	112	33.3	84.8	-	-	-	-	-	-	-
Hb (<8 g/dL)	200	54	22	32	768	178	40.7	81.2	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	200	166	62	104	696	138	37.3	83.5	-	-	-	-	-	-	-
Lymphadenopathy	200	132	60	72	728	140	45.5	83.9	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	200	828	188	640	160	12	22.7	93	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	200	846	190	656	144	10	22.5	93.5	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	200	456	168	288	512	32	36.8	94.1	-	-	-	-	-	-	-
Xpert alone*†	200	-	-	-	-	-	-	-	148	792	8	52	94.9	93.8	7

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	200	-	-	-	-	-	-	-	128	792	8	72	94.1	91.7	8

§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e. the number who need subsequent Xpert testing)

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

†The test accuracy of Xpert in those who were W4SS positive was: 10 studies; 3909 participants; sensitivity 0.76 (0.67-0.83), specificity 0.98 (0.97-0.99).

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-10D - Yield of different screening and diagnostic algorithms at different prevalences when screening a population of 1000 participants with a CD4 cell count <=200 cells/µL

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
	1% prevalence														
W4SS	10	680	44	636	314	6	6.4	98	-	-	-	-	-	-	-
CRP (>=10 mg/L)	10	405	44	361	589	6	10.9	99	-	-	-	-	-	-	-
CRP (>=8 mg/L)	10	444	45	399	551	5	10.1	99.1	-	-	-	-	-	-	-
CRP (>=5 mg/L)	10	540	46	494	456	4	8.5	99.1	-	-	-	-	-	-	-
CXR (abnormal)	10	435	36	399	551	14	8.4	97.6	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	10	279	32	247	703	18	11.5	97.5	-	-	-	-	-	-	-
Cough (any)	10	353	30	323	627	20	8.4	96.8	-	-	-	-	-	-	-
Cough (>=2 weeks)	10	230	21	209	741	29	9.1	96.2	-	-	-	-	-	-	-
Hb (<10 g/dL)	10	274	27	247	703	23	9.9	96.8	-	-	-	-	-	-	-
Hb (<8 g/dL)	10	54	7	47	903	43	12.8	95.5	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	10	179	18	161	789	32	9.8	96	-	-	-	-	-	-	-
Lymphadenopathy	10	120	16	104	846	34	13.6	96.2	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	10	778	46	732	218	4	6	98.4	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	10	865	48	817	133	2	5.5	98.5	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	10	363	40	323	627	10	11	98.4	-	-	-	-	-	-	-
Xpert alone*†	10	-	-	-	-	-	-	-	38	931	19	12	67	98.8	26
WHO screen then Xpert*†	10	-	-	-	-	-	-	-	34	931	19	16	64.5	98.4	29
	5% prevalence														
W4SS	50	690	87	603	297	13	12.6	95.8	-	-	-	-	-	-	-
CRP (>=10 mg/L)	50	430	88	342	558	12	20.5	97.9	-	-	-	-	-	-	-
CRP (>=8 mg/L)	50	468	90	378	522	10	19.2	98.1	-	-	-	-	-	-	-
CRP (>=5 mg/L)	50	560	92	468	432	8	16.4	98.2	-	-	-	-	-	-	-
CXR (abnormal)	50	451	73	378	522	27	16.2	95.1	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	50	298	64	234	666	36	21.5	94.9	-	-	-	-	-	-	-
Cough (any)	50	365	59	306	594	41	16.2	93.5	-	-	-	-	-	-	-
Cough (>=2 weeks)	50	240	42	198	702	58	17.5	92.4	-	-	-	-	-	-	-
Hb (<10 g/dL)	50	288	54	234	666	46	18.8	93.5	-	-	-	-	-	-	-
Hb (<8 g/dL)	50	59	14	45	855	86	23.7	90.9	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	50	188	35	153	747	65	18.6	92	-	-	-	-	-	-	-
Lymphadenopathy	50	132	33	99	801	67	25	92.3	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	50	786	93	693	207	7	11.8	96.7	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	50	870	96	774	126	4	11	96.9	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	50	386	80	306	594	20	20.7	96.7	-	-	-	-	-	-	-
Xpert alone*†	50	-	-	-	-	-	-	-	77	882	18	23	81.1	97.5	13

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	50	-	-	-	-	-	-	-	69	882	18	31	79.3	96.6	14
	10% prevalence														
W4SS	100	710	174	536	264	26	24.5	91	-	-	-	-	-	-	-
CRP (>=10 mg/L)	100	480	176	304	496	24	36.7	95.4	-	-	-	-	-	-	-
CRP (>=8 mg/L)	100	516	180	336	464	20	34.9	95.9	-	-	-	-	-	-	-
CRP (>=5 mg/L)	100	600	184	416	384	16	30.7	96	-	-	-	-	-	-	-
CXR (abnormal)	100	482	146	336	464	54	30.3	89.6	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	100	336	128	208	592	72	38.1	89.2	-	-	-	-	-	-	-
Cough (any)	100	390	118	272	528	82	30.3	86.6	-	-	-	-	-	-	-
Cough (>=2 weeks)	100	260	84	176	624	116	32.3	84.3	-	-	-	-	-	-	-
Hb (<10 g/dL)	100	316	108	208	592	92	34.2	86.5	-	-	-	-	-	-	-
Hb (<8 g/dL)	100	68	28	40	760	172	41.2	81.5	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	100	206	70	136	664	130	34	83.6	-	-	-	-	-	-	-
Lymphadenopathy	100	154	66	88	712	134	42.9	84.2	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	100	802	186	616	184	14	23.2	92.9	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	100	880	192	688	112	8	21.8	93.3	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	100	432	160	272	528	40	37	93	-	-	-	-	-	-	-
Xpert alone*†	100	-	-	-	-	-	-	-	154	784	16	46	90.6	94.5	6
WHO screen then Xpert*†	100	-	-	-	-	-	-	-	138	784	16	62	89.6	92.7	7
	20% prevalence														
W4SS	200	730	261	469	231	39	35.8	85.6	-	-	-	-	-	-	-
CRP (>=10 mg/L)	200	530	264	266	434	36	49.8	92.3	-	-	-	-	-	-	-
CRP (>=8 mg/L)	200	564	270	294	406	30	47.9	93.1	-	-	-	-	-	-	-
CRP (>=5 mg/L)	200	640	276	364	336	24	43.1	93.3	-	-	-	-	-	-	-
CXR (abnormal)	200	513	219	294	406	81	42.7	83.4	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	200	374	192	182	518	108	51.3	82.7	-	-	-	-	-	-	-
Cough (any)	200	415	177	238	462	123	42.7	79	-	-	-	-	-	-	-
Cough (>=2 weeks)	200	280	126	154	546	174	45	75.8	-	-	-	-	-	-	-
Hb (<10 g/dL)	200	344	162	182	518	138	47.1	79	-	-	-	-	-	-	-
Hb (<8 g/dL)	200	77	42	35	665	258	54.5	72	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	200	224	105	119	581	195	46.9	74.9	-	-	-	-	-	-	-
Lymphadenopathy	200	176	99	77	623	201	56.3	75.6	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	200	818	279	539	161	21	34.1	88.5	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	200	890	288	602	98	12	32.4	89.1	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	200	478	240	238	462	60	50.2	88.5	-	-	-	-	-	-	-
Xpert alone*†	200	-	-	-	-	-	-	-	231	686	14	69	94.3	90.9	4

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	200	-	-	-	-	-	-	-	207	686	14	93	93.7	88.1	5

§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e. the number who need subsequent Xpert testing)

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

†The test accuracy of Xpert in those who were W4SS positive was: 12 studies; 2315 participants; sensitivity 0.79 (0.7-0.86), specificity 0.97 (0.95-0.98).

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-10E - Yield of different screening and diagnostic algorithms at different prevalences when screening a population of 1000 participants with a CD4 cell count >200 cells/μL

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
	1% prevalence														
W4SS	10	512	7	505	485	3	1.4	99.4	-	-	-	-	-	-	-
CRP (>=10 mg/L)	10	214	6	208	782	4	3	99.6	-	-	-	-	-	-	-
CRP (>=8 mg/L)	10	245	7	238	752	3	2.7	99.6	-	-	-	-	-	-	-
CRP (>=5 mg/L)	10	354	8	346	644	2	2.2	99.7	-	-	-	-	-	-	-
CXR (abnormal)	10	363	7	356	634	3	1.8	99.5	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	10	194	6	188	802	4	3	99.5	-	-	-	-	-	-	-
Cough (any)	10	243	5	238	752	5	2.1	99.4	-	-	-	-	-	-	-
Cough (>=2 weeks)	10	132	3	129	861	7	2.6	99.2	-	-	-	-	-	-	-
Hb (<10 g/dL)	10	160	2	158	832	8	1.6	99.1	-	-	-	-	-	-	-
Hb (<8 g/dL)	10	41	1	40	950	9	1.5	99	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	10	81	2	79	911	8	2.5	99.1	-	-	-	-	-	-	-
Lymphadenopathy	10	102	3	99	891	7	2.9	99.2	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	10	651	8	643	347	2	1.2	99.4	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	10	722	9	713	277	1	1.3	99.7	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	10	204	6	198	792	4	2.8	99.5	-	-	-	-	-	-	-
Xpert alone*†	10	-	-	-	-	-	-	-	6	980	10	4	36.5	99.6	167
WHO screen then Xpert*†	10	-	-	-	-	-	-	-	5	980	10	5	31.7	99.5	200
	5% prevalence														
W4SS	50	520	36	484	466	14	6.8	97	-	-	-	-	-	-	-
CRP (>=10 mg/L)	50	231	32	199	751	18	14	97.7	-	-	-	-	-	-	-
CRP (>=8 mg/L)	50	262	34	228	722	16	12.8	97.8	-	-	-	-	-	-	-
CRP (>=5 mg/L)	50	371	39	332	618	11	10.5	98.2	-	-	-	-	-	-	-
CXR (abnormal)	50	375	33	342	608	17	8.8	97.3	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	50	210	30	180	770	20	14	97.4	-	-	-	-	-	-	-
Cough (any)	50	254	26	228	722	24	10.1	96.7	-	-	-	-	-	-	-
Cough (>=2 weeks)	50	141	17	124	826	33	12.1	96.2	-	-	-	-	-	-	-
Hb (<10 g/dL)	50	164	12	152	798	38	7.6	95.5	-	-	-	-	-	-	-
Hb (<8 g/dL)	50	41	3	38	912	47	7.3	95.1	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	50	86	10	76	874	40	11.6	95.6	-	-	-	-	-	-	-
Lymphadenopathy	50	110	15	95	855	35	13.6	96.1	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	50	656	39	617	333	11	5.9	96.8	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	50	730	46	684	266	4	6.2	98.3	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	50	219	29	190	760	21	13	97.2	-	-	-	-	-	-	-
Xpert alone*†	50	-	-	-	-	-	-	-	29	940	10	21	75	97.8	34

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	50	-	-	-	-	-	-	-	23	940	10	27	70.8	97.2	43
	10% prevalence														
W4SS	100	530	71	459	441	29	13.4	93.8	-	-	-	-	-	-	-
CRP (>=10 mg/L)	100	254	65	189	711	35	25.6	95.3	-	-	-	-	-	-	-
CRP (>=8 mg/L)	100	283	67	216	684	33	23.7	95.4	-	-	-	-	-	-	-
CRP (>=5 mg/L)	100	393	78	315	585	22	19.8	96.4	-	-	-	-	-	-	-
CXR (abnormal)	100	390	66	324	576	34	16.9	94.4	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	100	230	59	171	729	41	25.7	94.7	-	-	-	-	-	-	-
Cough (any)	100	267	51	216	684	49	19.1	93.3	-	-	-	-	-	-	-
Cough (>=2 weeks)	100	151	34	117	783	66	22.5	92.2	-	-	-	-	-	-	-
Hb (<10 g/dL)	100	169	25	144	756	75	14.8	91	-	-	-	-	-	-	-
Hb (<8 g/dL)	100	42	6	36	864	94	14.3	90.2	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	100	92	20	72	828	80	21.7	91.2	-	-	-	-	-	-	-
Lymphadenopathy	100	120	30	90	810	70	25	92	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	100	663	78	585	315	22	11.8	93.5	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	100	739	91	648	252	9	12.3	96.6	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	100	237	57	180	720	43	24.1	94.4	-	-	-	-	-	-	-
Xpert alone*†	100	-	-	-	-	-	-	-	57	891	9	43	86.4	95.4	18
WHO screen then Xpert*†	100	-	-	-	-	-	-	-	46	891	9	54	83.6	94.3	22
	20% prevalence														
W4SS	200	550	142	408	392	58	25.8	87.1	-	-	-	-	-	-	-
CRP (>=10 mg/L)	200	298	130	168	632	70	43.6	90	-	-	-	-	-	-	-
CRP (>=8 mg/L)	200	326	134	192	608	66	41.1	90.2	-	-	-	-	-	-	-
CRP (>=5 mg/L)	200	436	156	280	520	44	35.8	92.2	-	-	-	-	-	-	-
CXR (abnormal)	200	420	132	288	512	68	31.4	88.3	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	200	270	118	152	648	82	43.7	88.8	-	-	-	-	-	-	-
Cough (any)	200	294	102	192	608	98	34.7	86.1	-	-	-	-	-	-	-
Cough (>=2 weeks)	200	172	68	104	696	132	39.5	84.1	-	-	-	-	-	-	-
Hb (<10 g/dL)	200	178	50	128	672	150	28.1	81.8	-	-	-	-	-	-	-
Hb (<8 g/dL)	200	44	12	32	768	188	27.3	80.3	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	200	104	40	64	736	160	38.5	82.1	-	-	-	-	-	-	-
Lymphadenopathy	200	140	60	80	720	140	42.9	83.7	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	200	676	156	520	280	44	23.1	86.4	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	200	758	182	576	224	18	24	92.6	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	200	274	114	160	640	86	41.6	88.2	-	-	-	-	-	-	-
Xpert alone*†	200	-	-	-	-	-	-	-	114	792	8	86	93.4	90.2	9

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	200	-	-	-	-	-	-	-	92	792	8	108	92	88	11

§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e. the number who need subsequent Xpert testing)

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

†The test accuracy of Xpert in those who were W4SS positive was: 12 studies; 2121 participants; sensitivity 0.64 (0.54-0.74), specificity 0.99 (0.98-0.99).

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-10F - Yield of different screening and diagnostic algorithms at different prevalences when screening a population of 1000 pregnant participants#

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
	1% prevalence														
W4SS	10	424	8	416	574	2	2	99.7	-	-	-	-	-	-	-
CRP (>=10 mg/L)	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=8 mg/L)	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=5 mg/L)	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CXR (abnormal)	10	315	8	307	683	2	2.4	99.6	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	10	77	8	69	921	2	9.8	99.7	-	-	-	-	-	-	-
Cough (any)	10	195	7	188	802	3	3.4	99.6	-	-	-	-	-	-	-
Cough (>=2 weeks)	10	84	5	79	911	5	5.6	99.4	-	-	-	-	-	-	-
Hb (<10 g/dL)	10	250	2	248	742	8	0.8	98.9	-	-	-	-	-	-	-
Hb (<8 g/dL)	10	20	0	20	970	10	0	99	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	10	40	0	40	950	10	0	99	-	-	-	-	-	-	-
Lymphadenopathy	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	10	444	8	436	554	2	1.7	99.6	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xpert alone*†	10	-	-	-	-	-	-	-	5	980	10	5	34.9	99.5	200
WHO screen then Xpert*†	10	-	-	-	-	-	-	-	4	990	0	6	100	99.4	250
	5% prevalence														
W4SS	50	441	42	399	551	8	9.5	98.6	-	-	-	-	-	-	-
CRP (>=10 mg/L)	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=8 mg/L)	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=5 mg/L)	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CXR (abnormal)	50	332	38	294	656	12	11.3	98.1	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	50	104	38	66	884	12	36.1	98.6	-	-	-	-	-	-	-
Cough (any)	50	214	34	180	770	16	15.7	97.9	-	-	-	-	-	-	-
Cough (>=2 weeks)	50	100	24	76	874	26	23.6	97.1	-	-	-	-	-	-	-
Hb (<10 g/dL)	50	248	10	238	712	40	4	94.7	-	-	-	-	-	-	-
Hb (<8 g/dL)	50	19	0	19	931	50	0	94.9	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	50	38	0	38	912	50	0	94.8	-	-	-	-	-	-	-
Lymphadenopathy	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	50	456	38	418	532	12	8.2	97.7	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xpert alone*†	50	-	-	-	-	-	-	-	26	940	10	24	73.6	97.6	38

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	50	-	-	-	-	-	-	-	18	950	0	32	100	96.7	56
	10% prevalence														
W4SS	100	462	84	378	522	16	18.2	97	-	-	-	-	-	-	-
CRP (>=10 mg/L)	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=8 mg/L)	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=5 mg/L)	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CXR (abnormal)	100	354	75	279	621	25	21.2	96.1	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	100	138	75	63	837	25	54.3	97.1	-	-	-	-	-	-	-
Cough (any)	100	238	67	171	729	33	28.2	95.7	-	-	-	-	-	-	-
Cough (>=2 weeks)	100	119	47	72	828	53	39.5	94	-	-	-	-	-	-	-
Hb (<10 g/dL)	100	245	20	225	675	80	8.2	89.4	-	-	-	-	-	-	-
Hb (<8 g/dL)	100	18	0	18	882	100	0	89.8	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	100	36	0	36	864	100	0	89.6	-	-	-	-	-	-	-
Lymphadenopathy	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	100	471	75	396	504	25	15.9	95.3	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xpert alone*†	100	-	-	-	-	-	-	-	53	891	9	47	85.5	95	19
WHO screen then Xpert*†	100	-	-	-	-	-	-	-	36	900	0	64	100	93.4	28
	20% prevalence														
W4SS	200	504	168	336	464	32	33.3	93.5	-	-	-	-	-	-	-
CRP (>=10 mg/L)	200	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=8 mg/L)	200	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRP (>=5 mg/L)	200	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CXR (abnormal)	200	398	150	248	552	50	37.7	91.7	-	-	-	-	-	-	-
CXR (suggests tuberculosis)	200	206	150	56	744	50	72.8	93.7	-	-	-	-	-	-	-
Cough (any)	200	286	134	152	648	66	46.9	90.8	-	-	-	-	-	-	-
Cough (>=2 weeks)	200	158	94	64	736	106	59.5	87.4	-	-	-	-	-	-	-
Hb (<10 g/dL)	200	240	40	200	600	160	16.7	78.9	-	-	-	-	-	-	-
Hb (<8 g/dL)	200	16	0	16	784	200	0	79.7	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	200	32	0	32	768	200	0	79.3	-	-	-	-	-	-	-
Lymphadenopathy	200	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶	200	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	200	502	150	352	448	50	29.9	90	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	200	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xpert alone*†	200	-	-	-	-	-	-	-	106	792	8	94	93	89.4	9

		Outcome of screening§							Outcome of screening then diagnosis§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
WHO screen then Xpert*†	200	-	-	-	-	-	-	-	72	800	0	128	100	86.2	14

#For lymphadenopathy and strategies containing CRP, there were insufficient data to perform meta-analysis

§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e. the number who need subsequent Xpert testing)

¶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

*Accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

†The test accuracy of Xpert in those who were W4SS positive was: 5 studies; 137 participants; sensitivity 0.71 (0.33-0.93), specificity 1 (0-1).

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-11: Sensitivity analyses using different reference standards

Table 8-11A - Sensitivity analyses using different reference standards in all participants

	Sensitivity analyses 1*				Sensitivity analyses 2†			
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)
W4SS	21	15652	82 (71-89)	42 (29-57)	12	8612	88 (76-94)	43 (27-61)
CRP (>=10 mg/L)	5	3582	73 (49-88)	74 (62-83)	4	3110	77 (35-95)	72 (56-84)
CRP (>=8 mg/L)	5	3582	77 (58-89)	70 (57-81)	4	3110	78 (31-96)	69 (52-82)
CRP (>=5 mg/L)	5	3582	85 (72-93)	60 (48-71)	4	3110	88 (62-97)	59 (43-72)
CXR (abnormal)	8	6222	72 (65-78)	62 (52-71)	4	4190	73 (65-79)	65 (55-74)
CXR (suggests tuberculosis)	8	6177	63 (56-70)	78 (67-86)	3	3749	64 (52-74)	84 (68-93)
Cough (any)	21	15623	56 (48-63)	73 (65-79)	12	8586	63 (55-71)	73 (64-80)
Cough (>=2 weeks)	17	10919	38 (28-49)	84 (77-90)	9	4545	38 (21-58)	86 (76-92)
Hb (<10 g/dL)	9	5118	43 (33-53)	80 (73-85)	5	2098	57 (41-72)	76 (63-86)
Hb (<8 g/dL)	9	5118	12 (9-15)	96 (93-97)	5	2098	13 (9-19)	94 (89-97)
BMI (<18.5 kg/m²)	18	12704	29 (21-37)	89 (84-92)	12	8464	38 (26-51)	87 (80-91)
Lymphadenopathy	4	2394	30 (14-54)	91 (75-97)	3	1404	23 (13-38)	94 (84-98)
W4SS with CRP (>=10 mg/L)¶	5	3582	87 (56-97)	31 (13-56)	4	3110	94 (46-100)	30 (10-63)
W4SS with CXR (abnormal)¶	8	6213	95 (90-97)	21 (10-38)	4	4182	95 (87-99)	31 (15-55)
W4SS then CRP (>=5 mg/L)¶	5	3582	66 (25-92)	75 (54-88)	4	3110	72 (24-96)	75 (47-91)

*Diagnostic accuracy estimates using culture or Xpert of sputum and/or other specimens as a reference standard

†Diagnostic accuracy estimates using only Xpert of sputum and/or other specimens as a reference standard

¶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

Table 8-11B - Sensitivity analyses using different reference standards in outpatients (on ART)

	Sensitivity analyses 1*				Sensitivity analyses 2†			
	No of studies	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**No of studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
W4SS	9	4328	56 (32-77)	71 (51-85)	4	2664	51 (20-82)	80 (67-88)
CRP (>=10 mg/L)	1	381	12 (2-54)	90 (93-87)	1	378	10 (1-67)	90 (93-87)
CRP (>=8 mg/L)	1	381	25 (6-62)	89 (92-85)	1	378	10 (1-67)	88 (91-85)
CRP (>=5 mg/L)	1	381	38 (13-72)	80 (84-76)	1	378	25 (3-76)	79 (83-75)
CXR (abnormal)	4	2679	72 (61-82)	64 (50-75)	2	2063	69 (56-80)	72 (64-79)
CXR (suggests tuberculosis)	4	2590	69 (52-82)	78 (62-89)	1	1745	55 (40-69)	92 (93-91)
Cough (any)	9	4328	41 (22-62)	83 (73-90)	4	2664	43 (29-58)	88 (85-91)
Cough (>=2 weeks)	6	1746	18 (4-51)	93 (79-98)	2	536	13 (0-86)	97 (94-98)
Hb (<10 g/dL)	4	844	49 (30-68)	83 (71-90)	2	629	18 (0-93)	77 (61-88)
Hb (<8 g/dL)	4	844	13 (4-31)	94 (91-96)	2	629	10 (2-37)	93 (90-94)
BMI (<18.5 kg/m²)	7	4054	17 (8-32)	93 (88-96)	4	2646	14 (1-76)	89 (82-94)
Lymphadenopathy	1	338	17 (7-38)	93 (95-90)	1	338	11 (3-34)	92 (95-89)
W4SS with CRP (>=10 mg/L)¶	1	381	12 (2-54)	79 (83-75)	1	378	10 (1-67)	79 (83-75)
W4SS with CXR (abnormal)¶	4	2679	90 (73-97)	33 (17-55)	2	2063	87 (62-97)	51 (33-70)
W4SS then CRP (>=5 mg/L)¶	1	381	6 (0-50)	96 (97-93)	1	378	10 (1-67)	96 (97-93)

*Diagnostic accuracy estimates using culture or Xpert of sputum and/or other specimens as a reference standard

†Diagnostic accuracy estimates using only Xpert of sputum and/or other specimens as a reference standard

¶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

Table 8-11C - Sensitivity analyses using different reference standards in outpatients (not on ART)

	Sensitivity analyses 1*				Sensitivity analyses 2†			
	No of studies	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**No of studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
W4SS	20	11196	85 (76-91)	37 (26-51)	11	5820	90 (80-96)	37 (24-53)
CRP (>=10 mg/L)	5	3198	82 (77-86)	67 (61-73)	4	2729	89 (85-92)	63 (56-70)
CRP (>=8 mg/L)	5	3198	84 (81-87)	64 (57-70)	4	2729	90 (86-94)	59 (51-67)
CRP (>=5 mg/L)	5	3198	89 (85-91)	54 (46-61)	4	2729	93 (89-96)	50 (42-59)
CXR (abnormal)	8	3543	72 (65-78)	62 (52-72)	4	2127	73 (64-80)	66 (54-76)
CXR (suggests tuberculosis)	8	3587	62 (55-69)	79 (67-87)	3	2004	64 (52-74)	85 (67-94)
Cough (any)	20	11167	58 (51-65)	70 (63-76)	11	5794	66 (59-73)	69 (62-75)
Cough (>=2 weeks)	16	9045	43 (34-52)	82 (75-87)	8	3881	48 (33-63)	81 (72-87)
Hb (<10 g/dL)	9	4271	44 (33-56)	79 (71-85)	5	1466	62 (43-78)	74 (58-85)
Hb (<8 g/dL)	9	4271	11 (8-15)	96 (93-98)	5	1466	14 (10-20)	94 (87-97)
BMI (<18.5 kg/m²)	17	8522	31 (23-40)	87 (82-92)	11	5690	42 (29-56)	85 (76-91)
Lymphadenopathy	4	2056	30 (13-54)	91 (76-97)	3	1066	25 (13-43)	95 (86-98)
W4SS with CRP (>=10 mg/L)¶	5	3198	94 (86-98)	20 (12-33)	4	2729	98 (93-100)	17 (9-30)
W4SS with CXR (abnormal)¶	8	3534	95 (92-98)	18 (9-34)	4	2119	96 (91-99)	29 (15-50)
W4SS then CRP (>=5 mg/L)¶	5	3198	83 (74-89)	64 (57-70)	4	2729	89 (84-92)	60 (52-68)

*Diagnostic accuracy estimates using culture or Xpert of sputum and/or other specimens as a reference standard

†Diagnostic accuracy estimates using only Xpert of sputum and/or other specimens as a reference standard

¶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

Table 8-11D - Sensitivity analyses using different reference standards in participants with CD4 cell count <=200 cells/µL

	Sensitivity analyses 1*				Sensitivity analyses 2†			
	No of studies	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**No of studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
W4SS	21	5636	88 (78-93)	33 (20-49)	12	3129	90 (78-96)	35 (19-55)
CRP (>=10 mg/L)	5	1597	88 (82-92)	63 (50-73)	4	1460	91 (84-95)	62 (48-75)
CRP (>=8 mg/L)	5	1597	90 (86-93)	59 (46-71)	4	1460	93 (86-97)	58 (43-72)
CRP (>=5 mg/L)	5	1597	92 (88-95)	48 (34-63)	4	1460	94 (88-97)	49 (33-65)
CXR (abnormal)	8	2210	74 (65-80)	58 (49-67)	4	1311	72 (62-80)	61 (49-71)
CXR (suggests tuberculosis)	8	2143	64 (56-71)	74 (64-82)	3	1157	61 (49-71)	81 (63-91)
Cough (any)	21	5626	59 (52-66)	67 (59-74)	12	3119	65 (56-73)	67 (58-75)
Cough (>=2 weeks)	17	4203	42 (33-50)	78 (70-85)	9	1889	38 (22-58)	80 (69-88)
Hb (<10 g/dL)	9	1971	53 (42-63)	74 (65-81)	5	916	65 (51-77)	71 (53-84)
Hb (<8 g/dL)	9	1971	14 (10-19)	95 (91-97)	5	916	15 (9-25)	93 (86-96)
BMI (<18.5 kg/m²)	18	4969	34 (26-43)	84 (77-88)	12	3098	42 (31-54)	80 (72-86)
Lymphadenopathy	4	1003	31 (13-56)	89 (73-96)	3	627	21 (11-35)	94 (87-97)
W4SS with CRP (>=10 mg/L)¶	5	1597	92 (52-99)	23 (10-47)	4	1460	97 (76-100)	24 (8-55)
W4SS with CXR (abnormal)¶	8	2208	97 (92-99)	14 (7-25)	4	1309	97 (88-99)	21 (11-36)
W4SS then CRP (>=5 mg/L)¶	5	1597	80 (51-94)	66 (37-86)	4	1460	95 (66-99)	69 (35-90)

*Diagnostic accuracy estimates using culture or Xpert of sputum and/or other specimens as a reference standard

†Diagnostic accuracy estimates using only Xpert of sputum and/or other specimens as a reference standard

¶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

Table 8-11E - Sensitivity analyses using different reference standards in participants with CD4 cell count >200 cells/µL

	Sensitivity analyses 1*				Sensitivity analyses 2†			
	No of studies	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**No of studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
W4SS	21	9633	71 (58-81)	50 (36-63)	12	5163	81 (63-91)	50 (33-67)
CRP (>=10 mg/L)	5	1829	61 (45-75)	79 (68-87)	4	1515	72 (45-89)	78 (62-88)
CRP (>=8 mg/L)##	5	1829	64 (47-78)	76 (64-85)	4	1515	78 (67-87)	74 (58-86)
CRP (>=5 mg/L)	5	1829	75 (59-87)	65 (54-75)	4	1515	86 (58-97)	64 (50-77)
CXR (abnormal)	8	3909	66 (59-73)	65 (53-74)	4	2784	73 (60-83)	67 (56-76)
CXR (suggests tuberculosis)	8	3932	60 (51-69)	81 (70-89)	3	2498	66 (46-82)	86 (69-94)
Cough (any)	21	9616	50 (41-60)	76 (68-82)	12	5149	61 (48-71)	76 (67-83)
Cough (>=2 weeks)	17	6416	34 (22-48)	87 (80-91)	9	2415	38 (16-65)	88 (80-93)
Hb (<10 g/dL)	9	3048	27 (18-38)	84 (79-89)	5	1112	38 (20-60)	82 (73-88)
Hb (<8 g/dL)	9	3048	6 (2-13)	97 (94-98)	5	1112	7 (2-25)	95 (91-97)
BMI (<18.5 kg/m²)	18	7396	20 (13-30)	92 (88-95)	12	5063	32 (20-47)	91 (85-94)
Lymphadenopathy	4	1374	31 (12-59)	91 (73-97)	3	770	27 (12-49)	94 (77-99)
W4SS with CRP (>=10 mg/L)¶	5	1829	76 (47-92)	35 (15-61)	4	1515	93 (4-100)	34 (12-66)
W4SS with CXR (abnormal)¶	8	3903	91 (82-96)	28 (16-44)	4	2778	93 (75-98)	35 (15-63)
W4SS then CRP (>=5 mg/L)¶	5	1829	53 (12-90)	80 (62-91)	4	1515	56 (5-97)	79 (55-92)

*Diagnostic accuracy estimates using culture or Xpert of sputum and/or other specimens as a reference standard

†Diagnostic accuracy estimates using only Xpert of sputum and/or other specimens as a reference standard

##For sensitivity analyses 2, the bivariate model did not converge; results from model assuming no correlation between sensitivity and 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

Table 8-11F - Sensitivity analyses using different reference standards in pregnant participants§

	Sensitivity analyses 1*				Sensitivity analyses 2†			
	No of studies	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**No of studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
W4SS	8	1935	77 (25-97)	58 (39-75)	5	489	71 (10-98)	55 (31-77)
CRP (>=10 mg/L)#	-	-	-	-	-	-	-	-
CRP (>=8 mg/L)#	-	-	-	-	-	-	-	-
CRP (>=5 mg/L)#	-	-	-	-	-	-	-	-
CXR (abnormal)	1	8	75 (11-99)	69 (91-33)	1	8	75 (11-99)	69 (91-33)
CXR (suggests tuberculosis)	1	7	75 (11-99)	93 (100-42)	1	7	75 (11-99)	93 (100-42)
Cough (any)	8	1933	62 (24-89)	81 (70-88)	5	487	53 (12-90)	81 (70-89)
Cough (>=2 weeks)	8	1933	43 (17-74)	92 (86-95)	5	488	17 (3-61)	92 (87-96)
Hb (<10 g/dL)#	5	1350	20 (10-36)	75 (61-85)	-	-	-	-
Hb (<8 g/dL)#	5	1350	0 (0-100)	98 (97-99)	-	-	-	-
BMI (<18.5 kg/m²)##	7	472	0 (0-98)	96 (94-98)	5	431	0 (0-100)	97 (94-98)
Lymphadenopathy#	-	-	-	-	-	-	-	-
W4SS with CRP (>=10 mg/L)¶#	-	-	-	-	-	-	-	-
W4SS with CXR (abnormal)¶	1	8	75 (11-99)	56 (84-24)	1	8	75 (11-99)	56 (84-24)
W4SS then CRP (>=5 mg/L)¶#	-	-	-	-	-	-	-	-

§For some analyses, all studies had 0% or 100% sensitivity/specificity; therefore, models may have given unreliable estimates such as 95% CIs that range from 0 to 100

*Diagnostic accuracy estimates using culture or Xpert of sputum and/or other specimens as a reference standard

†Diagnostic accuracy estimates using only Xpert of sputum and/or other specimens as a reference standard

#Insufficient data to perform meta-analysis

##Bivariate model did not converge; results from model assuming no correlation between sensitivity and 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

Table 8-12: Sensitivity analyses comparing W4SS followed by Xpert with CRP (≥10mg) followed by Xpert in all participants and by subgroups†

									Number of Xpert tests per 1000 PLHIV*							
	CRP (>=10 mg/L) then Xpert				W4SS then Xpert				1% prevalence		5% prevalence		10% prevalence		20% prevalence	
Subgroup	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	W4SS	CRP	W4SS	CRP	W4SS	CRP	W4SS	CRP
All participants	4	3099	54 (45-63)	100 (99-100)	4	3099	50 (28-71)	100 (98-100)	612	265	620	286	631	311	652	362
Outpatients (Not on ART)	4	2718	56 (51-60)	99 (99-100)	4	2718	59 (51-66)	99 (99-100)	732	345	738	365	747	390	764	440
Outpatients (On ART)	1	378	8 (1-62)	100 (100-98)	1	378	8 (1-62)	100 (100-98)	120	101	118	105	116	110	112	120
CD4 >200 cells/μL	4	1508	43 (34-53)	100 (7-100)	4	1508	44 (32-58)	100 (99-100)	572	214	578	230	585	251	600	292
CD4 <=200 cells/μL	4	1456	61 (51-71)	99 (98-100)	4	1456	63 (53-71)	99 (98-99)	682	365	689	386	698	411	716	462
Pregnant participants#	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

†Using culture as a reference standard; accuracy measures for entire algorithm using total Xpert (sputum and/or non-sputum Xpert result)

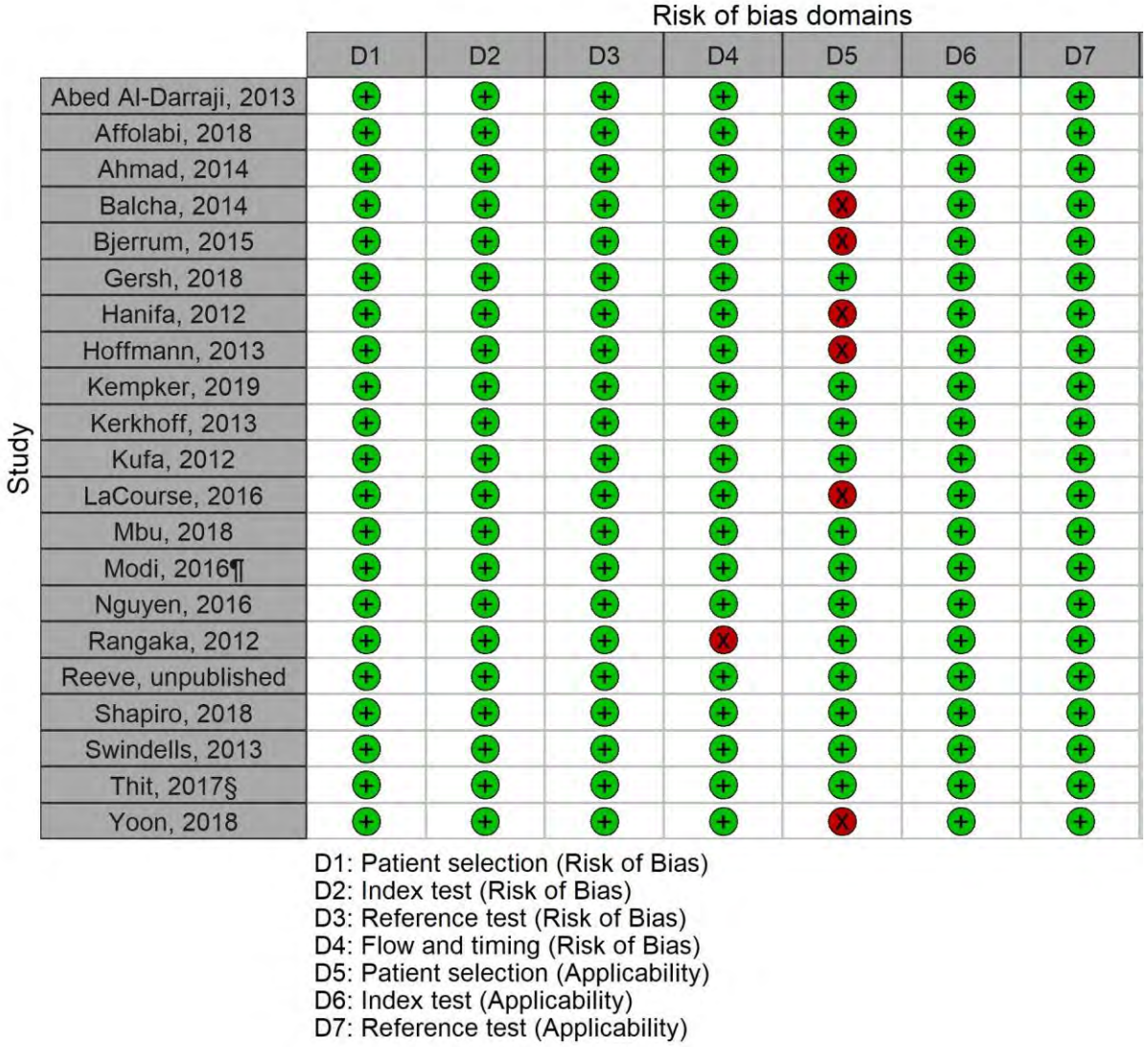
*Calculated using direct comparison estimates

#Insufficient data to perform meta-analysis

Definition of abbreviations: ART = antiretroviral therapy, CRP = C-reactive protein, PLHIV = people living with HIV, W4SS = WHO four-symptom screen

Figure 8-1: Risk of bias and applicability results on the QUADAS-2 criteria tool

Figure 8-1A - Risk of bias and applicability results on the QUADAS-2 criteria tool by study†



†In general, domains were assessed for all participants. Green = low risk of bias, red = high risk of bias
¶For the domain D4 (flow and timing), the risk of bias was judged high for chest X-ray as the index test (>20% missing data).
§For the domain D4 (flow and timing), the risk of bias was judged high for haemoglobin as the index test (>20% missing data).

Figure 8-1B - Clustered bar graphs of risk of bias and applicability results on the QUADAS-2 criteria tool

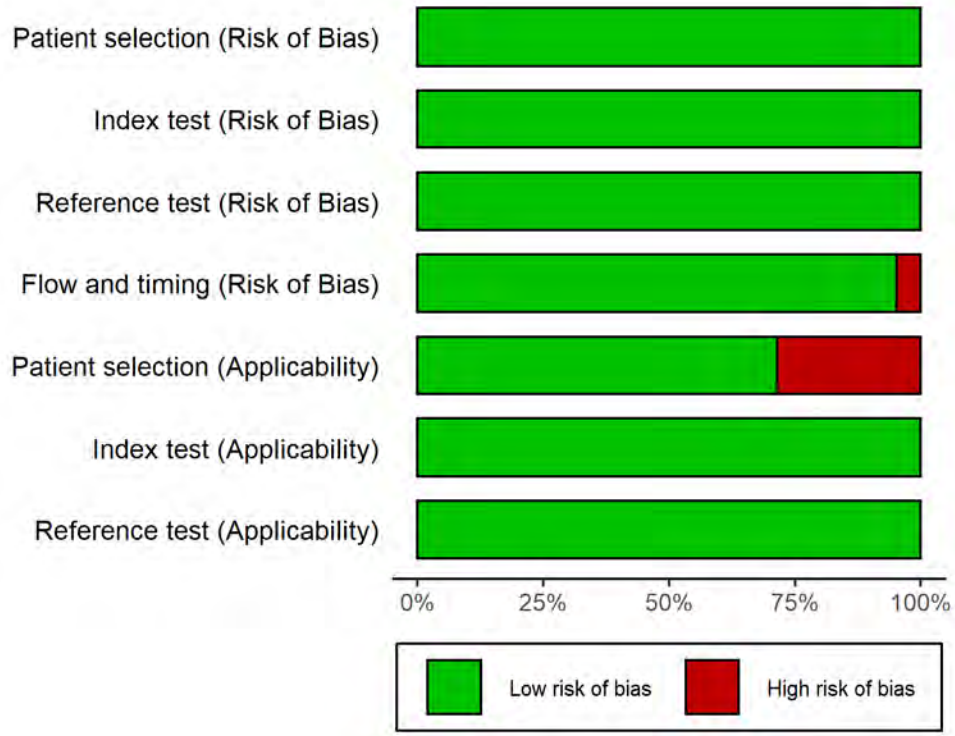
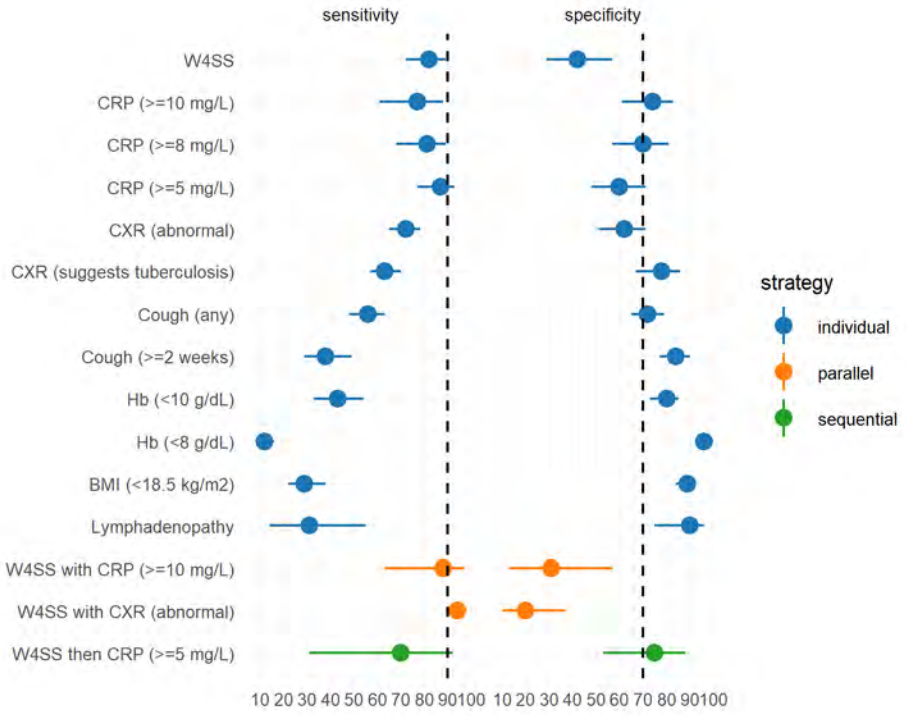


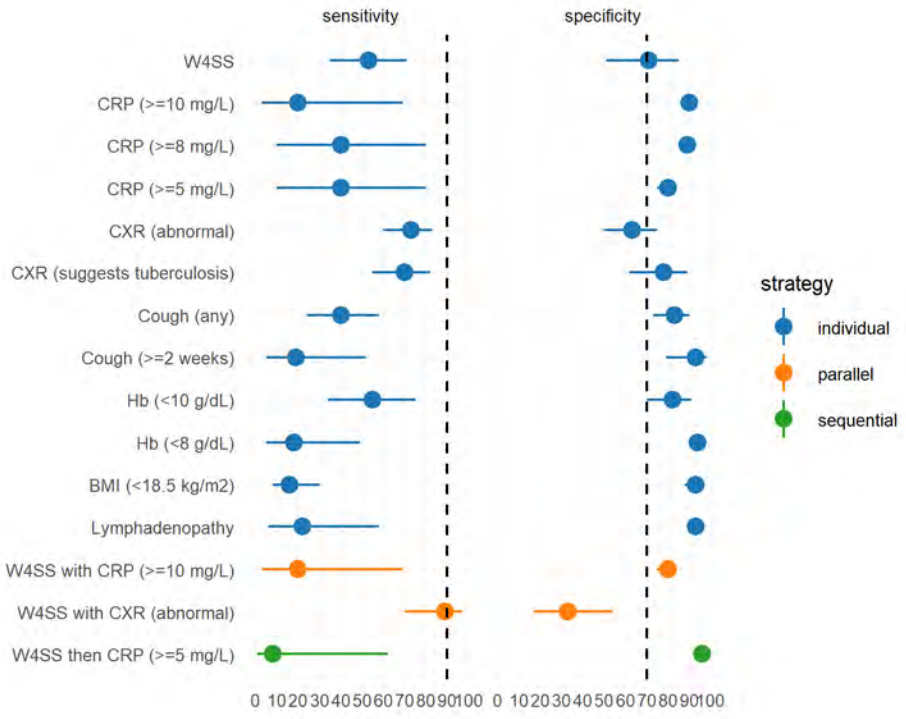
Figure 8-2: Plots of sensitivity and specificity for each screening test for the detection of tuberculosis

Figure 8-2A - Plots of sensitivity and specificity for each screening test for the detection of tuberculosis in all participants†



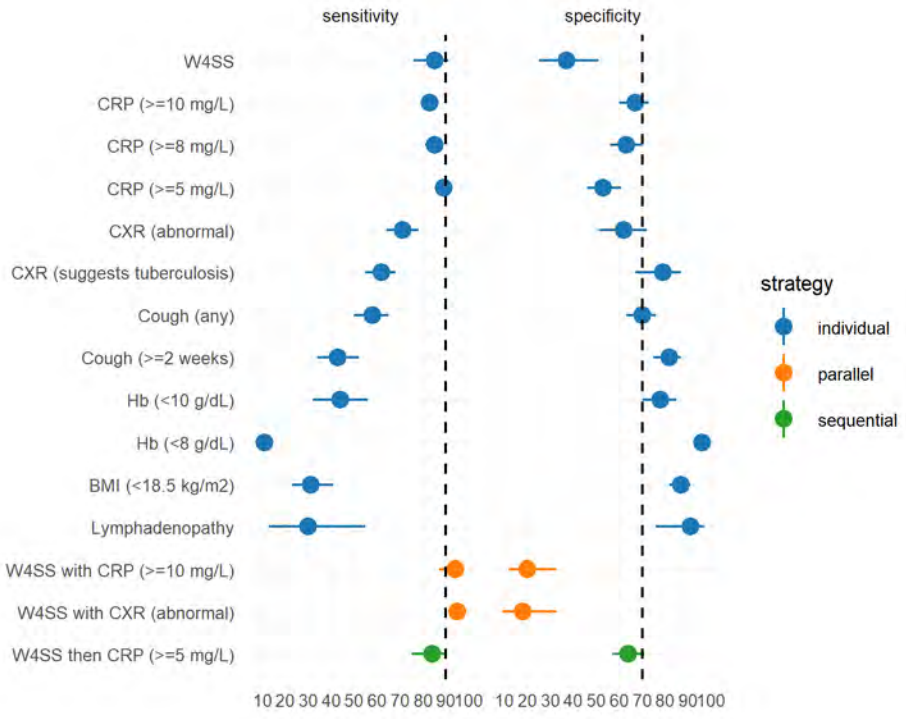
†Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity)
Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Figure 8-2B - Plots of sensitivity and specificity for each screening test for the detection of tuberculosis in outpatients (on ART)†



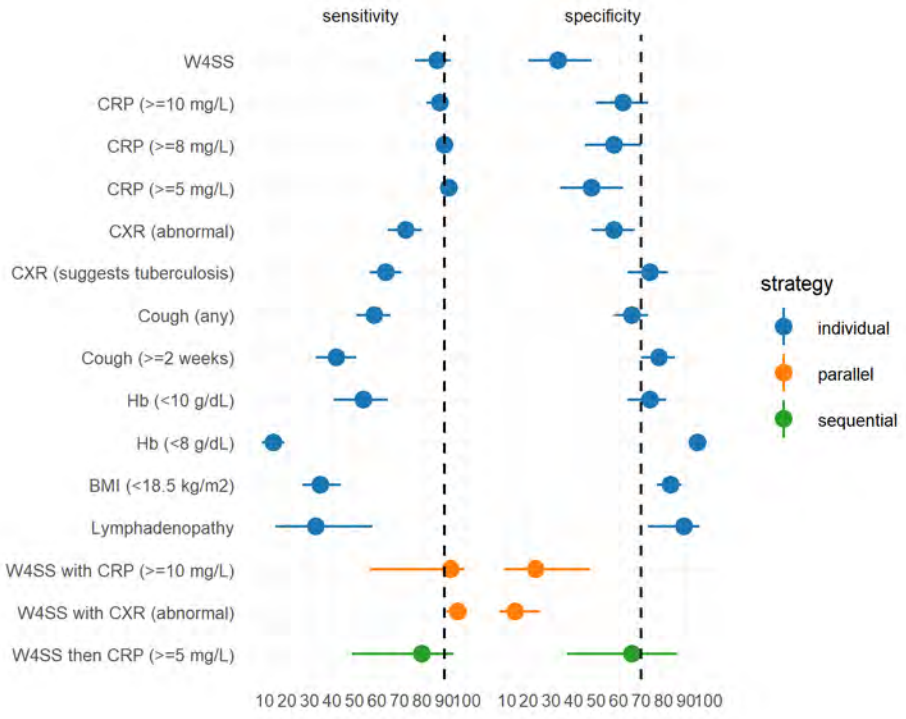
†Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity)
Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Figure 8-2C - Plots of sensitivity and specificity for each screening test for the detection of tuberculosis in outpatients (not on ART)†



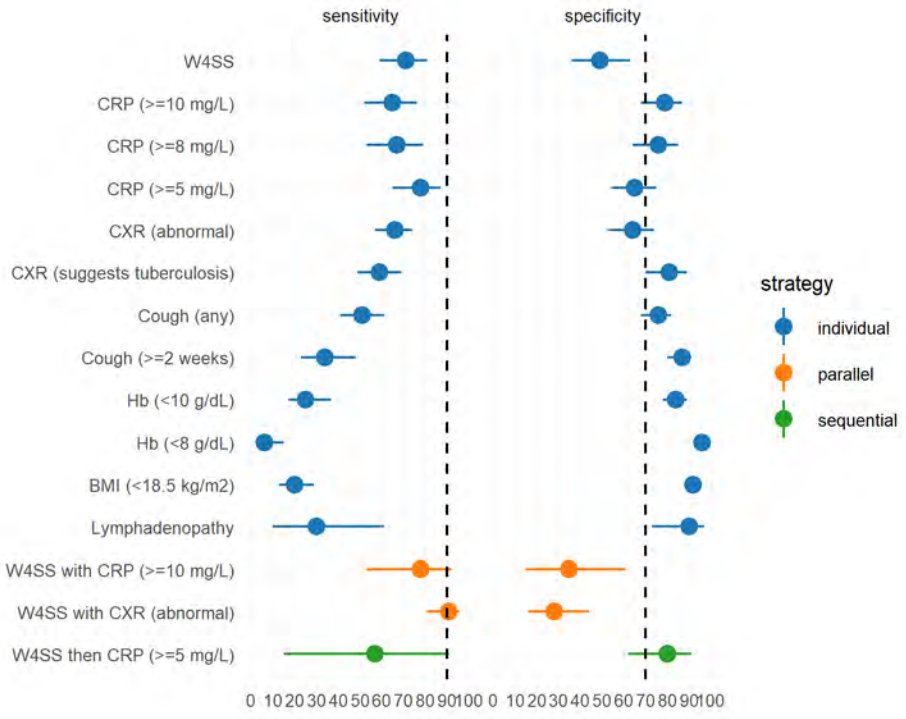
†Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity)
Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Figure 8-2D - Plots of sensitivity and specificity for each screening test for the detection of tuberculosis in participants with CD4 cell count <=200 cells/μL†



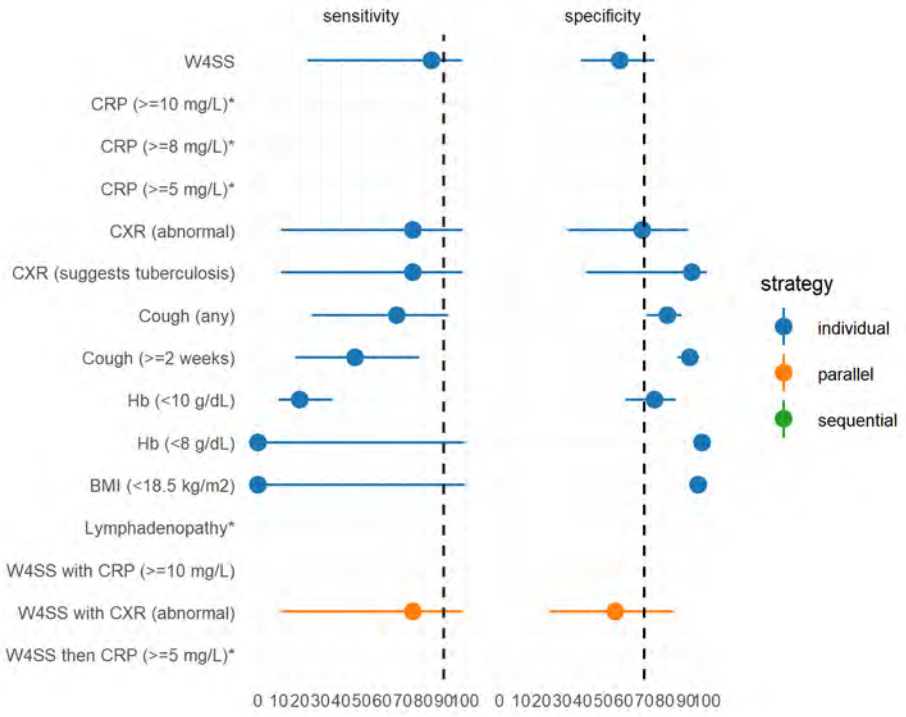
†Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity)
Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Figure 8-2E - Plots of sensitivity and specificity for each screening test for the detection of tuberculosis in participants with CD4 cell count >200 cells/μL†



†Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity)
Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

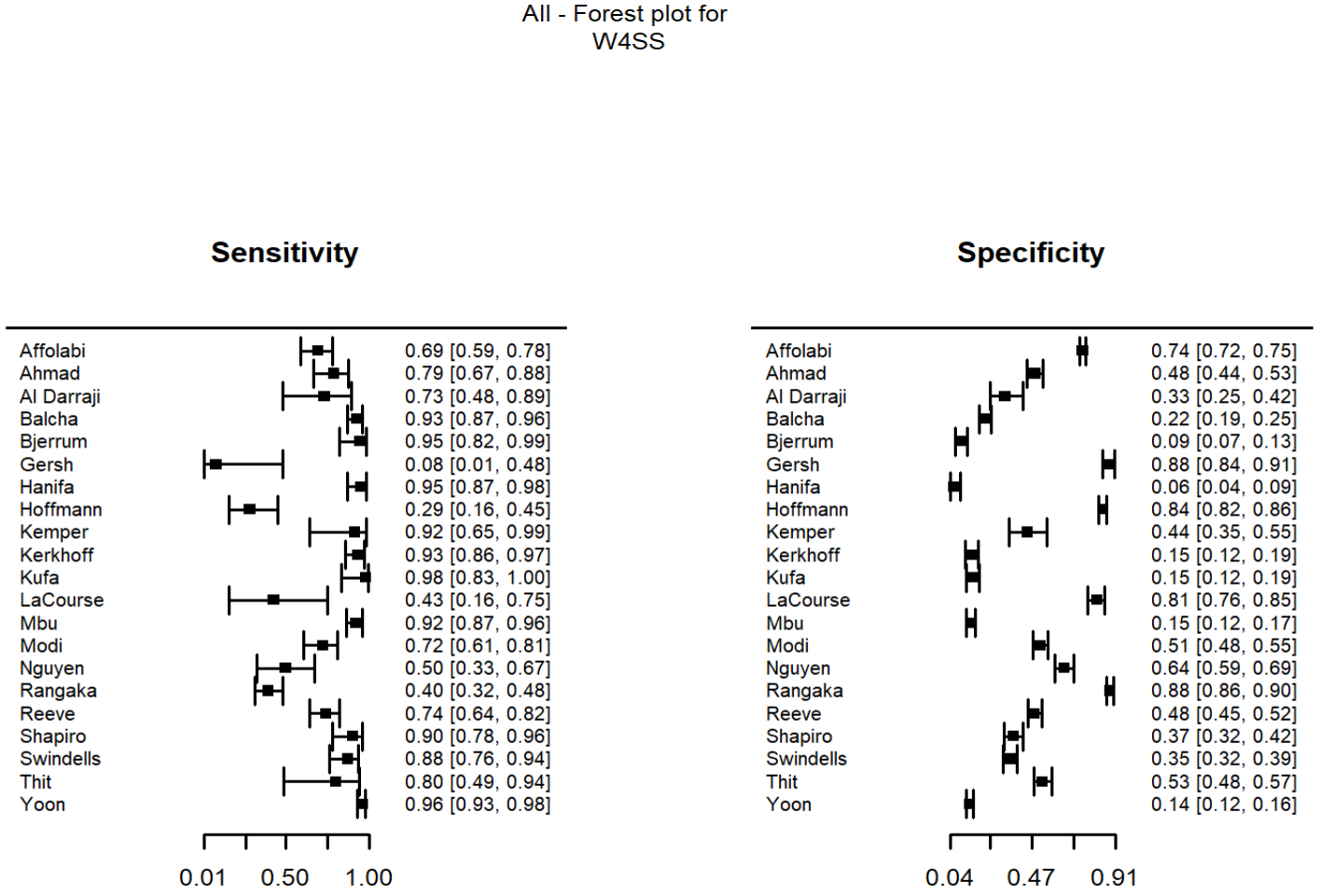
Figure 8-2F - Plots of sensitivity and specificity for each screening test for the detection of tuberculosis in in pregnant participants†*



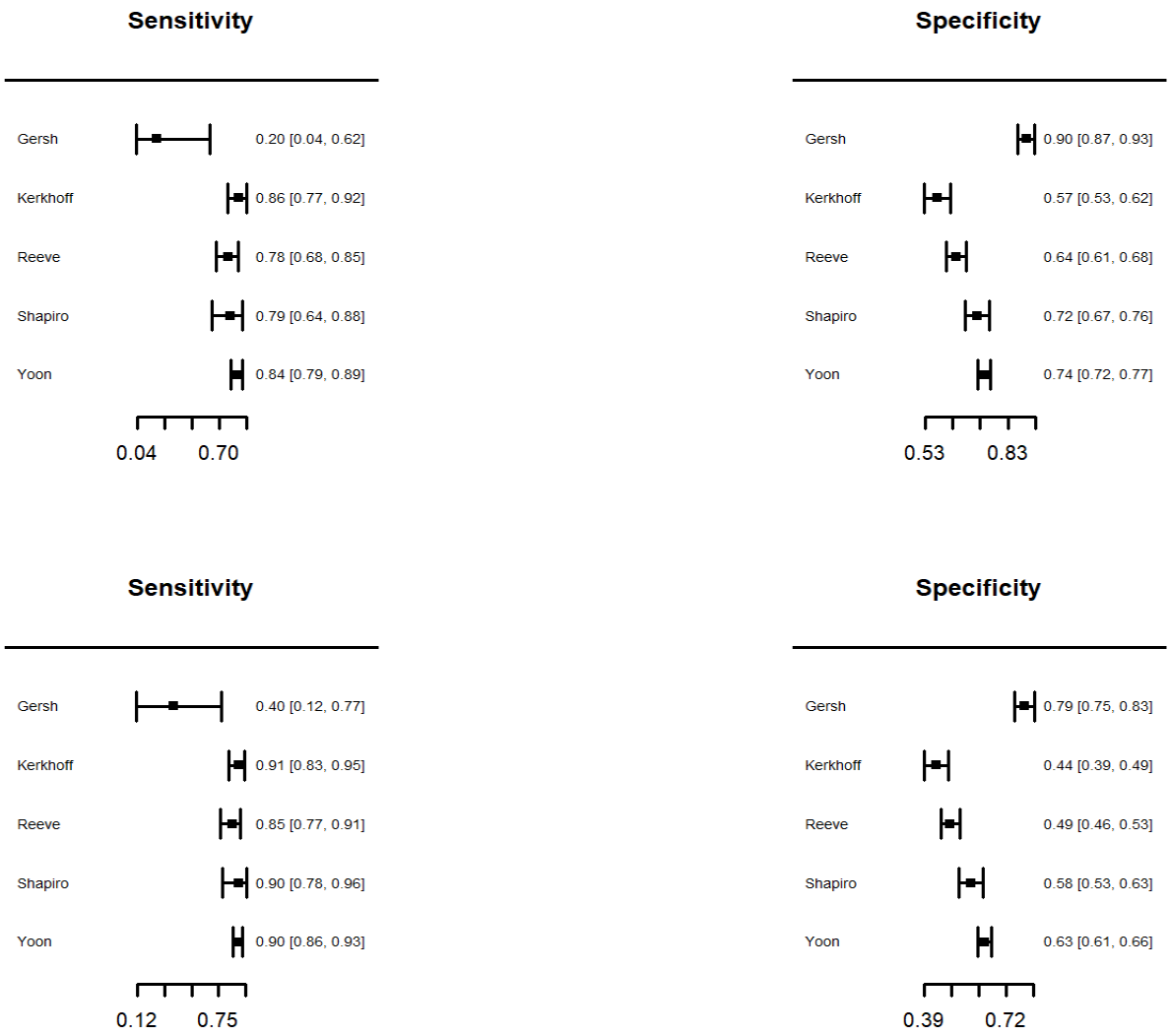
†Dashed lines indicate WHO's minimum requirements for a tuberculosis screening test (90% sensitivity and 70% specificity)
*Insufficient data to perform meta-analysis
Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, Hb = haemoglobin, W4SS = WHO four-symptom screen

Figure 8-3: Forest plots of sensitivity and specificity estimates in all participants and subgroups

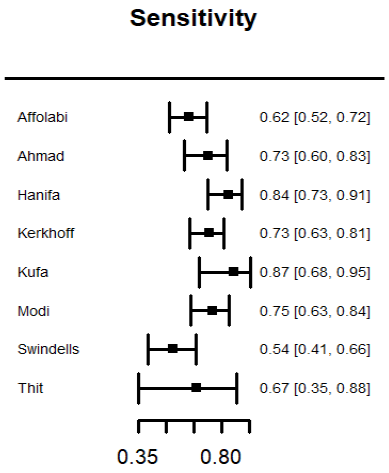
Figure 8-3A - Forest plots of sensitivity and specificity estimates in all participants

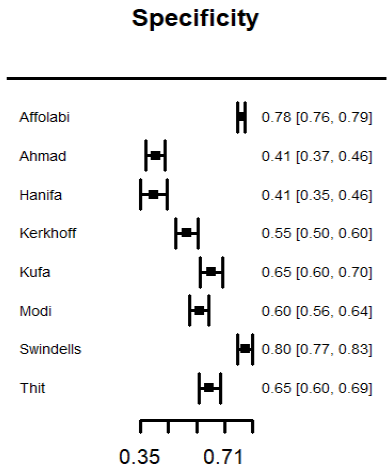


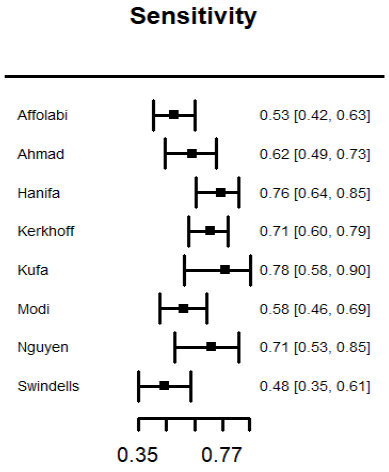
All - Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)

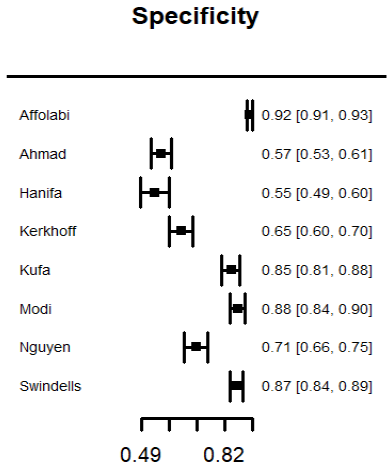


All - Forest plot for
Top: CXR (abnormal) and Bottom: CXR (suggests tuberculosis)



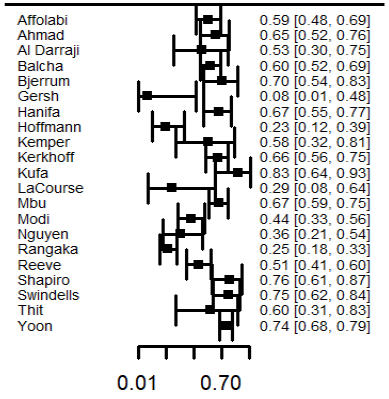




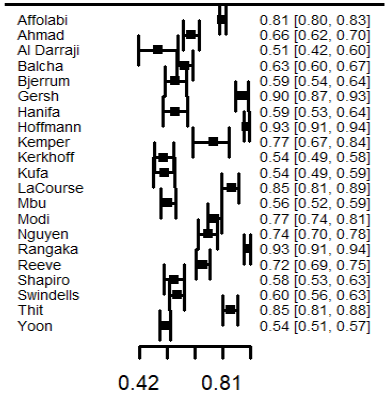


All - Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

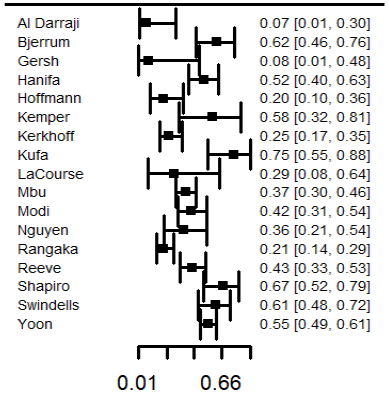
Sensitivity



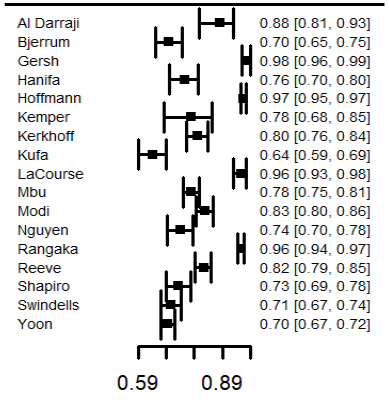
Specificity



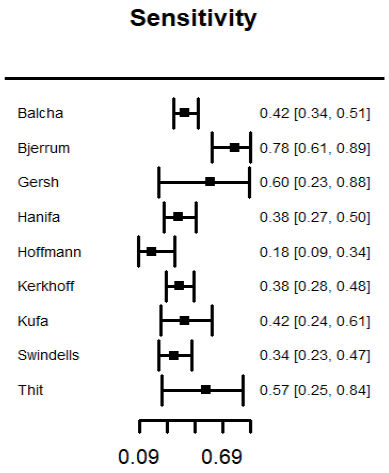
Sensitivity

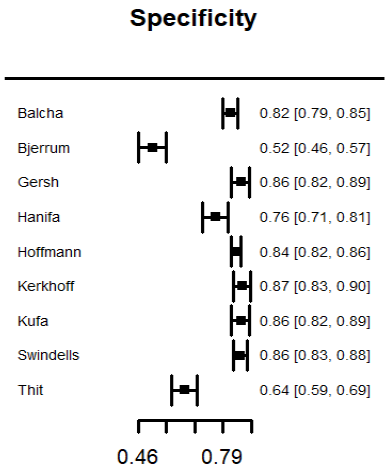


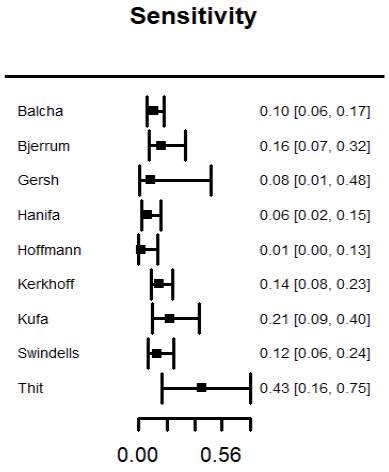
Specificity

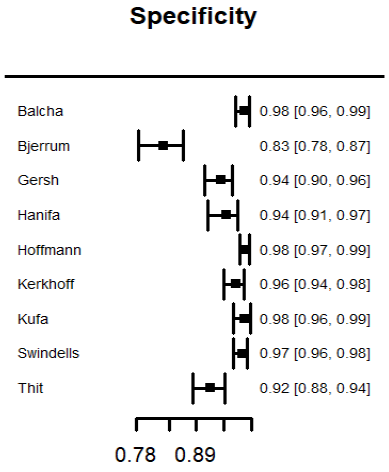


All - Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)

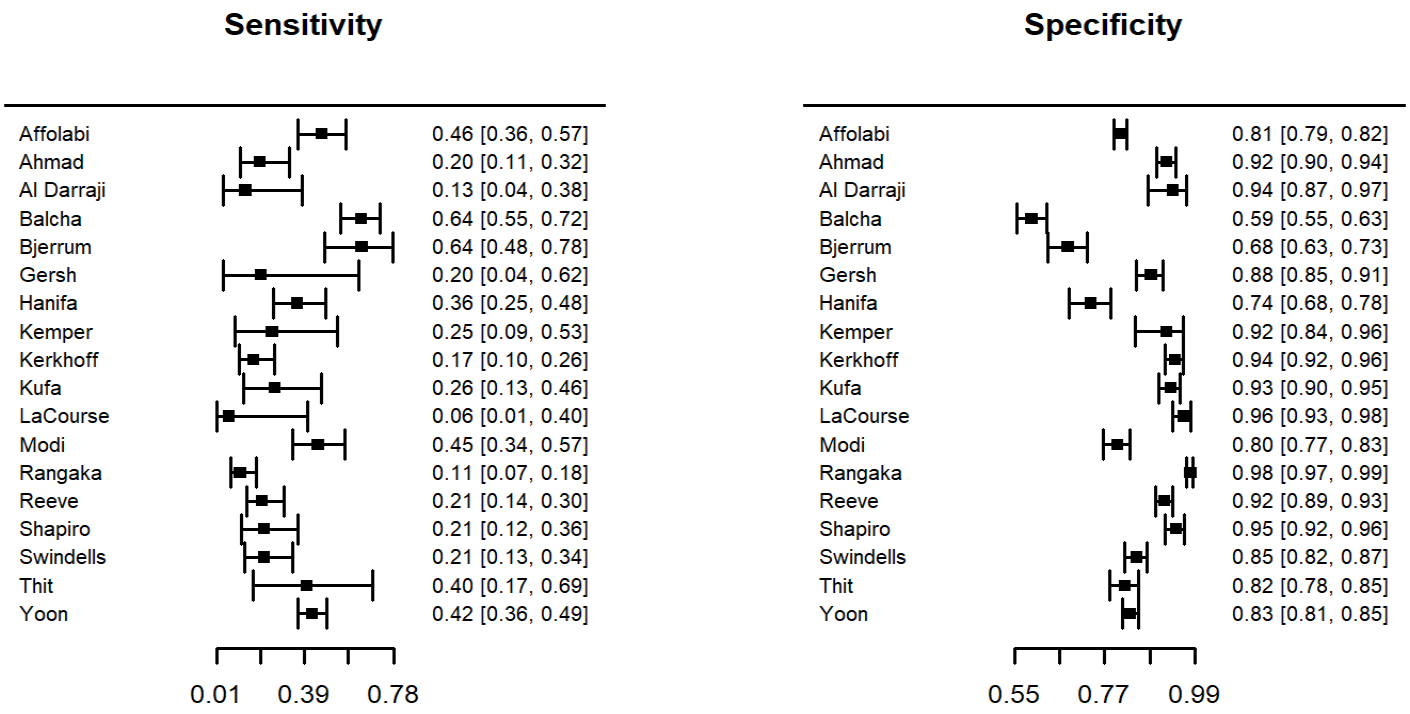




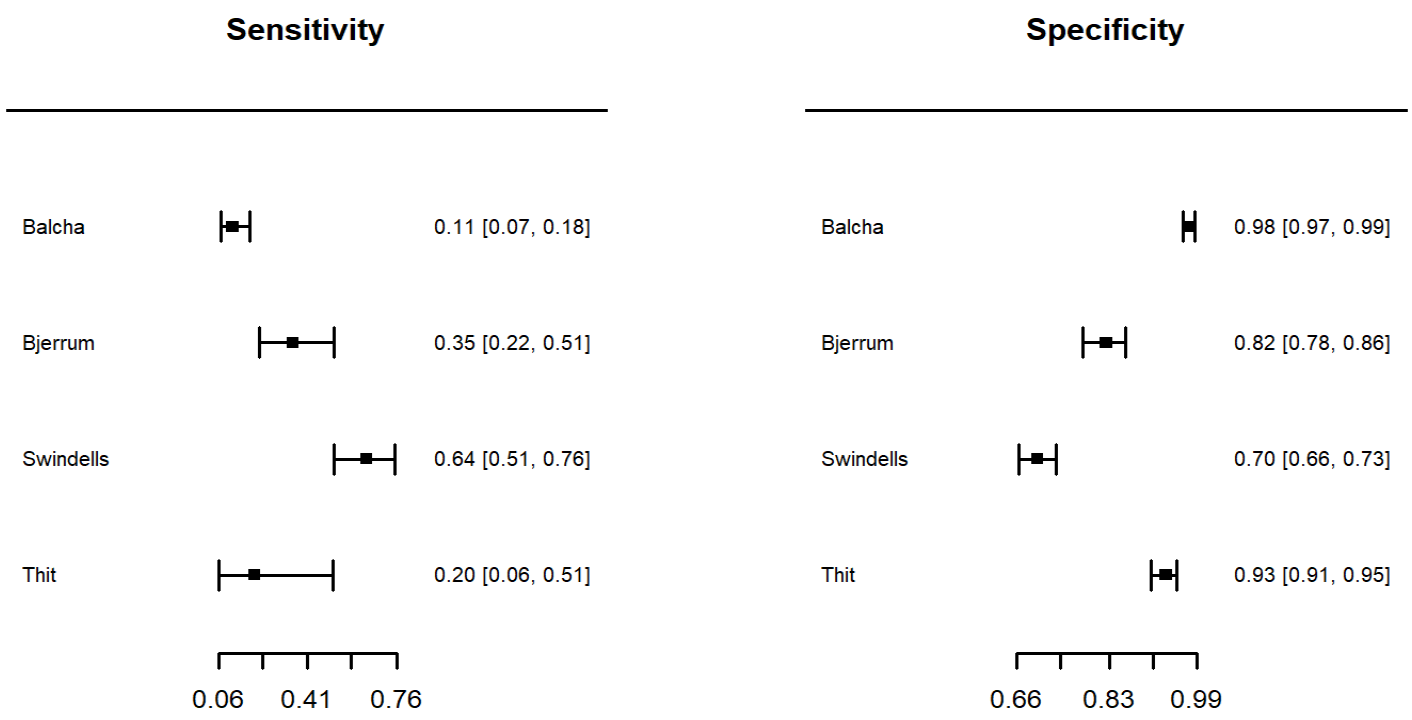




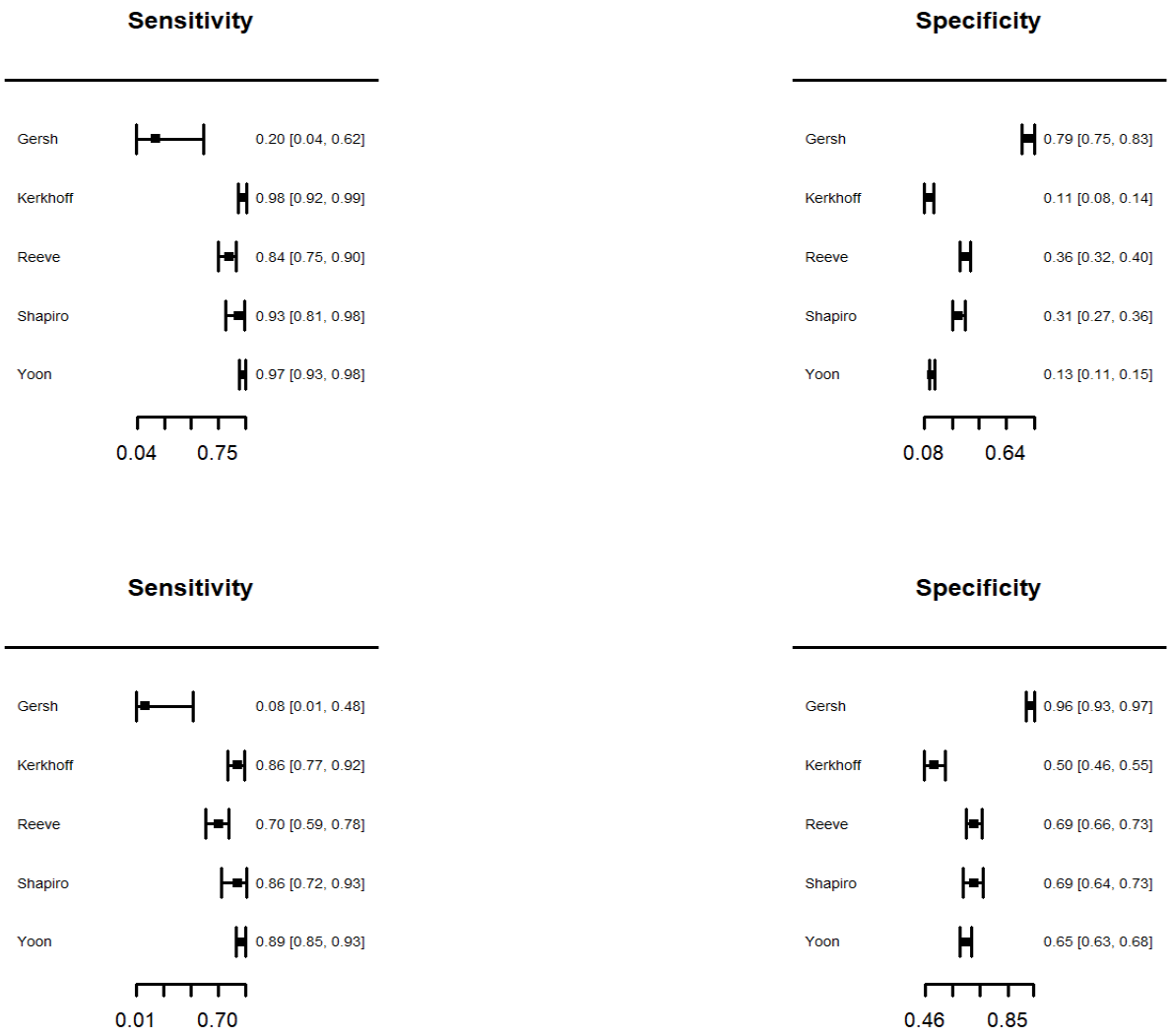
All - Forest plot for
BMI (<18.5 kg/m²)



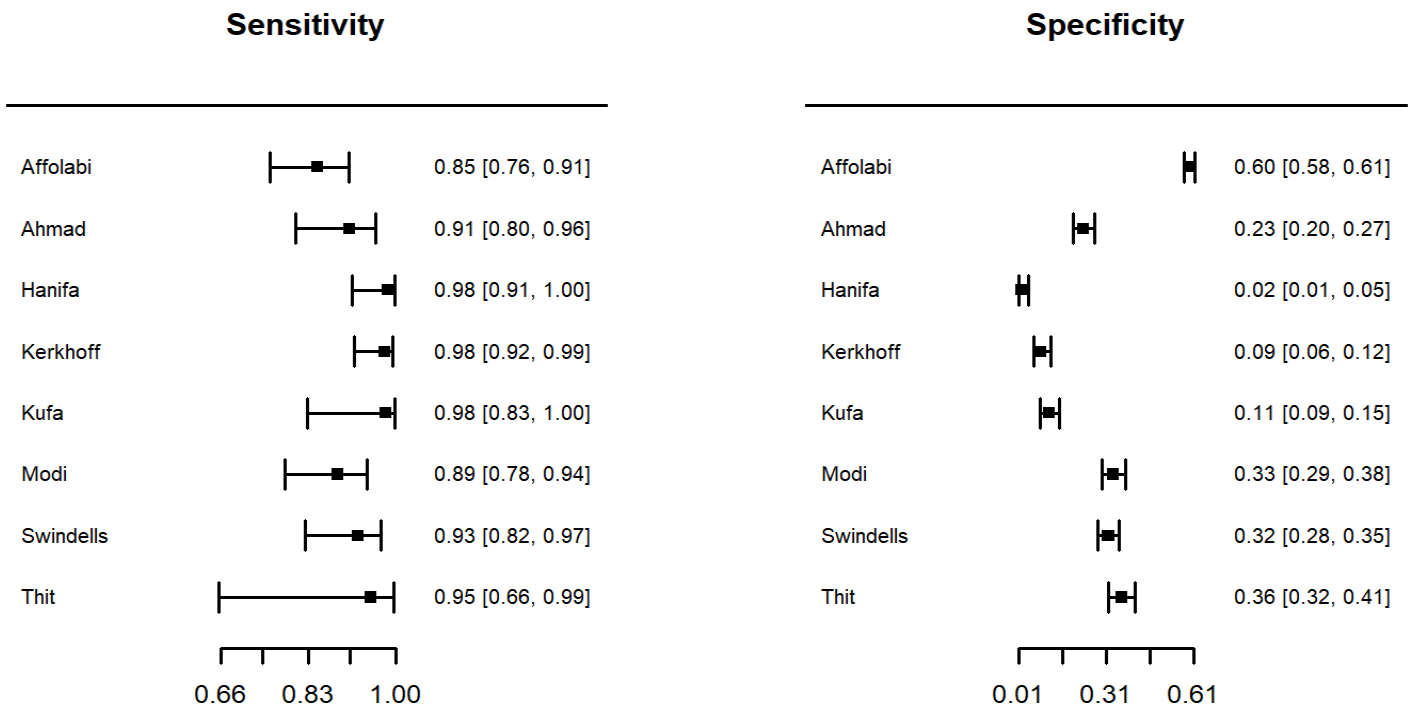
All - Forest plot for
Lymphadenopathy



All - Forest plot for
Top: W4SS with CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)



All - Forest plot for
W4SS with CXR (abnormal)



All - Forest plot for
Top: W4SS then Xpert and Bottom: Xpert alone

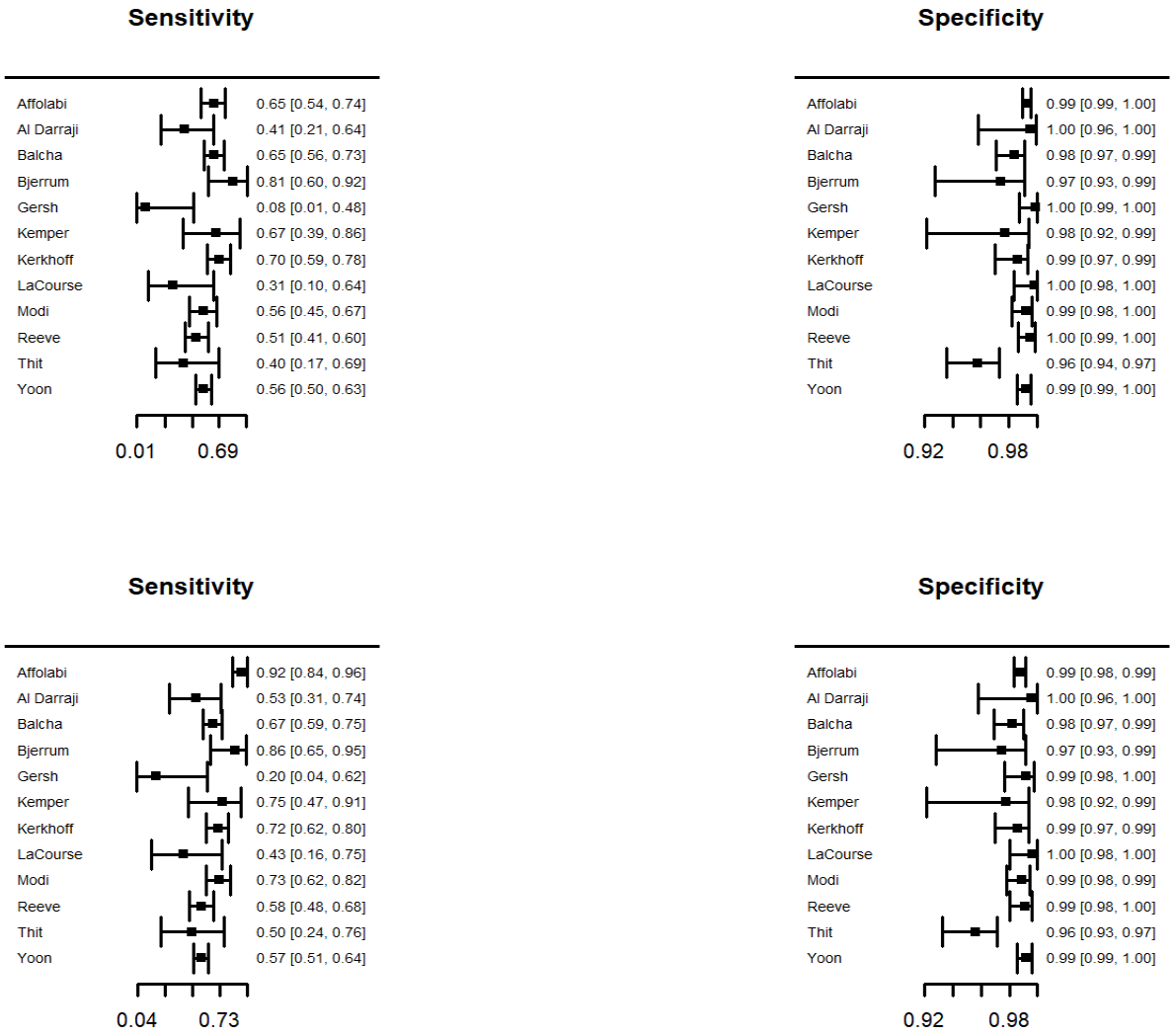
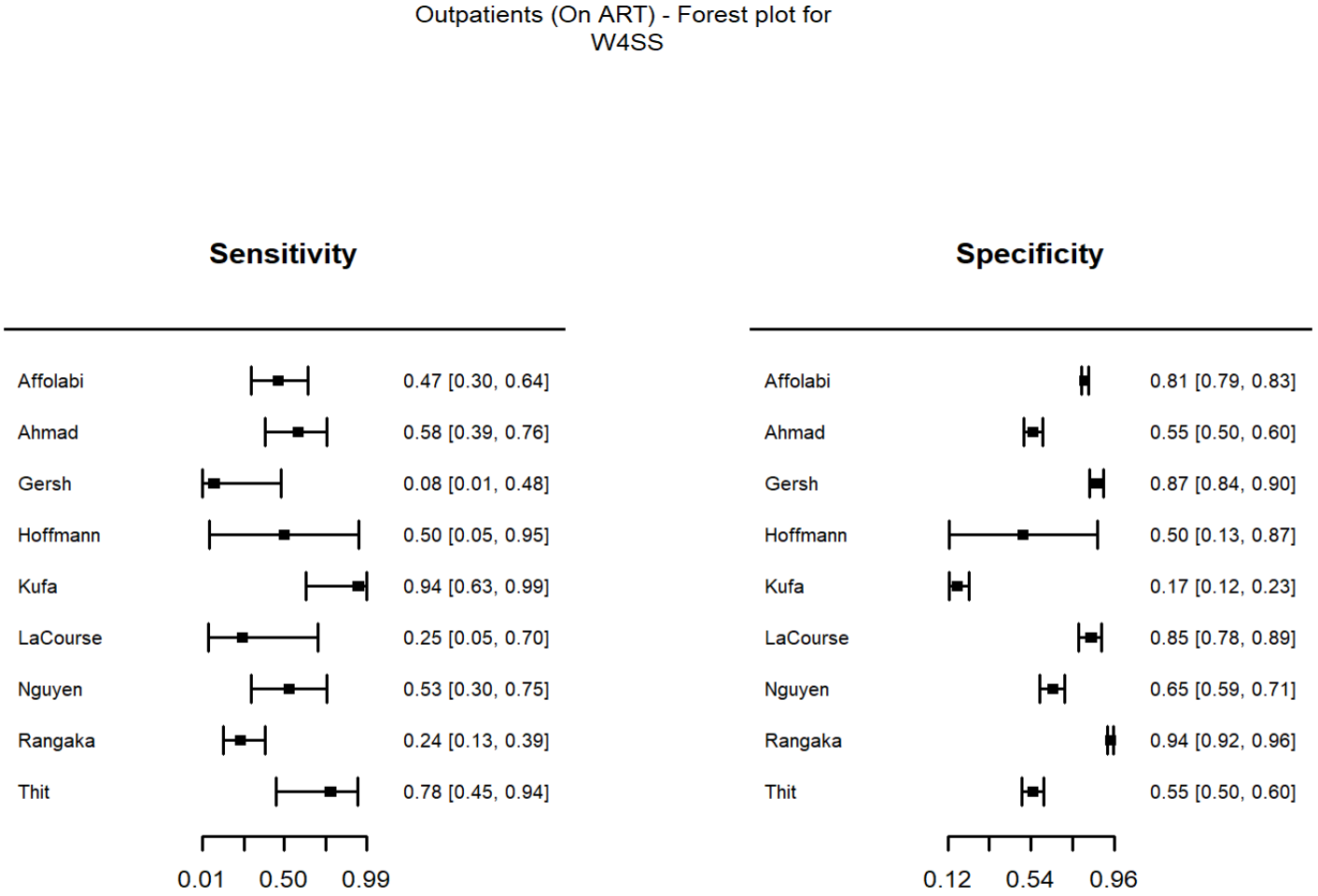
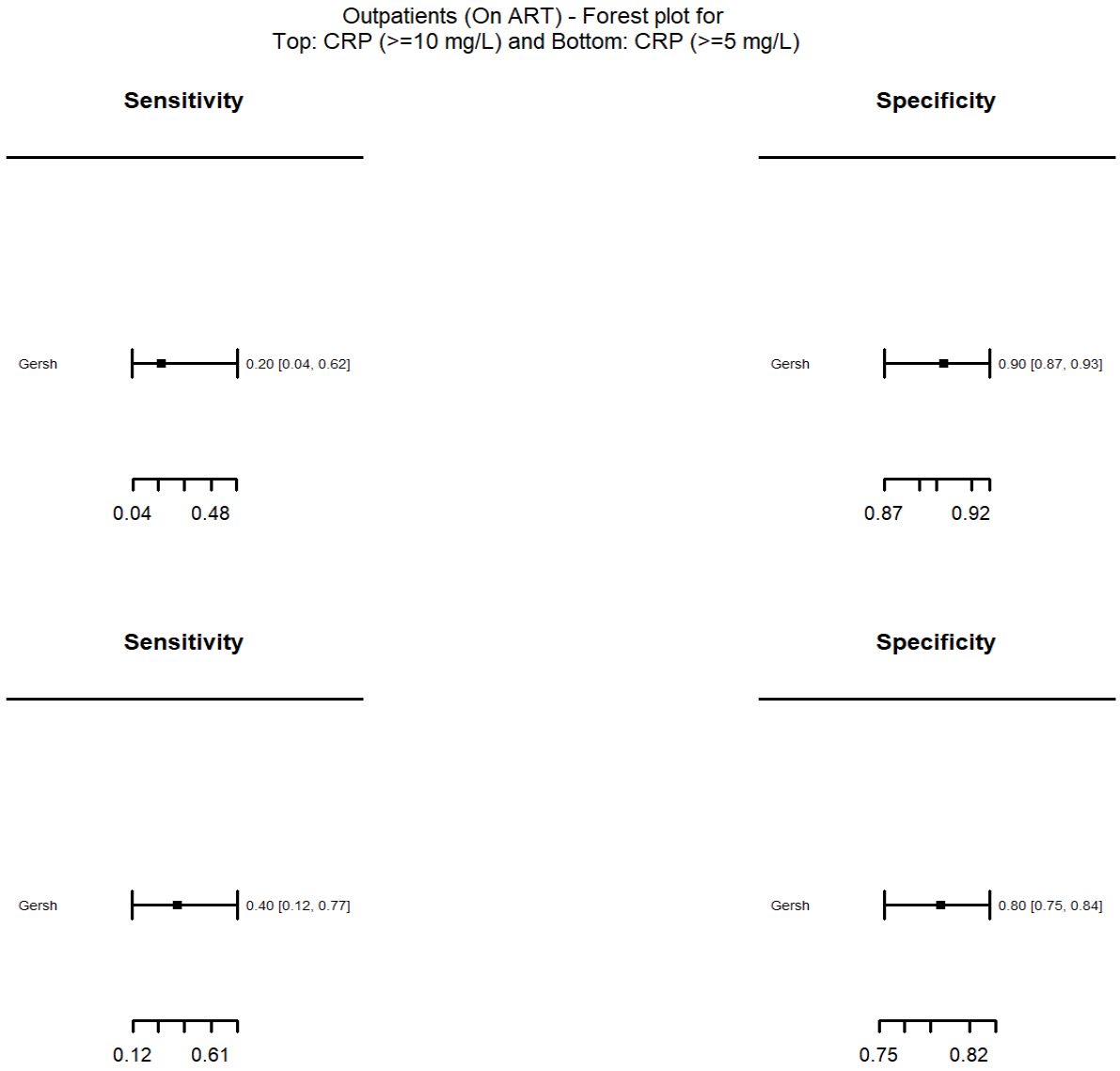
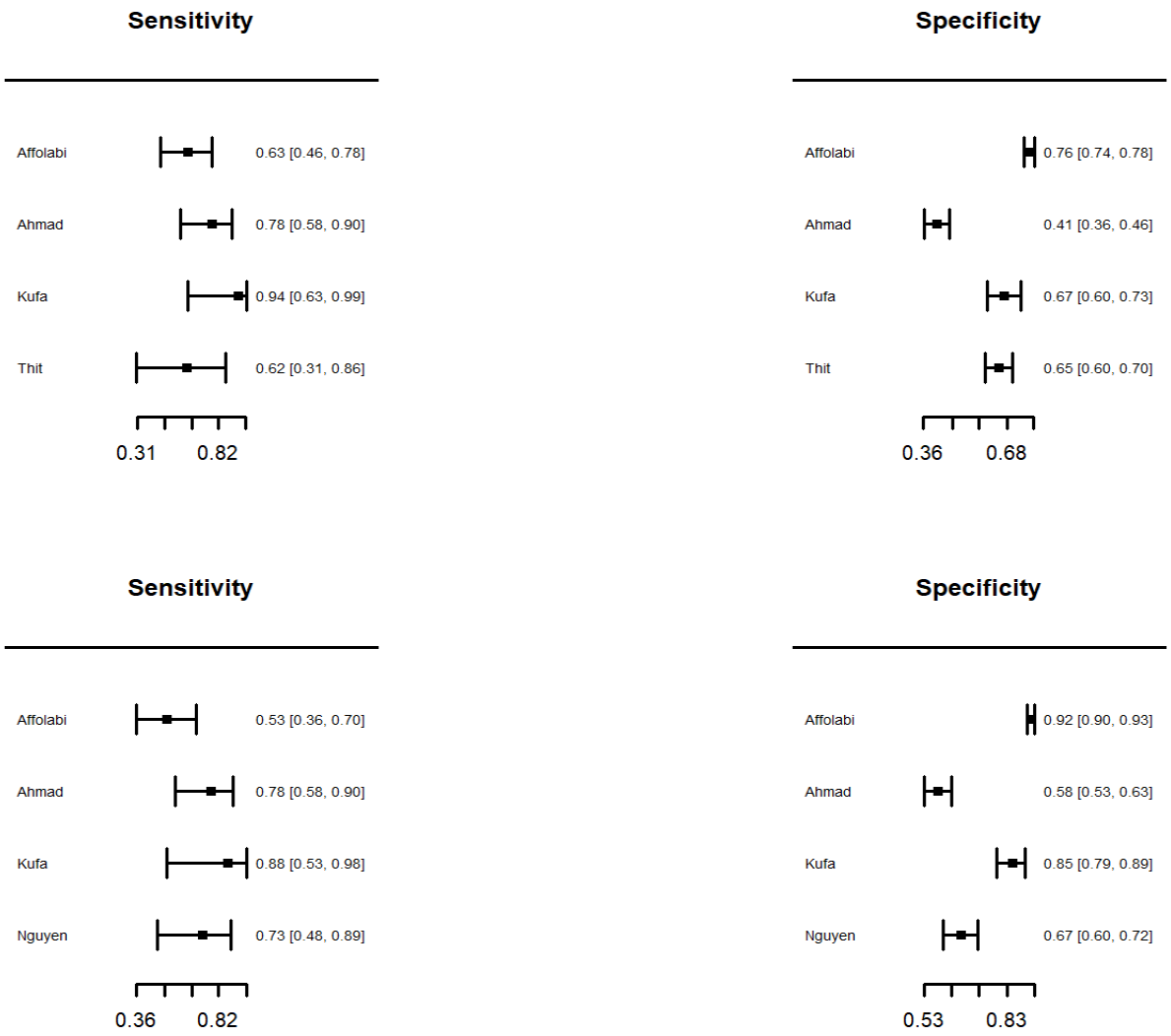


Figure 8-3B - Forest plots of sensitivity and specificity estimates in outpatients (on ART)

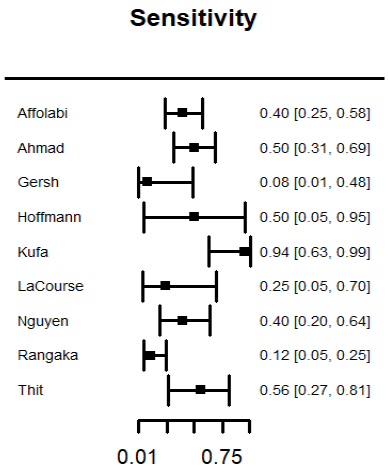


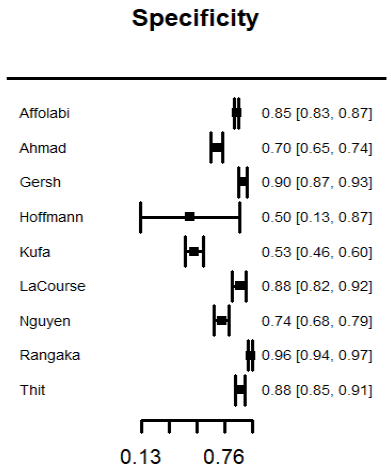

Outpatients (On ART) - Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)
Sensitivity
Gersh
0.20 [0.04, 0.62]
0.04
0.48
Specificity
Gersh
0.90 [0.87, 0.93]
0.87
0.92
Sensitivity
Gersh
0.40 [0.12, 0.77]
0.12
0.61
Specificity
Gersh
0.80 [0.75, 0.84]
0.75
0.82

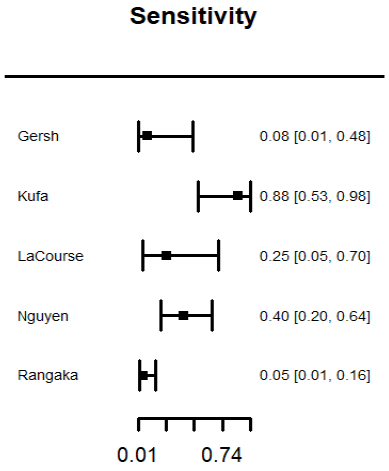
Outpatients (On ART) - Forest plot for
Top: CXR (abnormal) and Bottom: CXR (suggests tuberculosis)

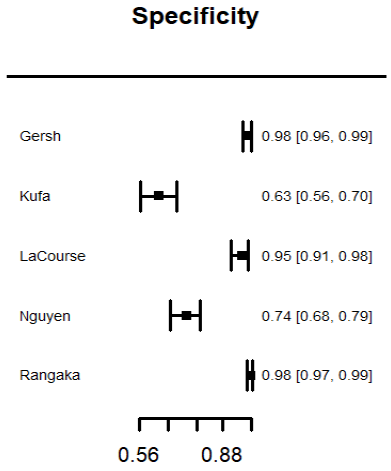


Outpatients (On ART) - Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

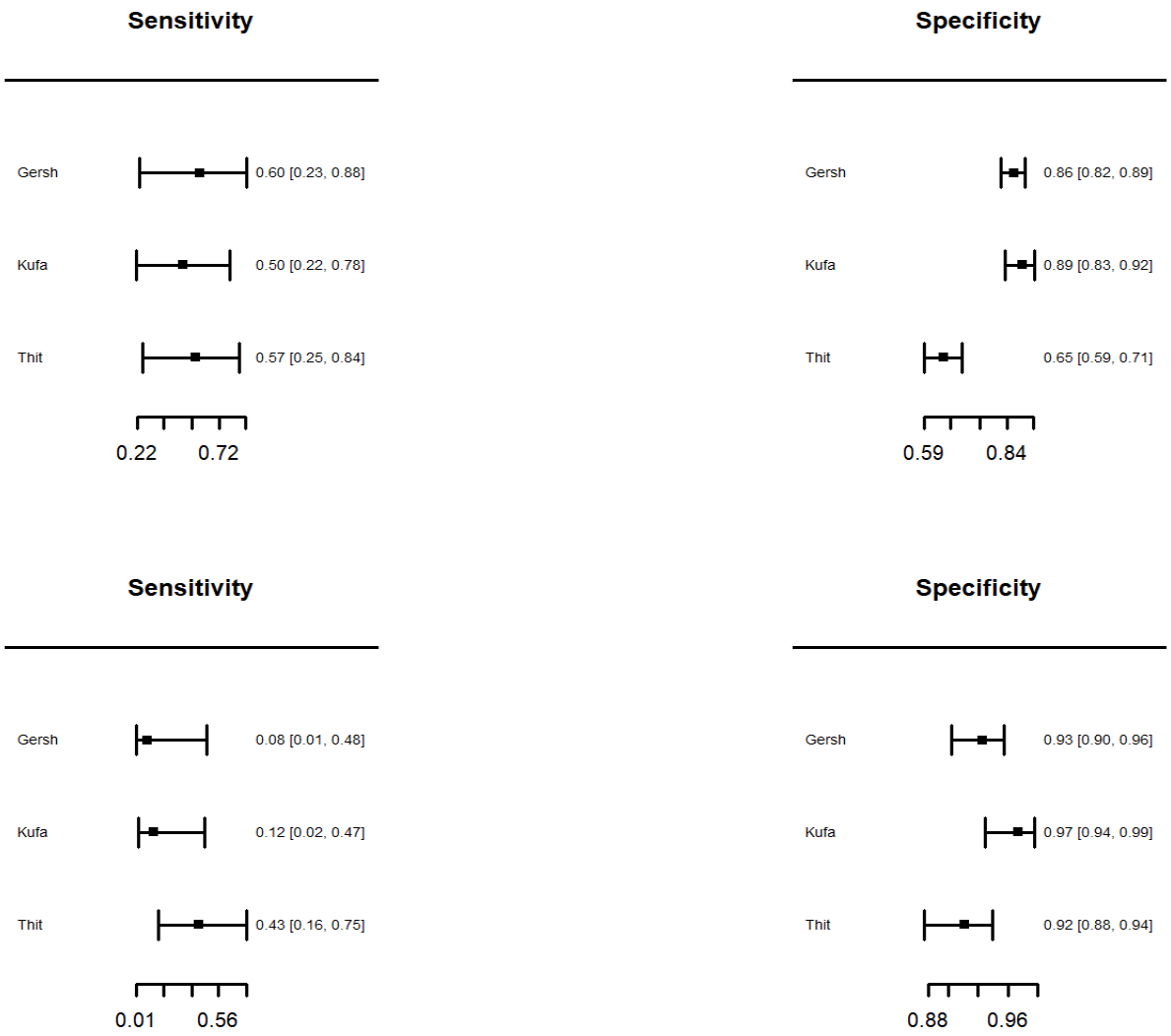




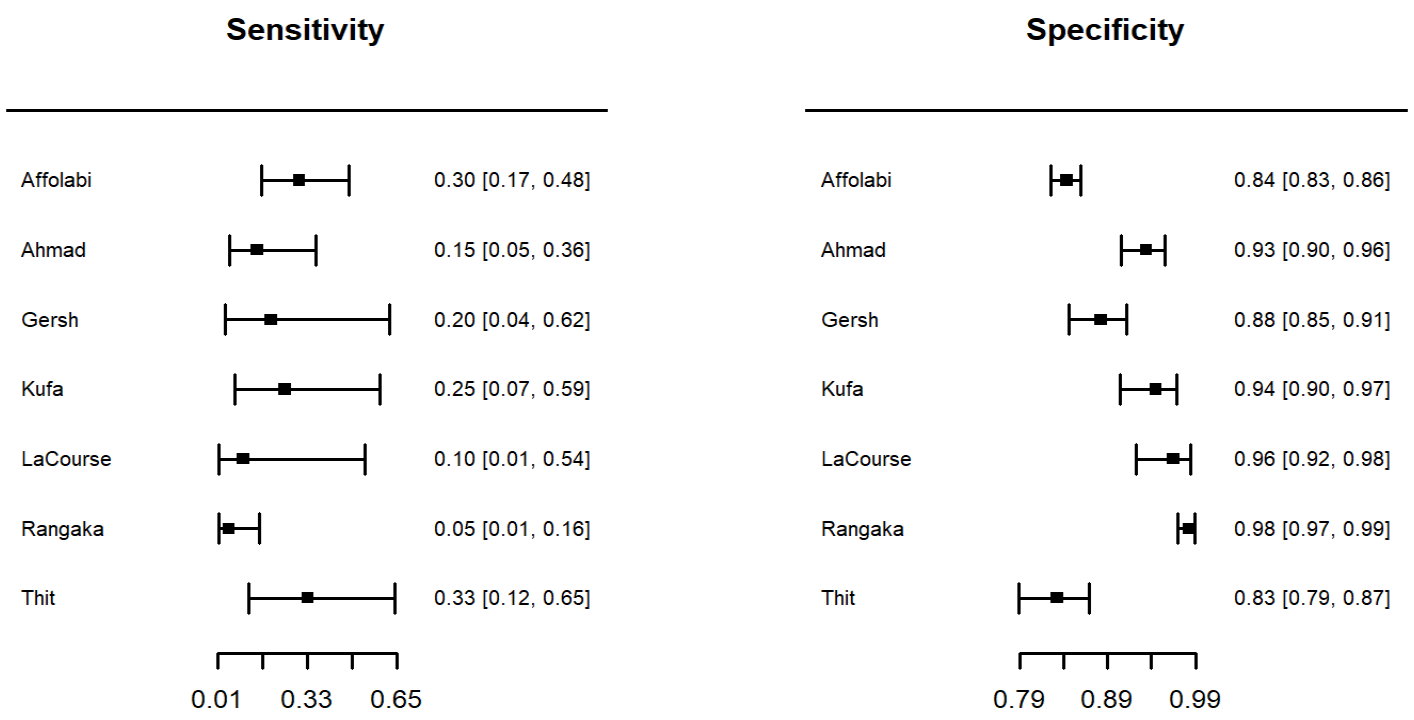




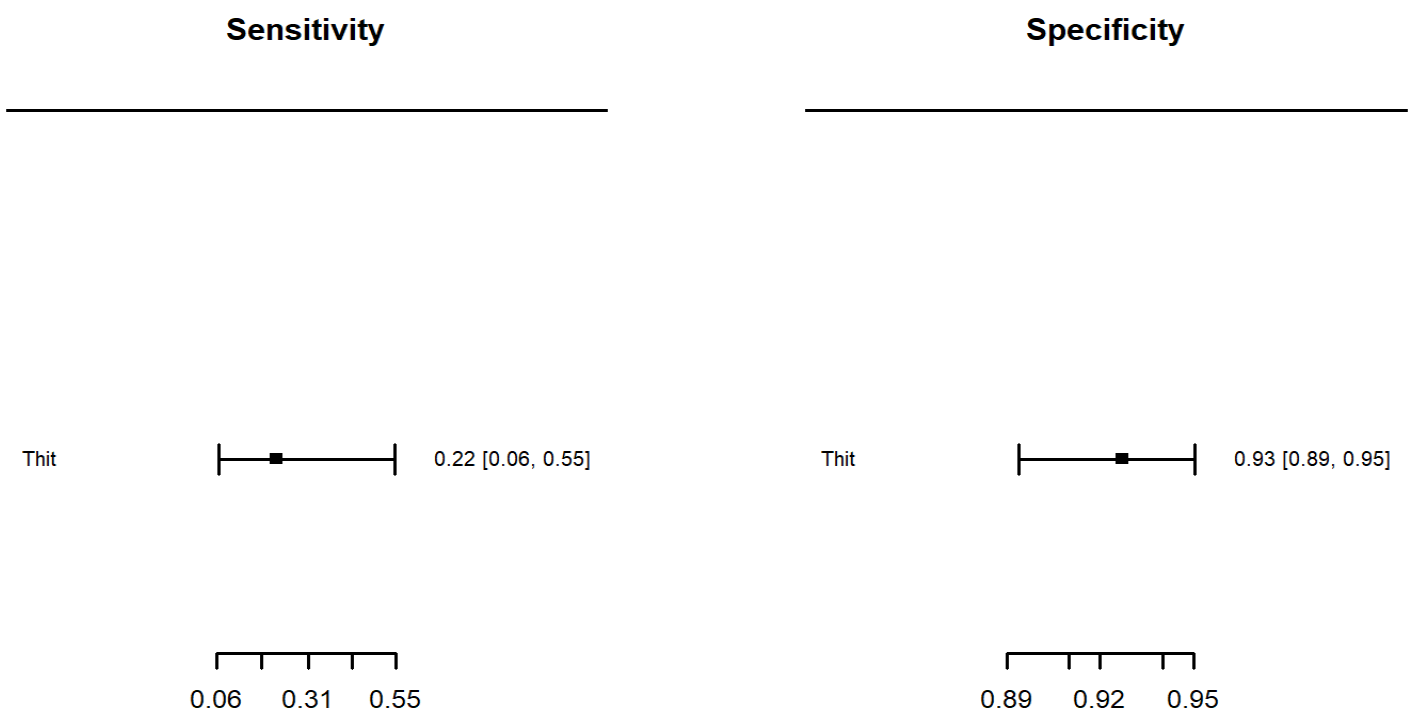
Outpatients (On ART) - Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)



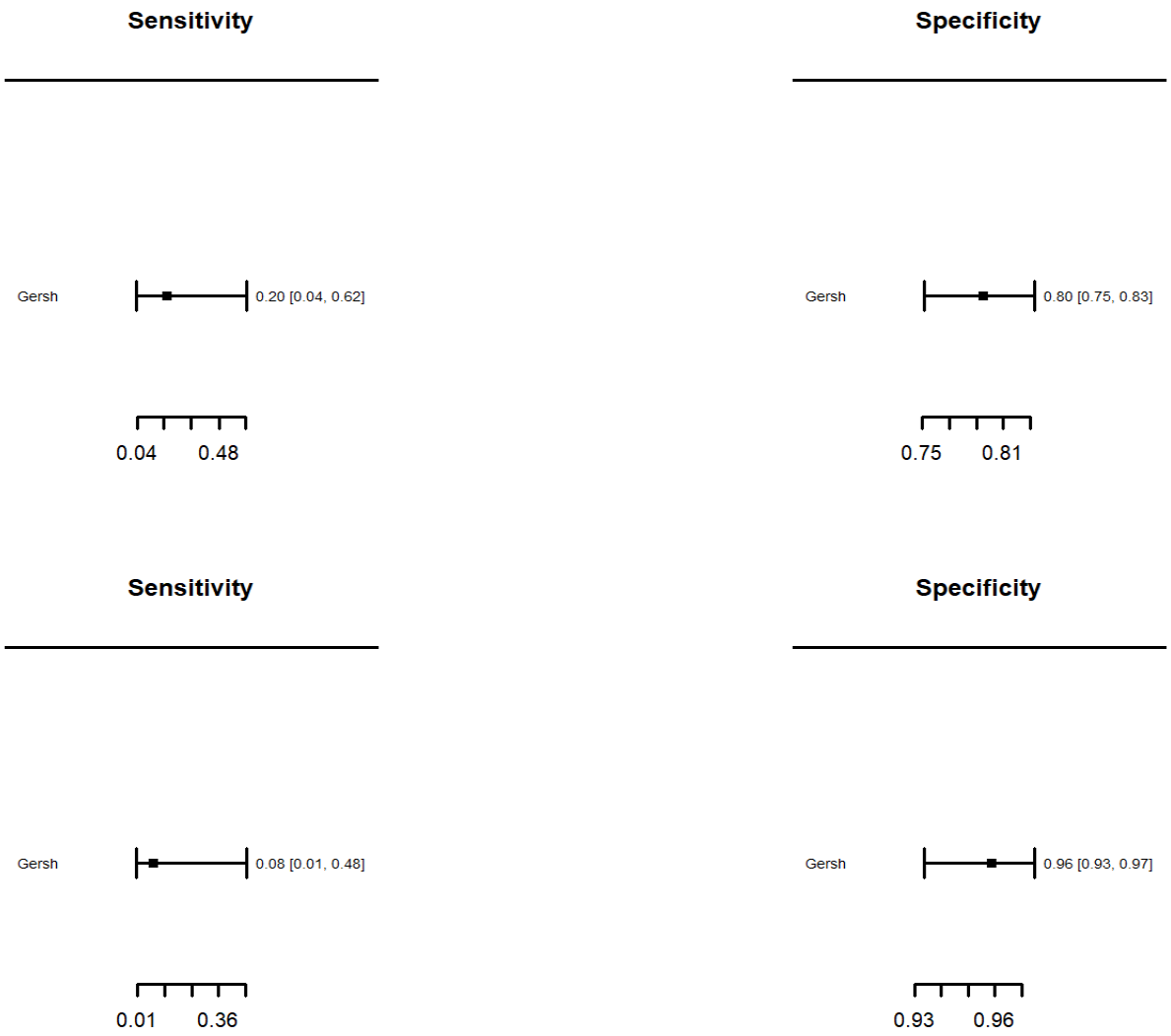
Outpatients (On ART) - Forest plot for
BMI (<18.5 kg/m²)



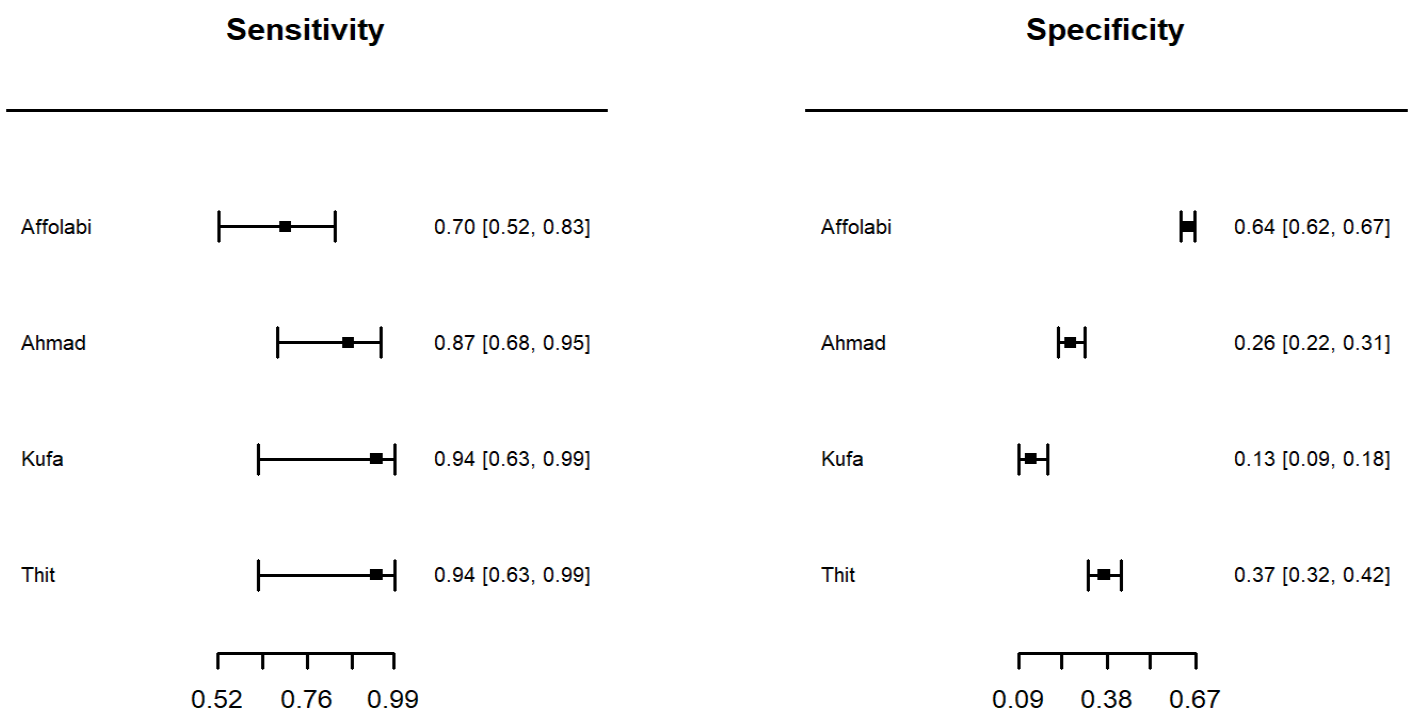
Outpatients (On ART) - Forest plot for
Lymphadenopathy



Outpatients (On ART) - Forest plot for
Top: W4SS with CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)



Outpatients (On ART) - Forest plot for
W4SS with CXR (abnormal)



Outpatients (On ART) - Forest plot for
Top: W4SS then Xpert and Bottom: Xpert alone

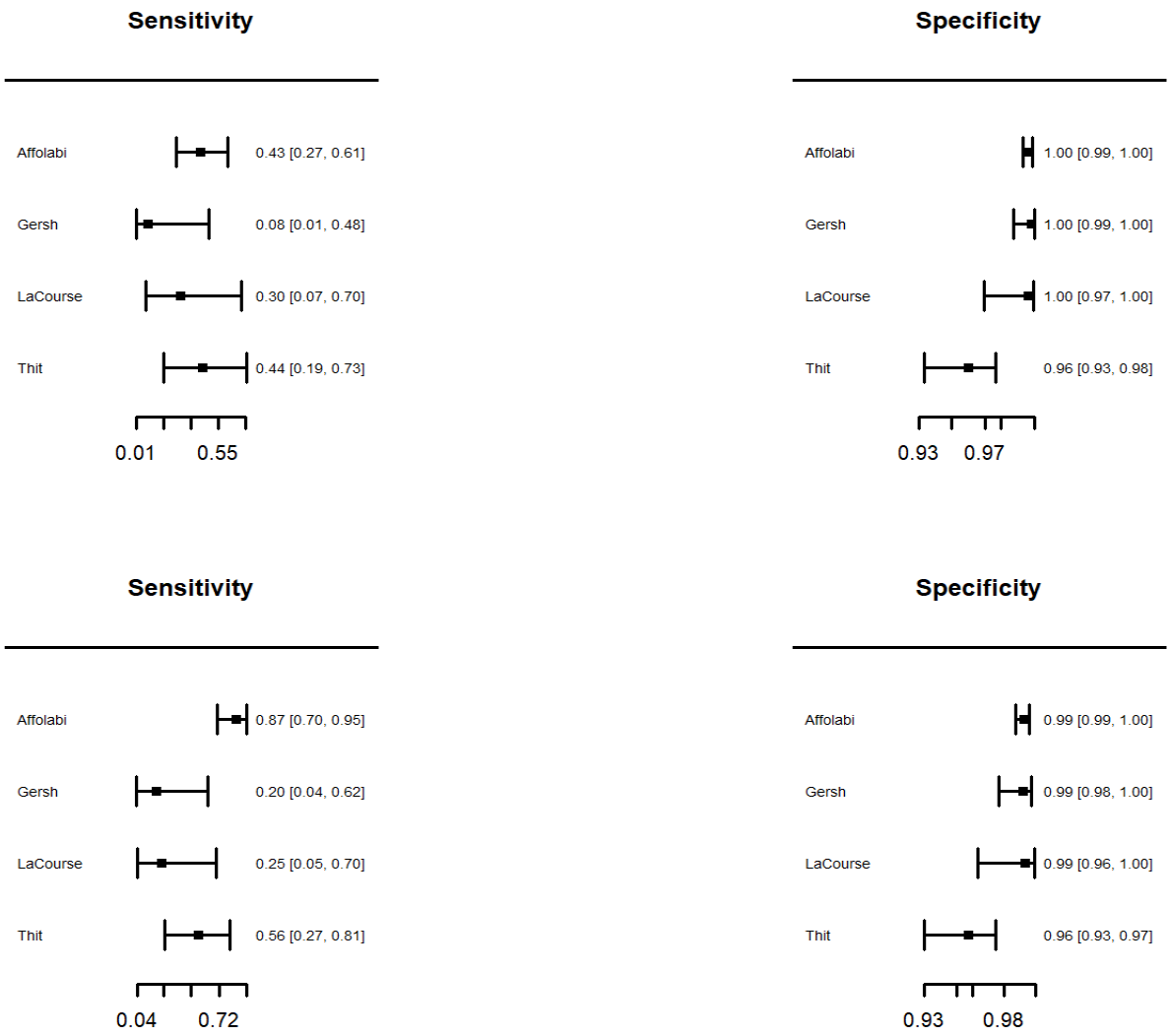
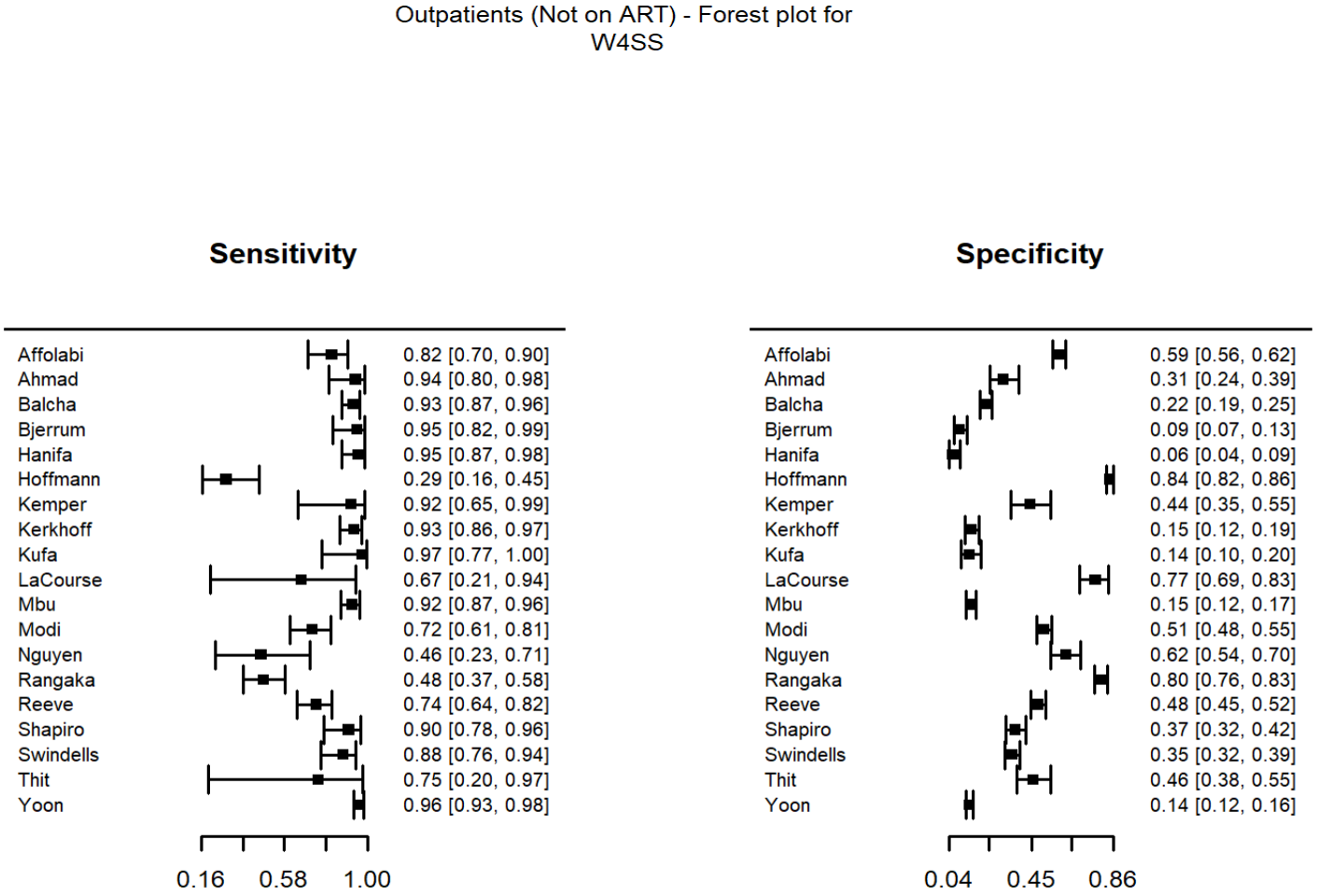
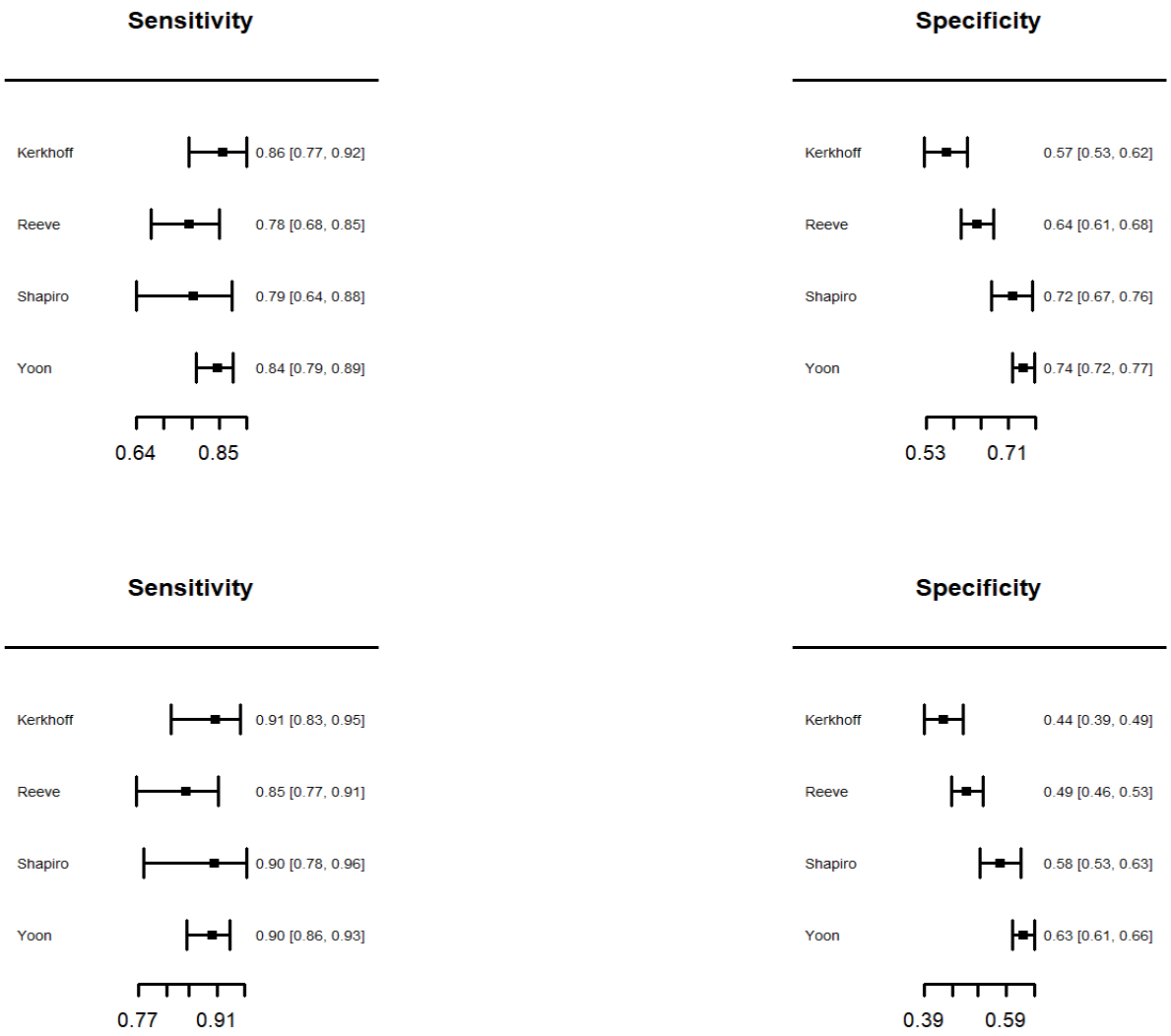


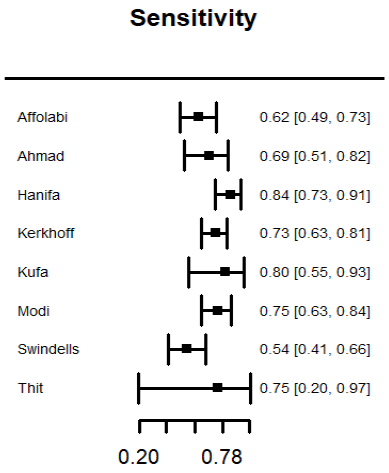
Figure 8-3C - Forest plots of sensitivity and specificity estimates in outpatients (not on ART)

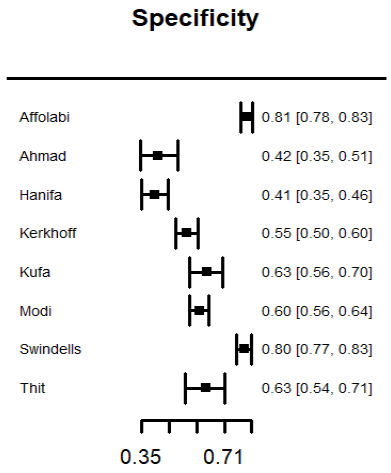


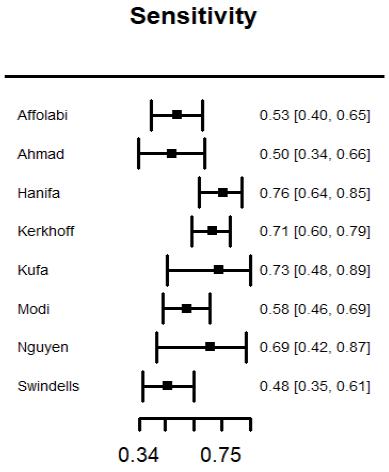
Outpatients (Not on ART) - Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)

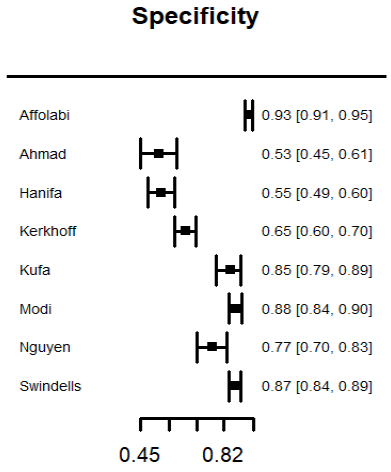


Outpatients (Not on ART) - Forest plot for
Top: CXR (abnormal) and Bottom: CXR (suggests tuberculosis)

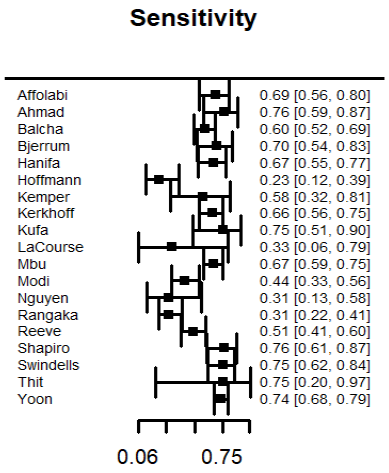


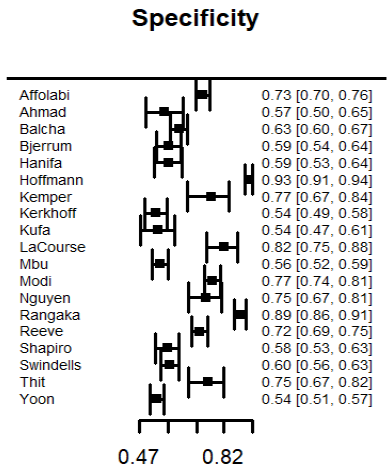


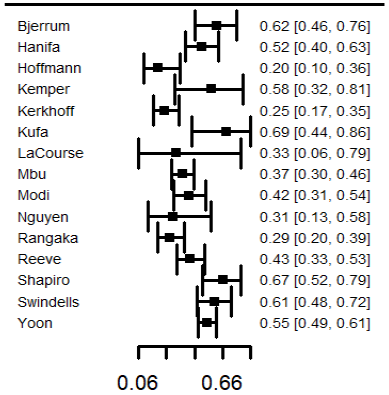


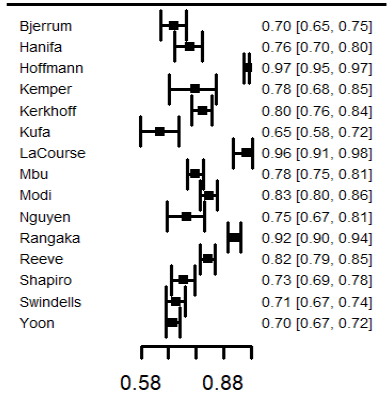


Outpatients (Not on ART) - Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

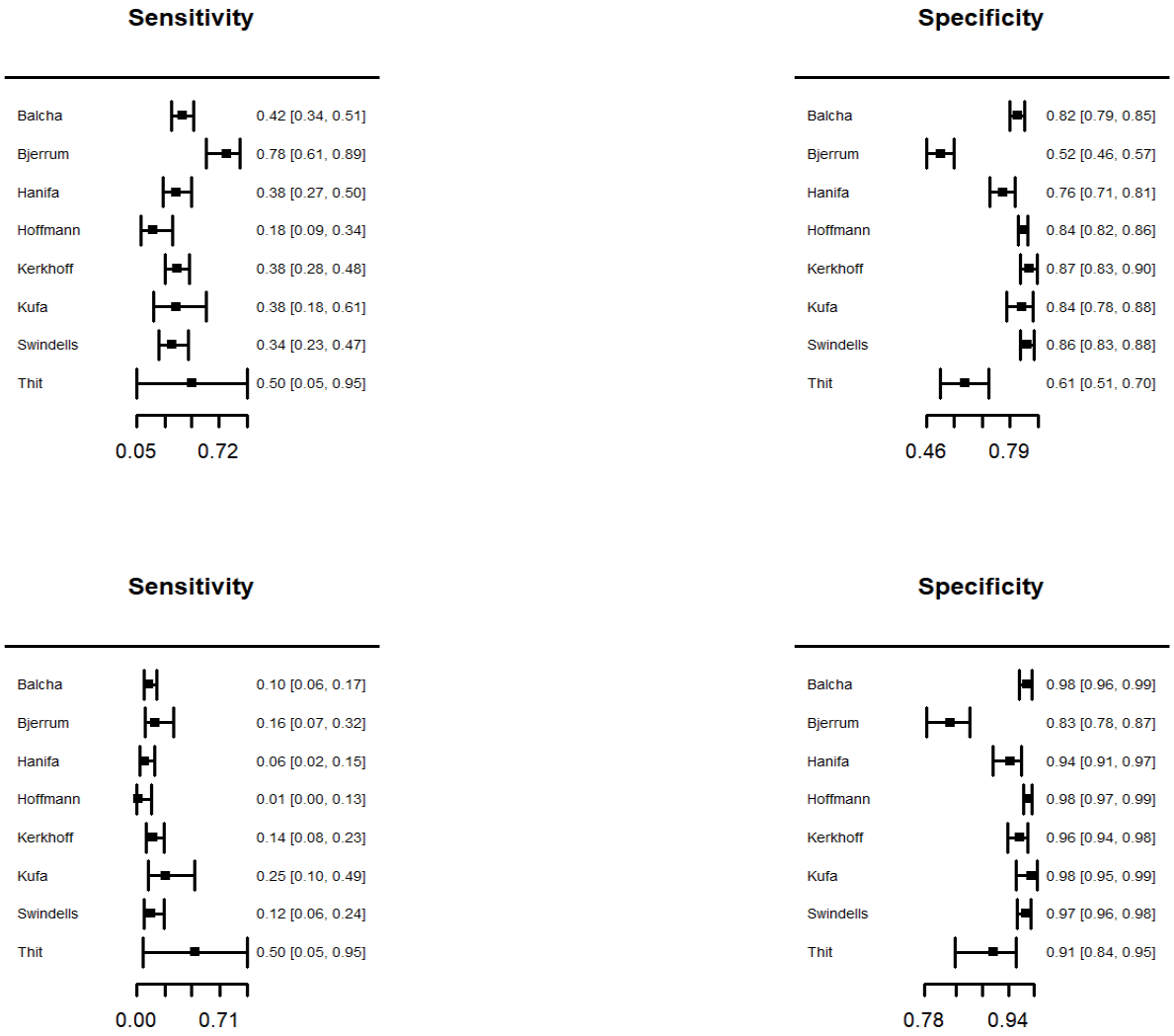




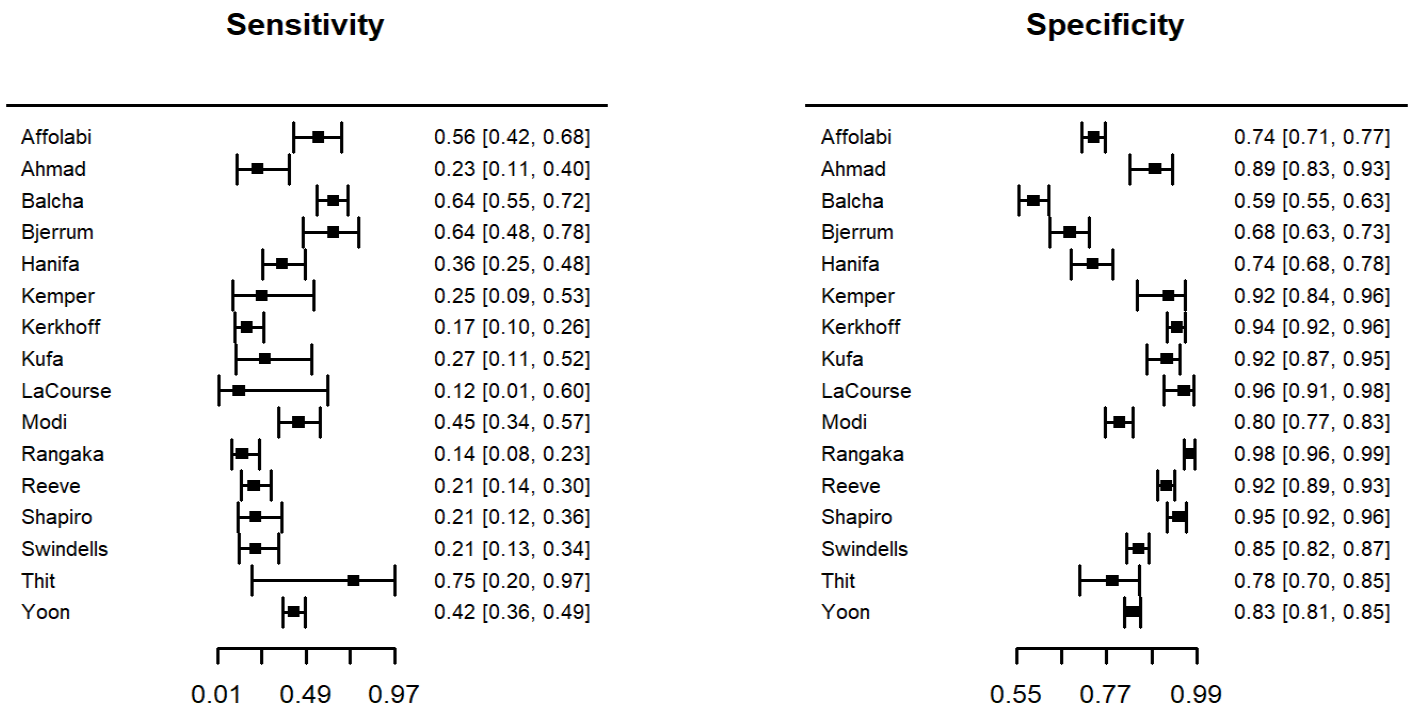




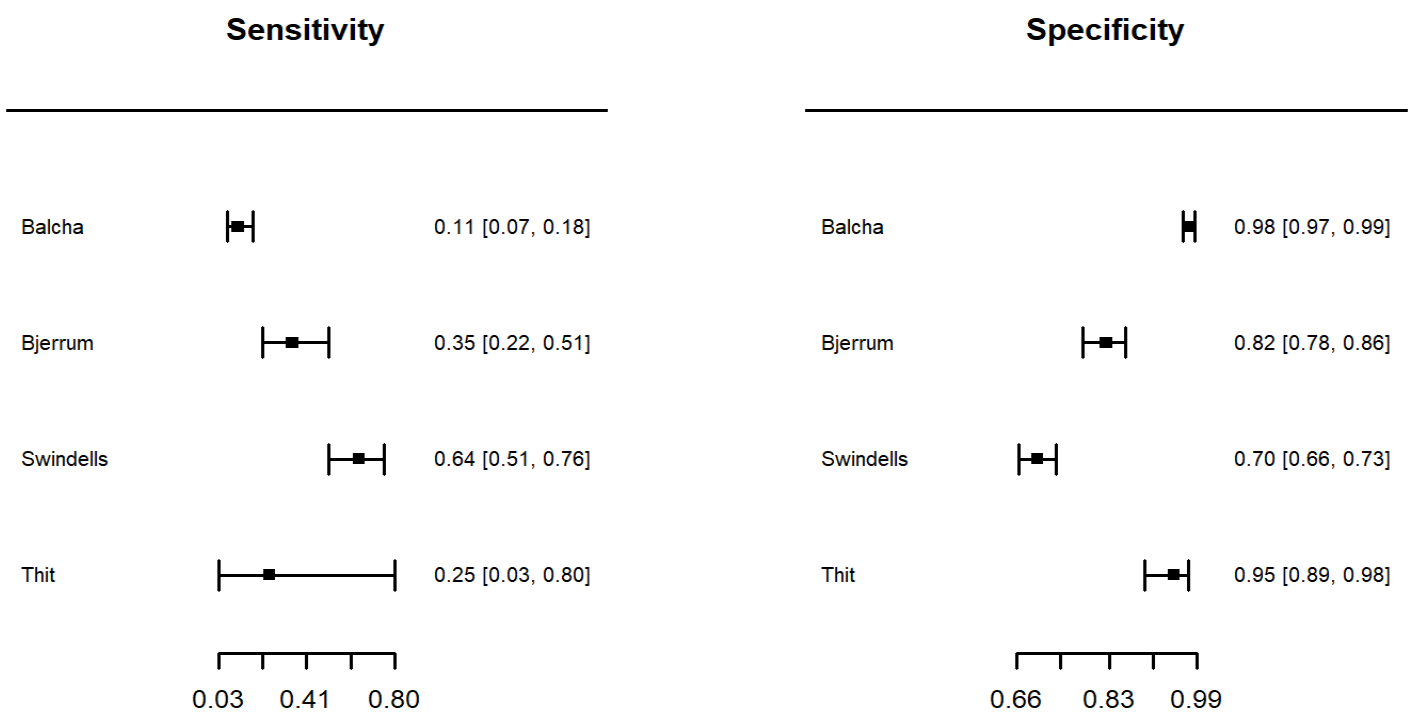
Outpatients (Not on ART) - Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)



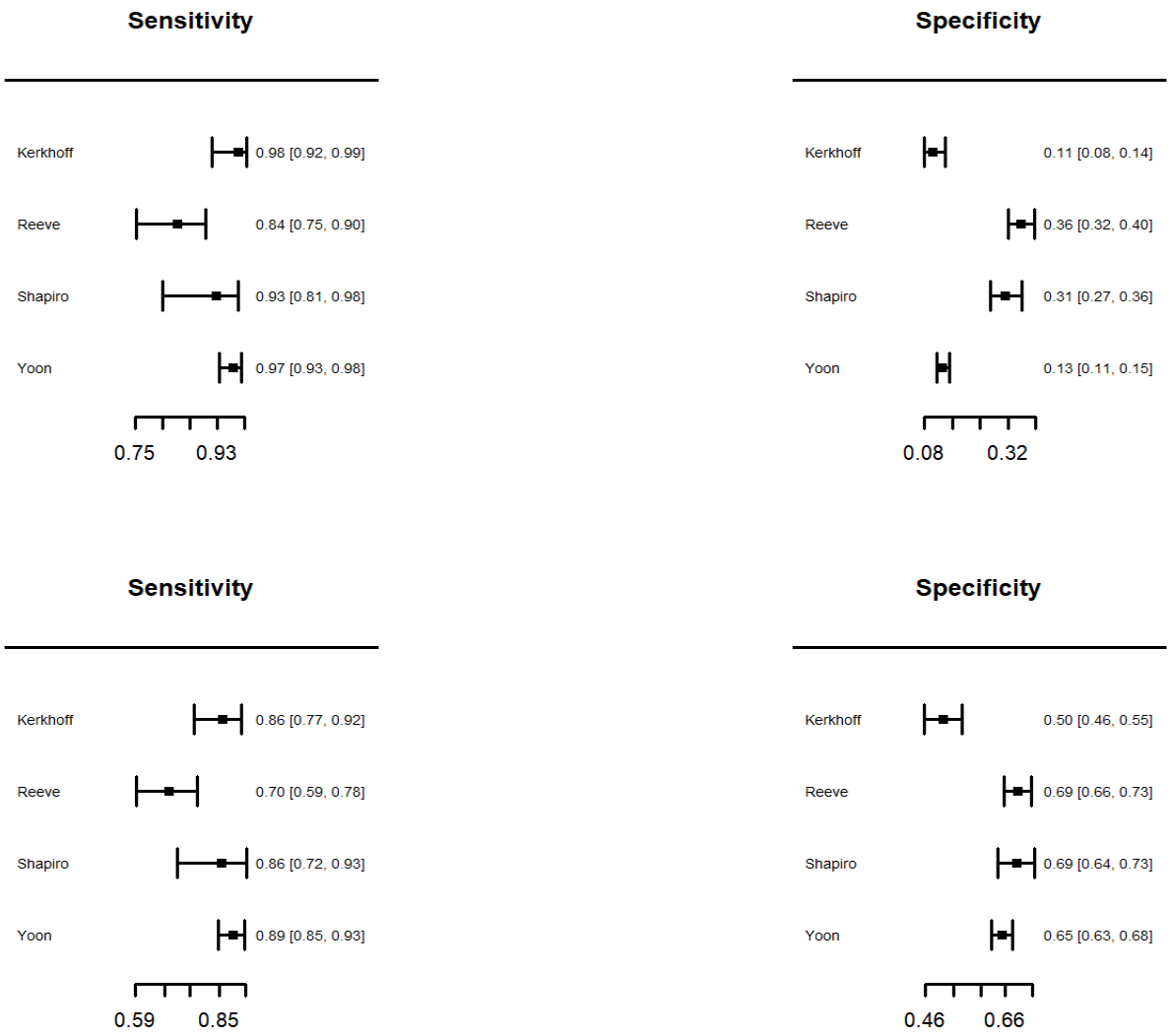
Outpatients (Not on ART) - Forest plot for
BMI (<18.5 kg/m²)



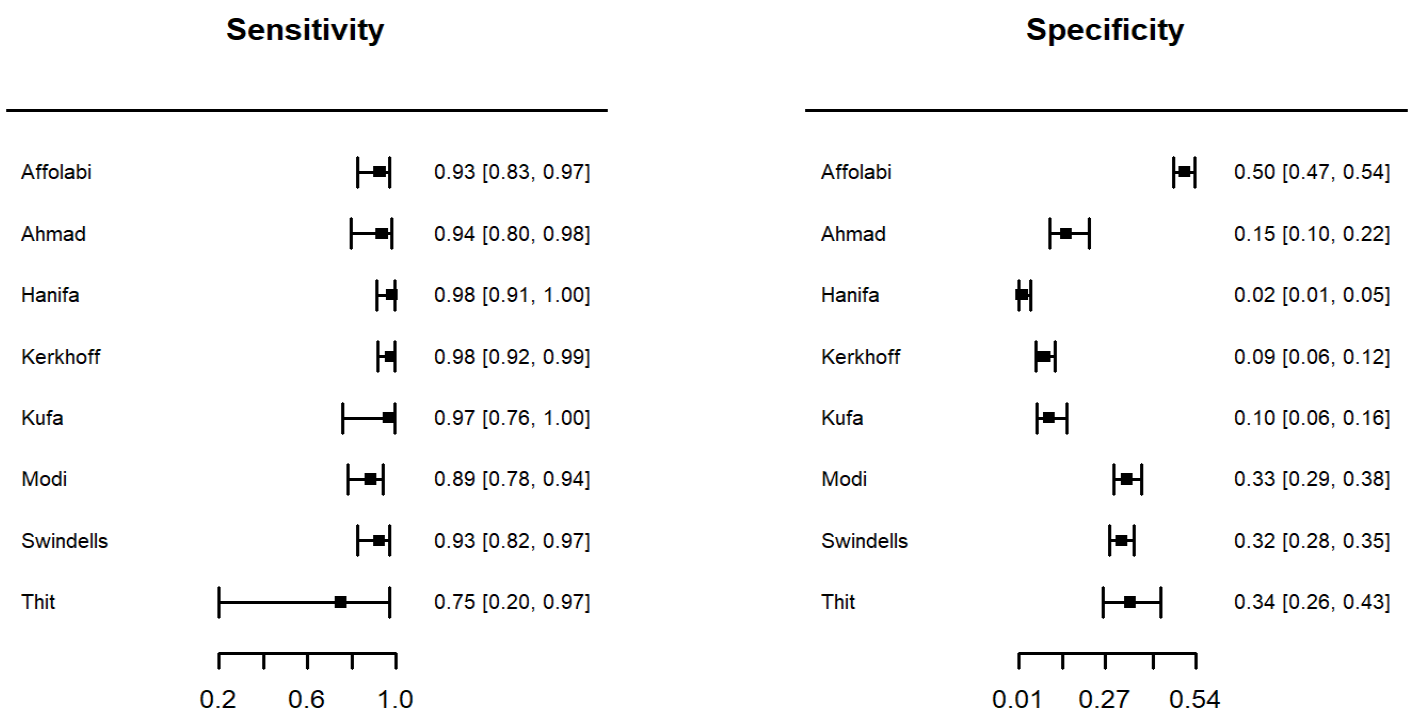
Outpatients (Not on ART) - Forest plot for Lymphadenopathy



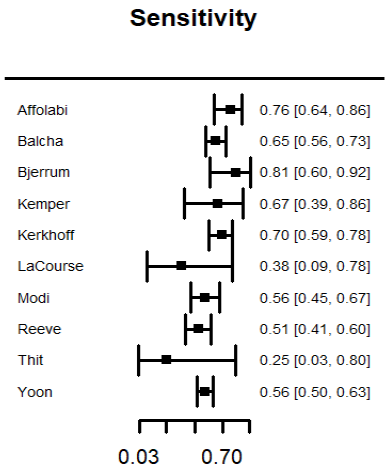
Outpatients (Not on ART) - Forest plot for
Top: W4SS with CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)

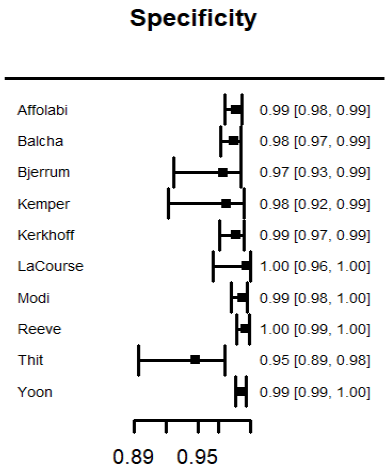


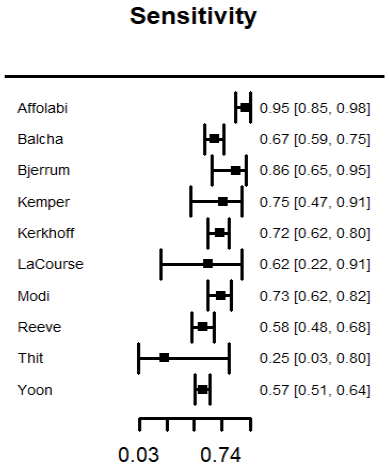
Outpatients (Not on ART) - Forest plot for
W4SS with CXR (abnormal)



Outpatients (Not on ART) - Forest plot for
Top: W4SS then Xpert and Bottom: Xpert alone







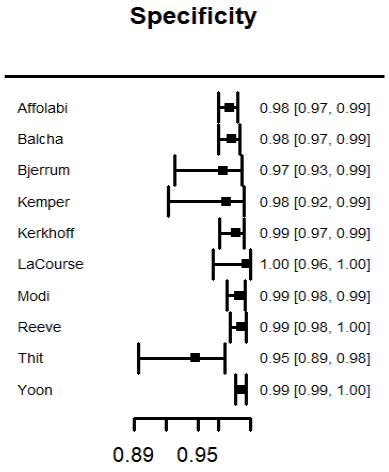
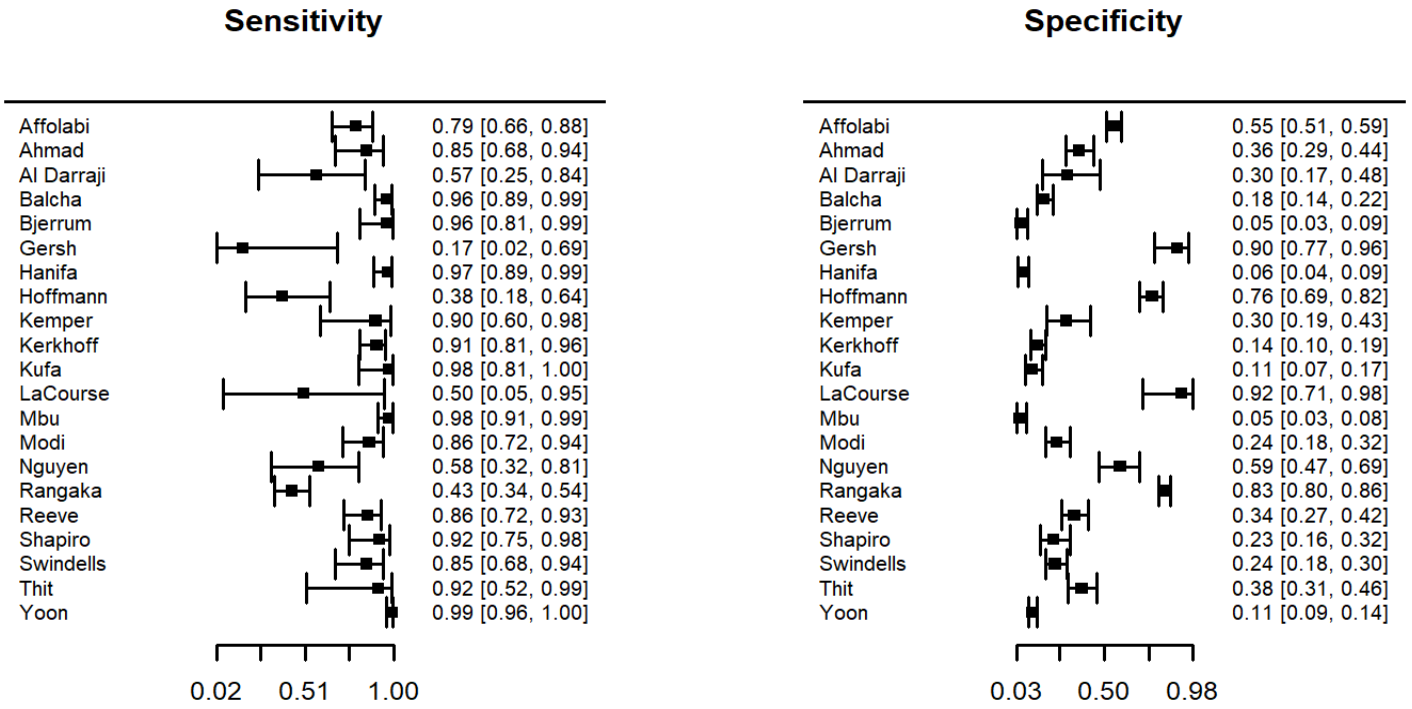
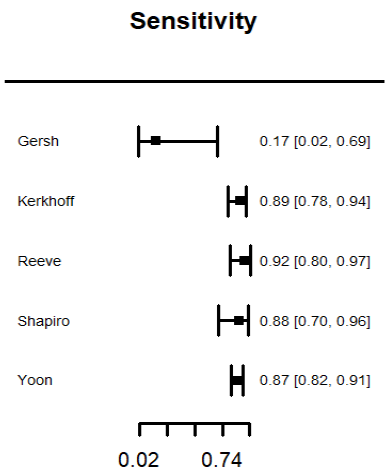


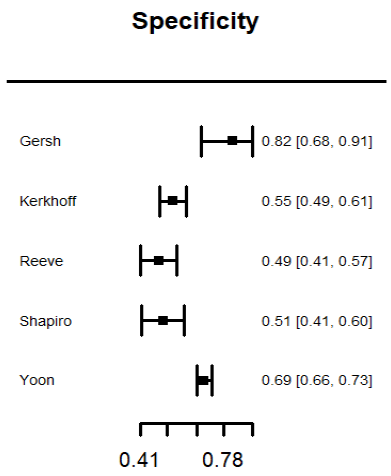
Figure 8-3D - Forest plots of sensitivity and specificity estimates in participants with a CD4 cell count <=200 cells/µL

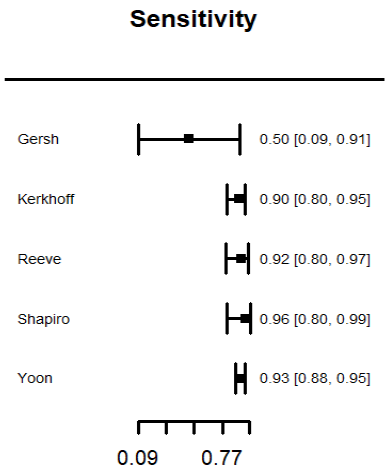
CD4 <=200 cells/µL - Forest plot for
W4SS

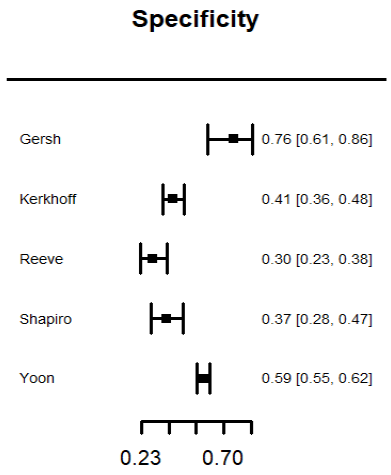


CD4 <=200 cells/μL - Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)

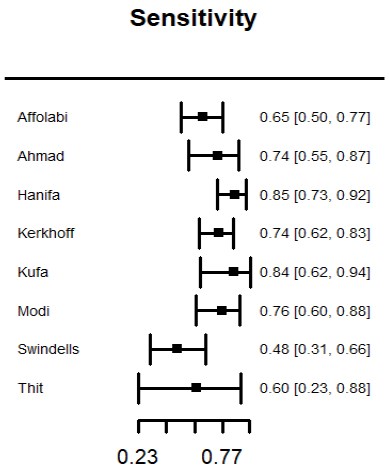


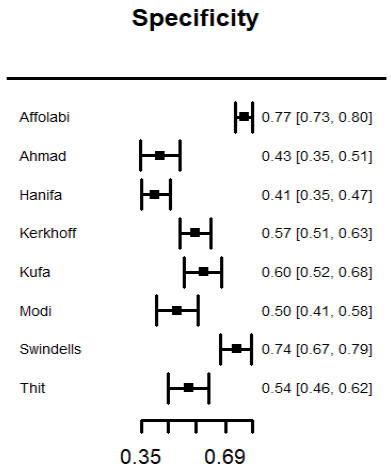


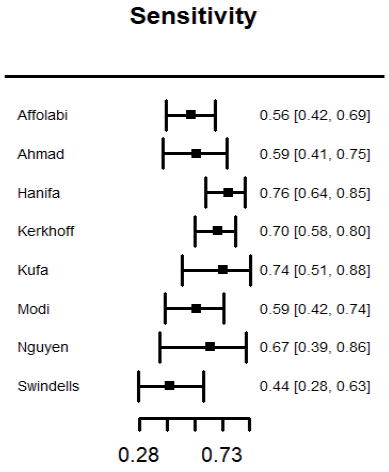


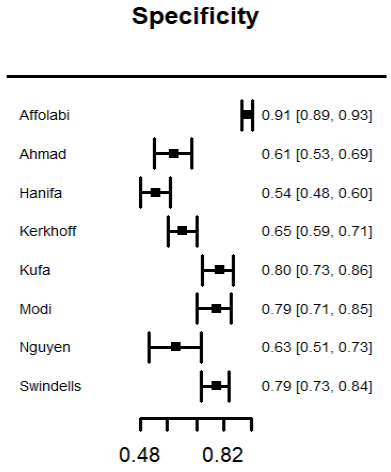


CD4 <=200 cells/μL - Forest plot for
Top: CXR (abnormal) and Bottom: CXR (suggests tuberculosis)



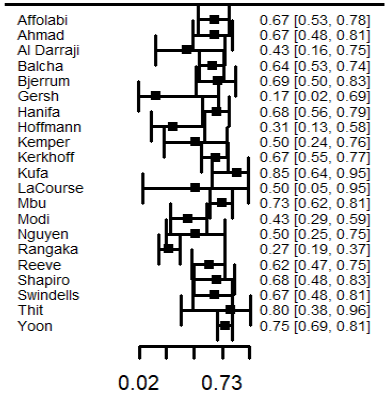




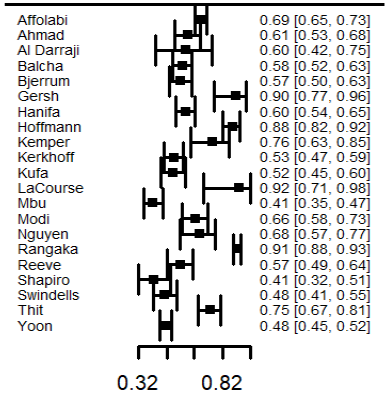


CD4 <=200 cells/µL - Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

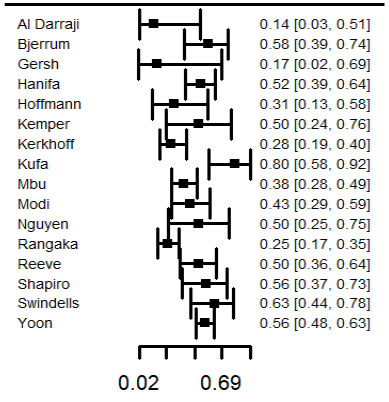
Sensitivity



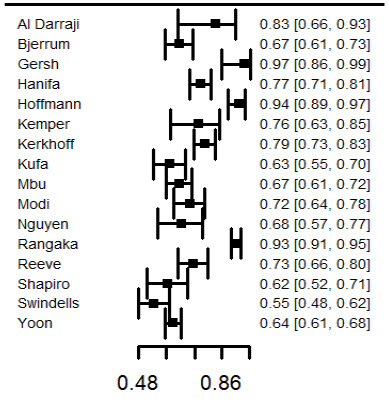
Specificity



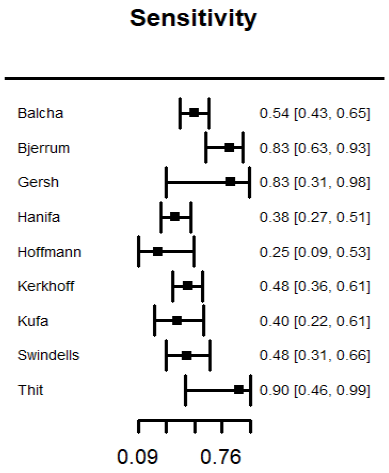
Sensitivity

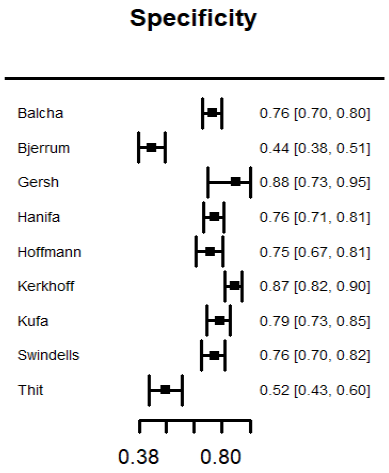


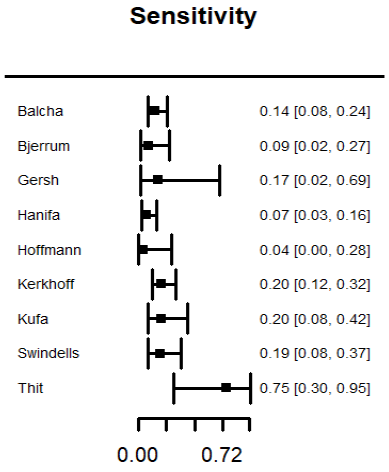
Specificity

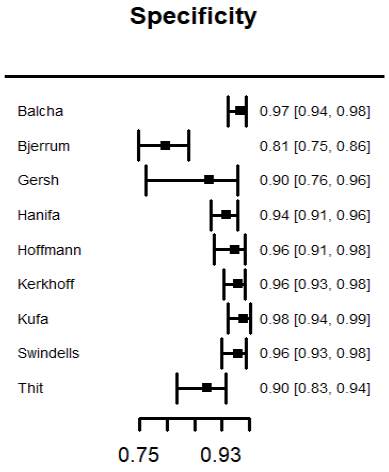


CD4 <=200 cells/µL - Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)

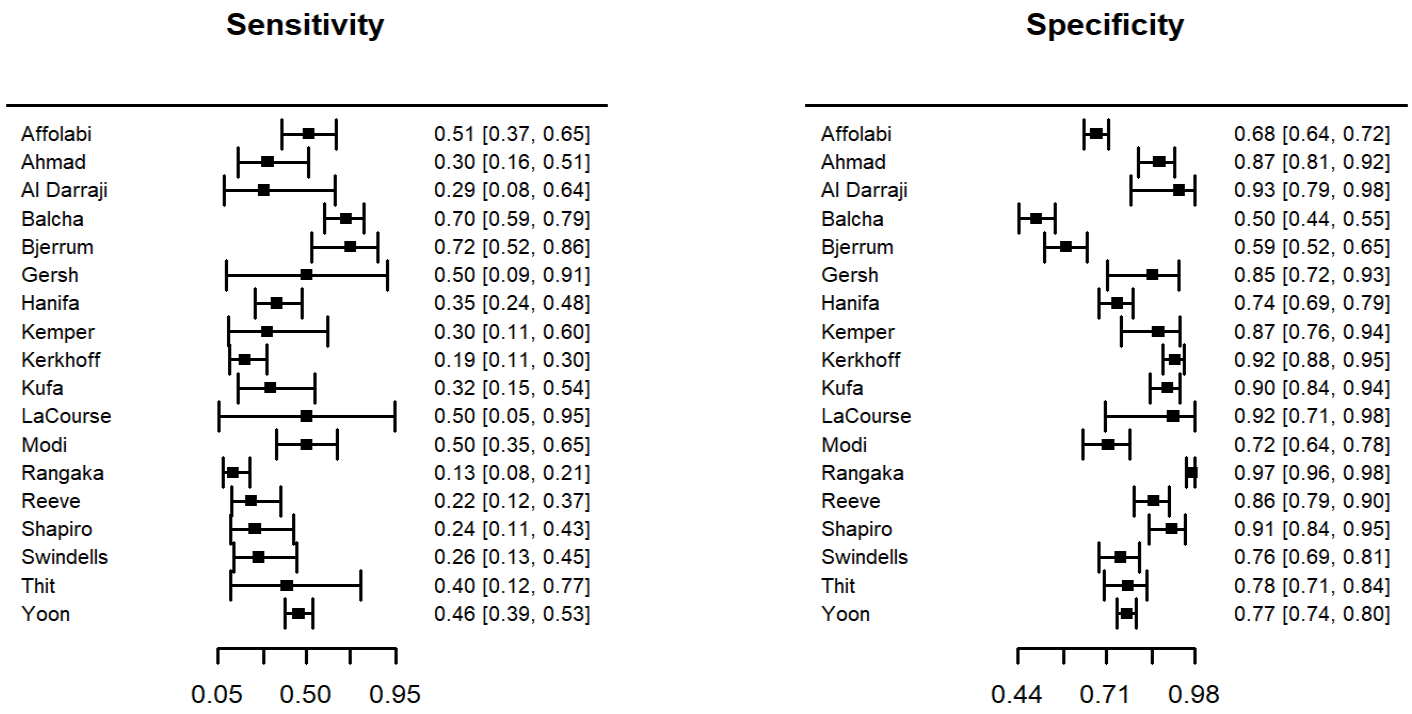




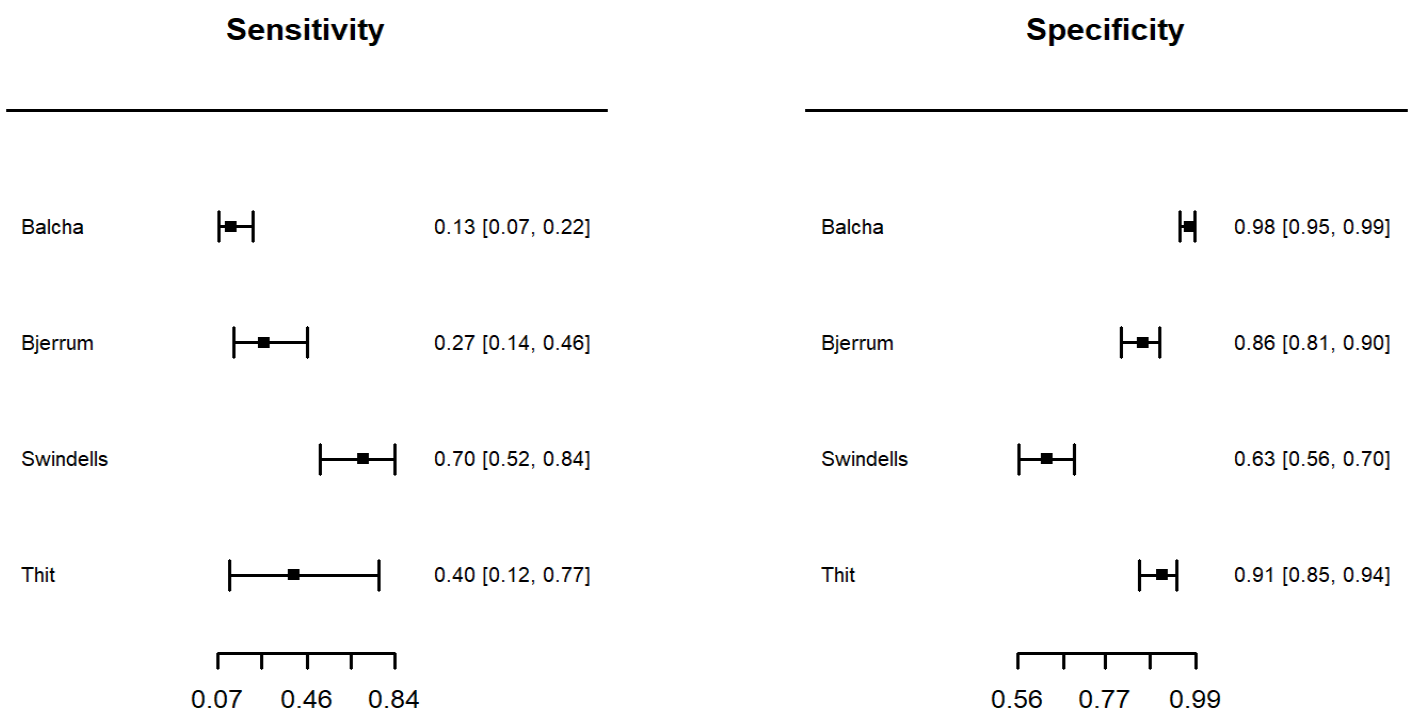




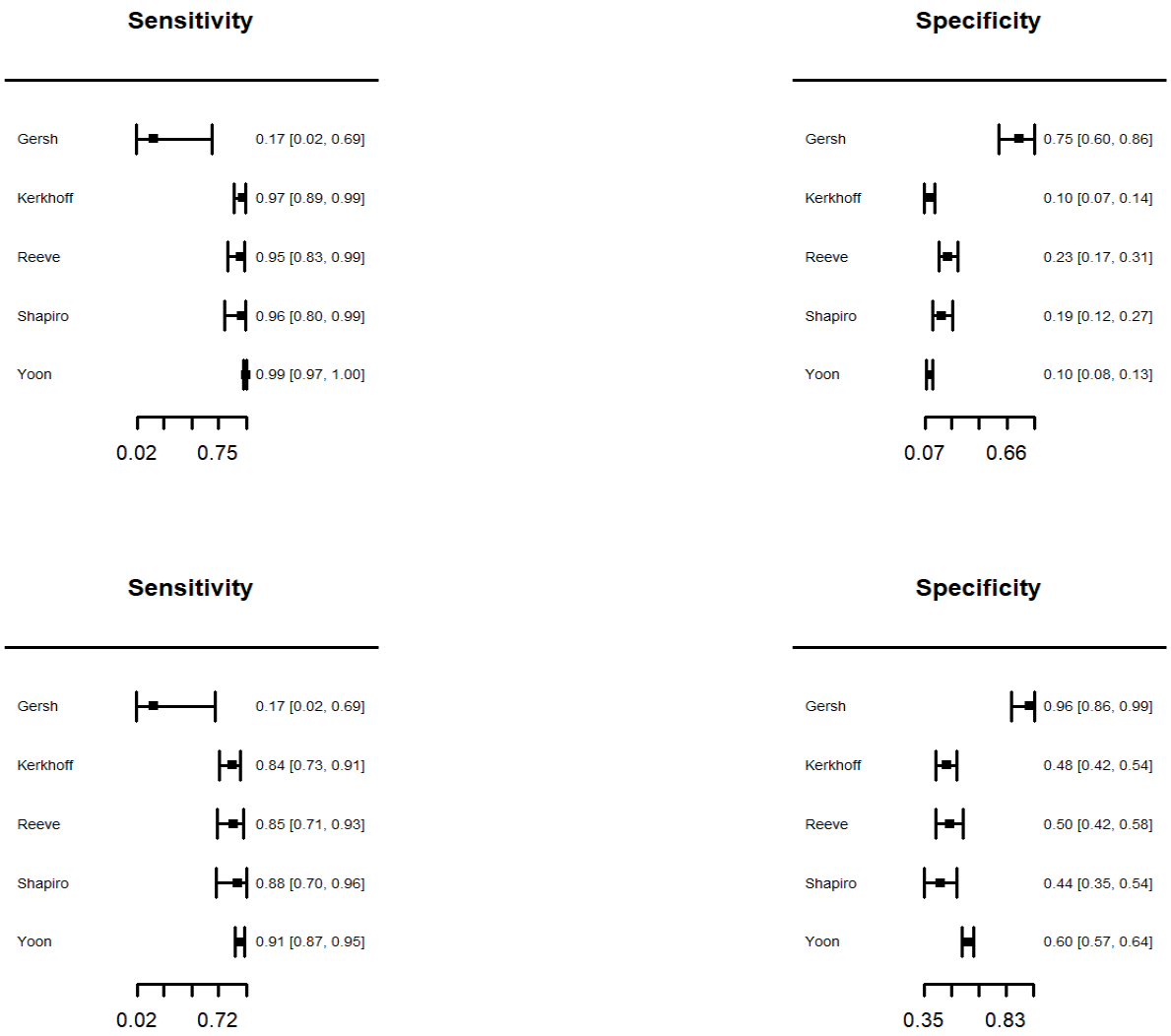
CD4 <=200 cells/µL - Forest plot for
BMI (<18.5 kg/m²)



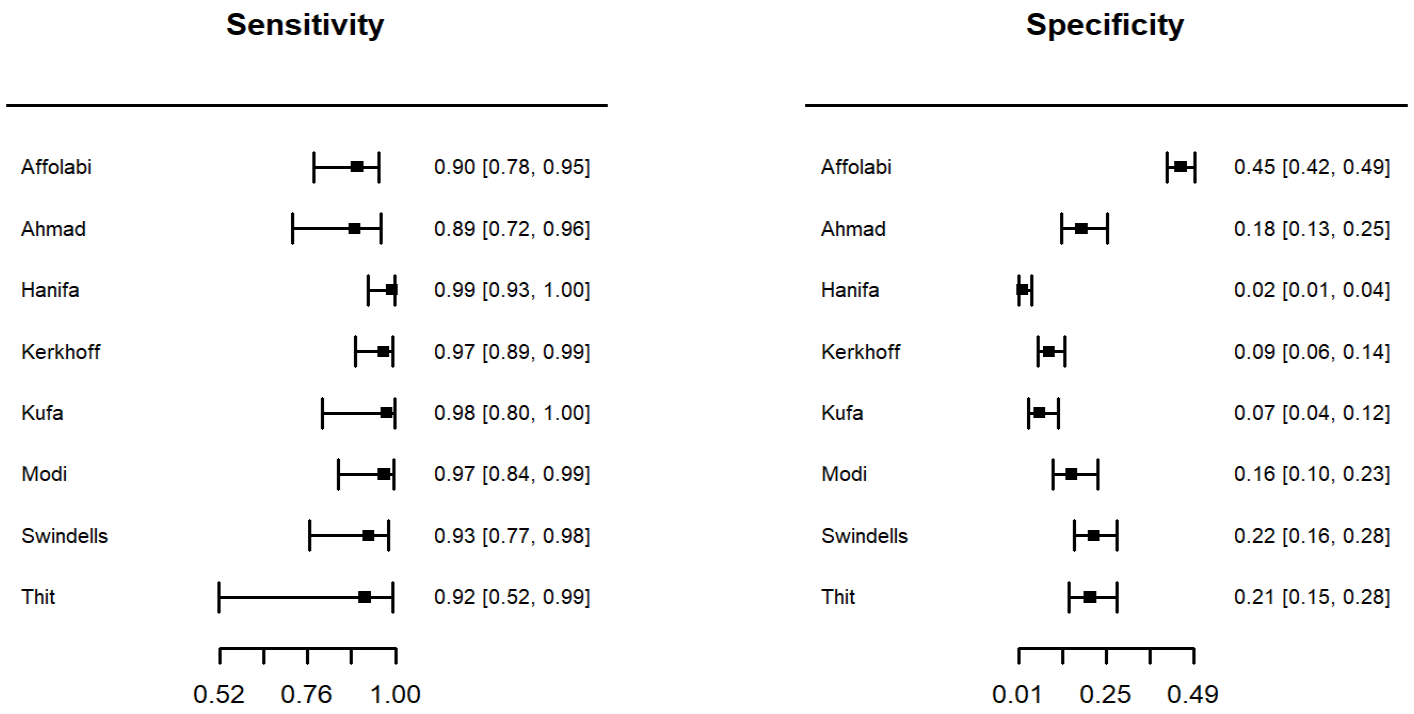
CD4 <=200 cells/µL - Forest plot for Lymphadenopathy



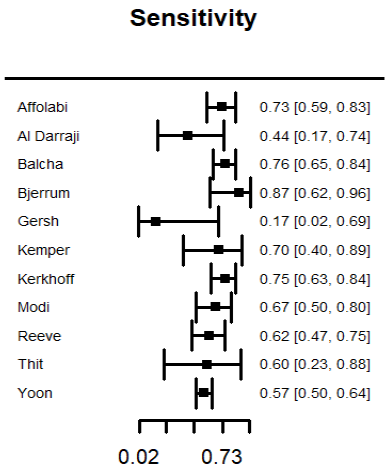
CD4 <=200 cells/μL - Forest plot for
Top: W4SS with CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)

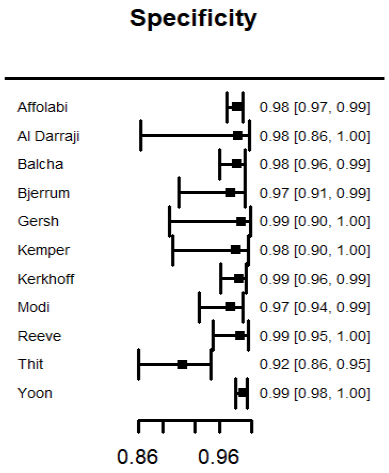


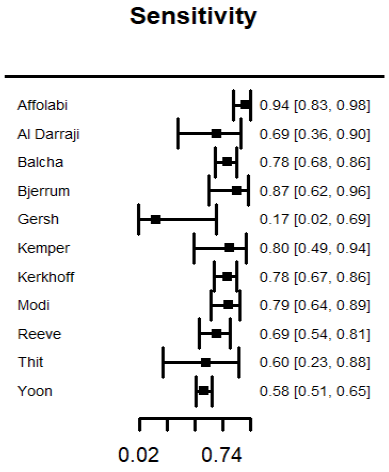
CD4 <=200 cells/µL - Forest plot for
W4SS with CXR (abnormal)



CD4 <=200 cells/μL - Forest plot for
Top: W4SS then Xpert and Bottom: Xpert alone







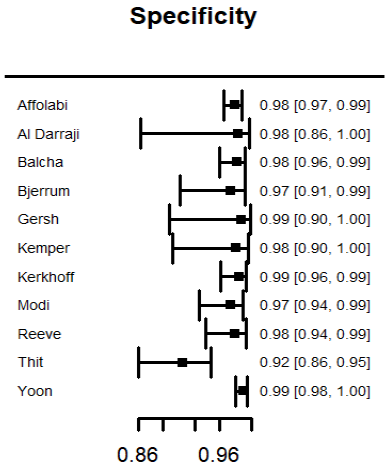
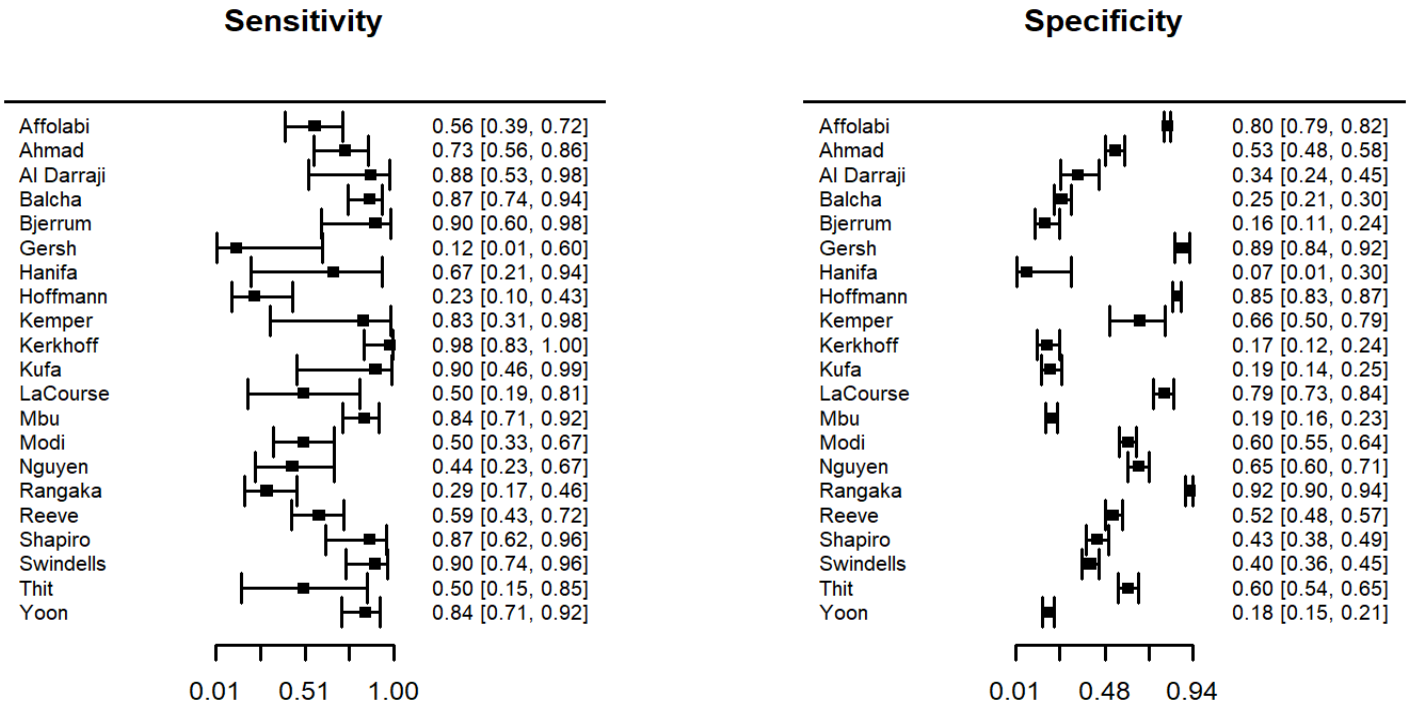
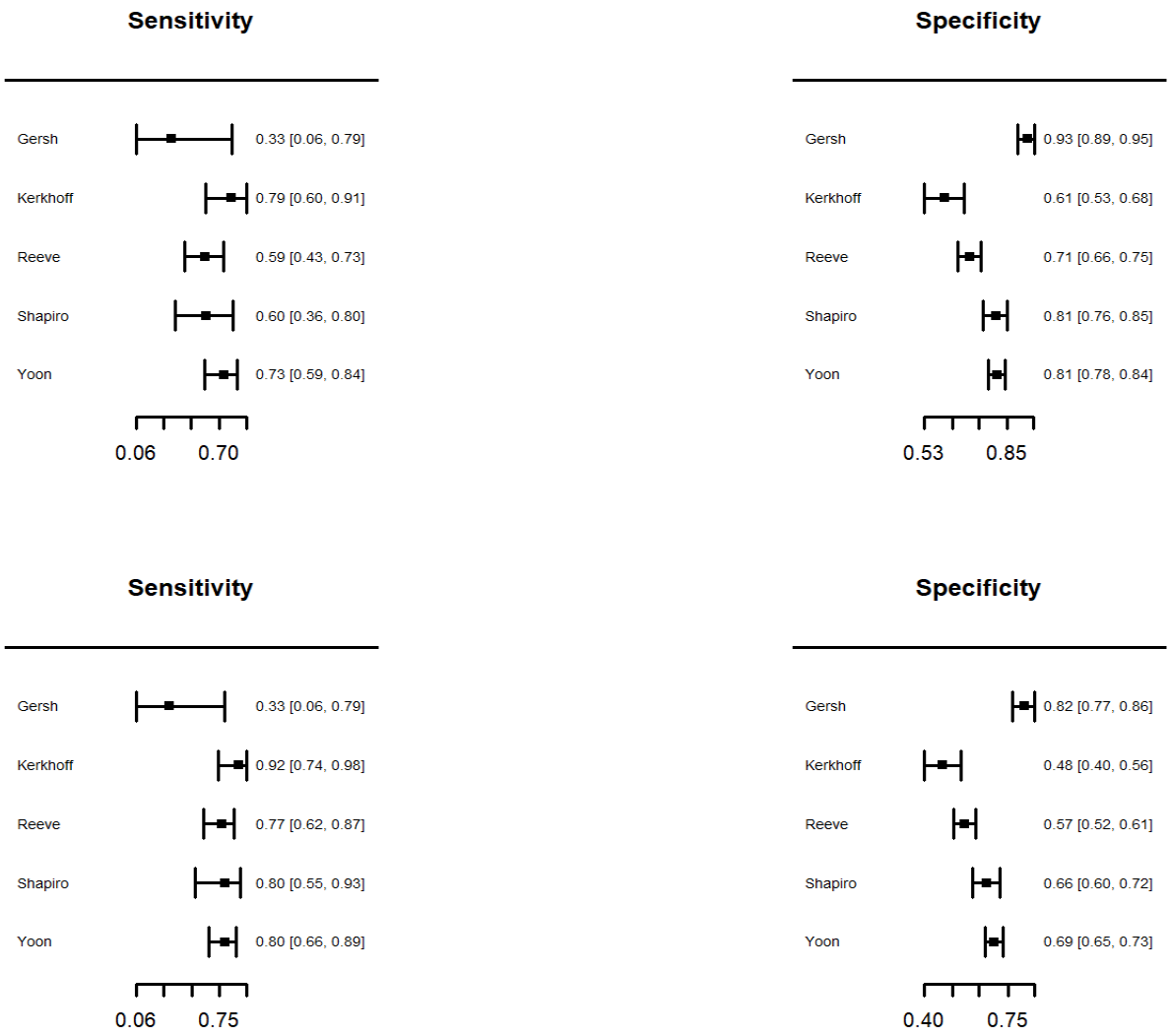


Figure 8-3E - Forest plots of sensitivity and specificity estimates in participants with a CD4 cell count >200 cells/µL

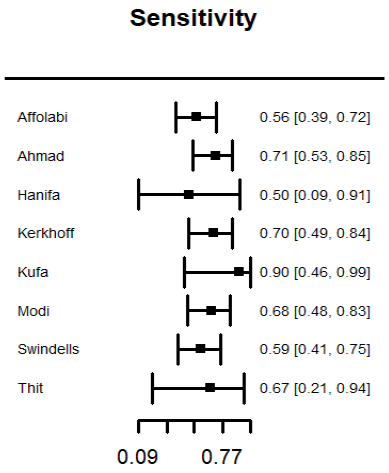
CD4 >200 cells/µL - Forest plot for
W4SS

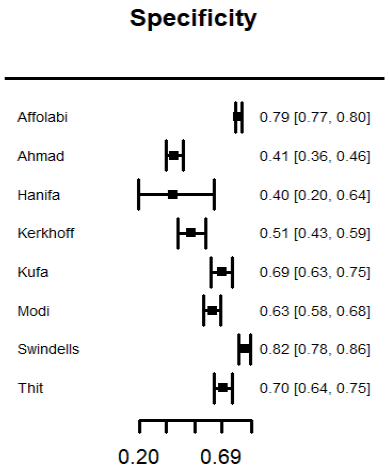


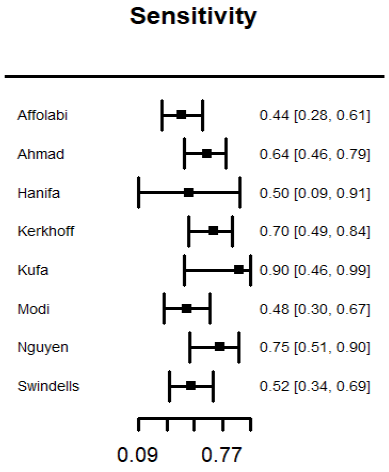
CD4 >200 cells/μL - Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)

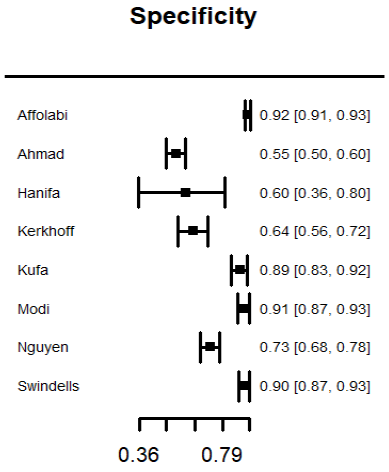


CD4 >200 cells/μL - Forest plot for
Top: CXR (abnormal) and Bottom: CXR (suggests tuberculosis)



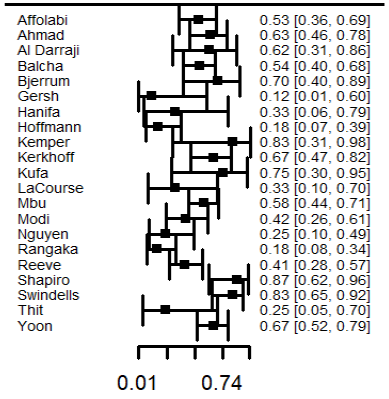




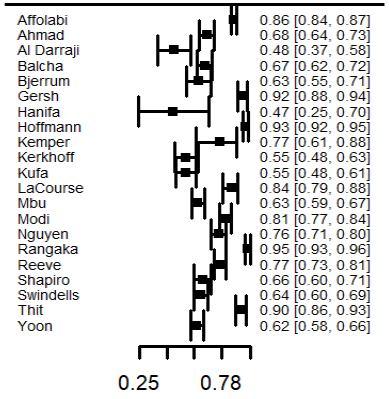


CD4 >200 cells/μL - Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

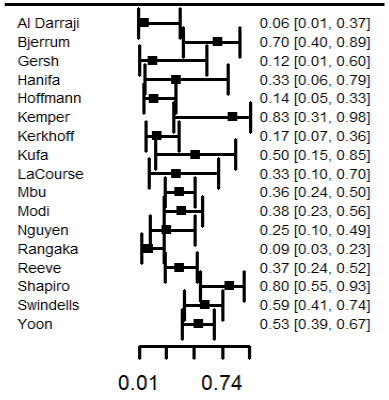
Sensitivity



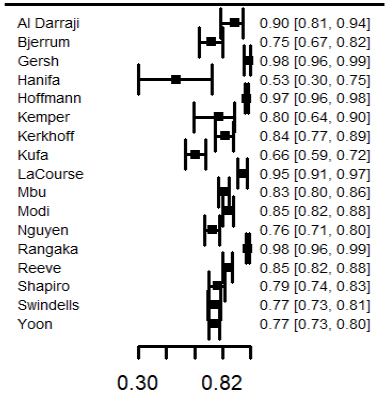
Specificity



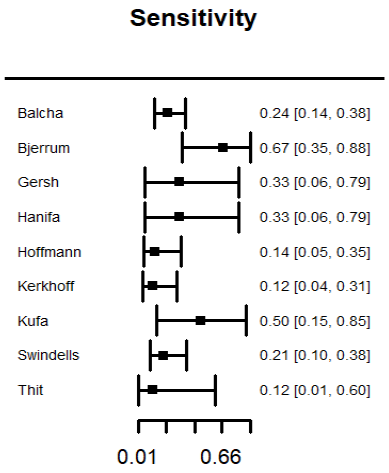
Sensitivity

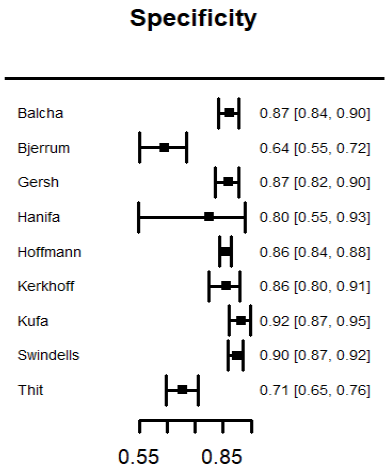


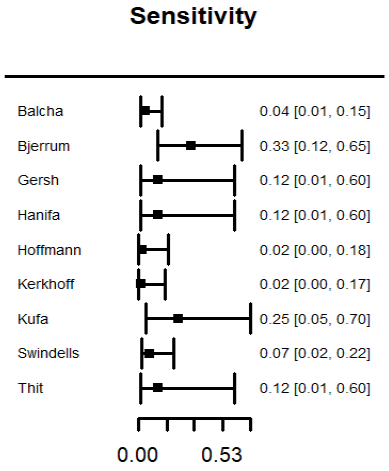
Specificity

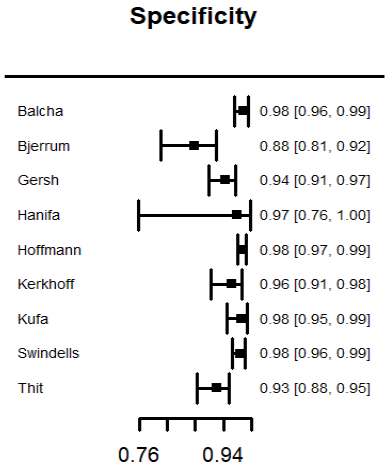


CD4 >200 cells/μL - Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)

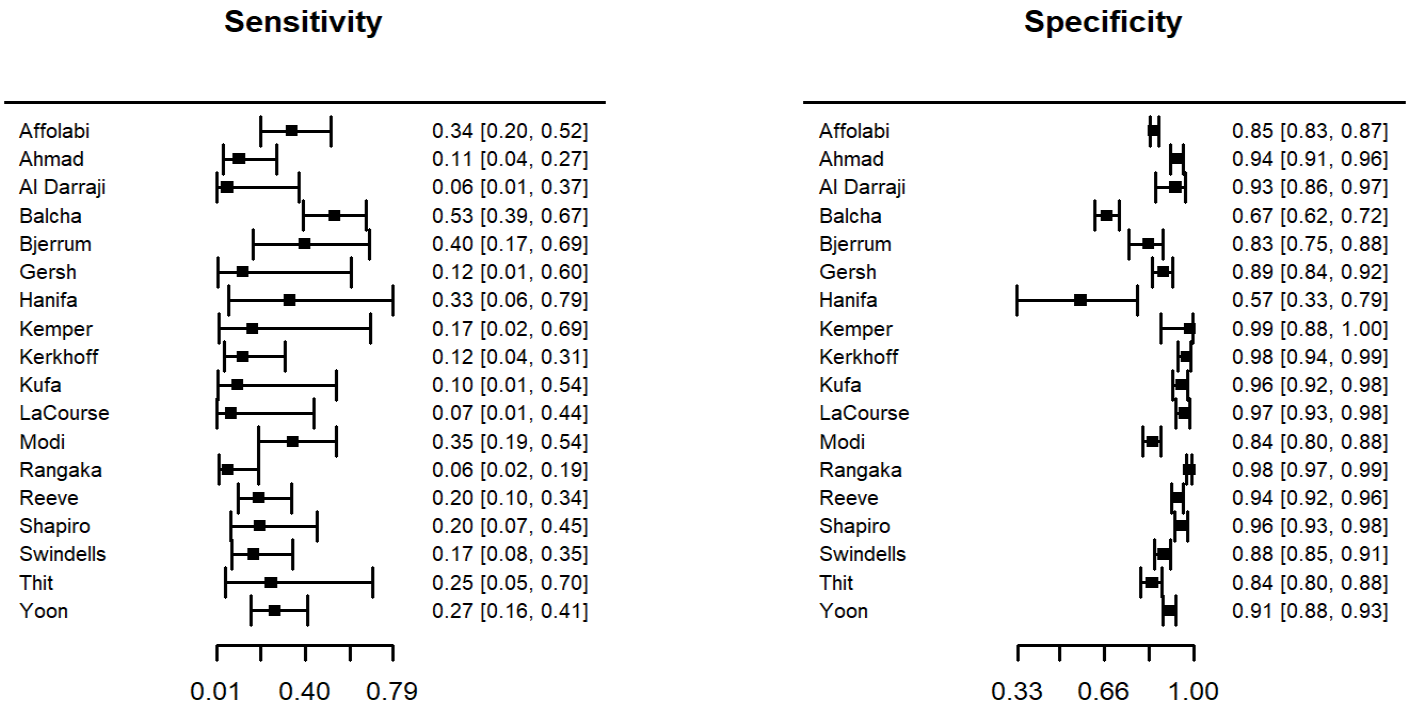




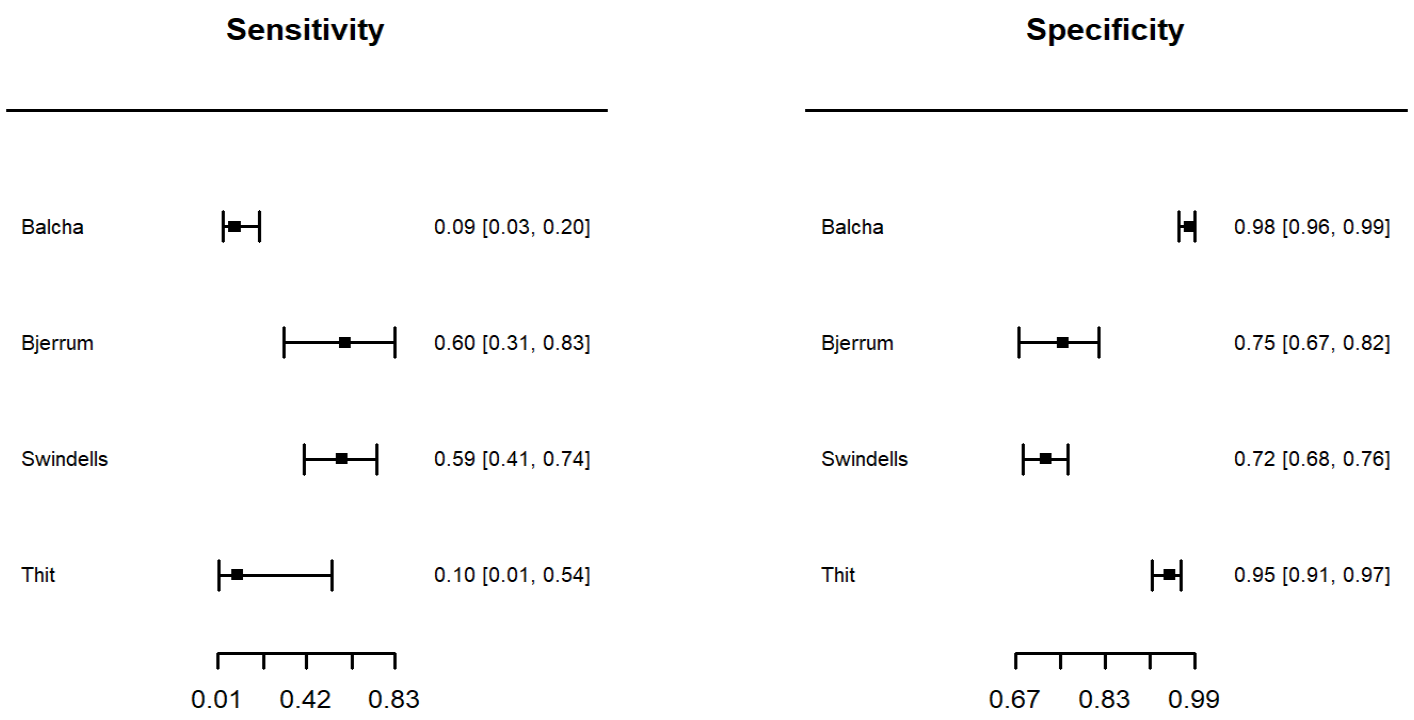




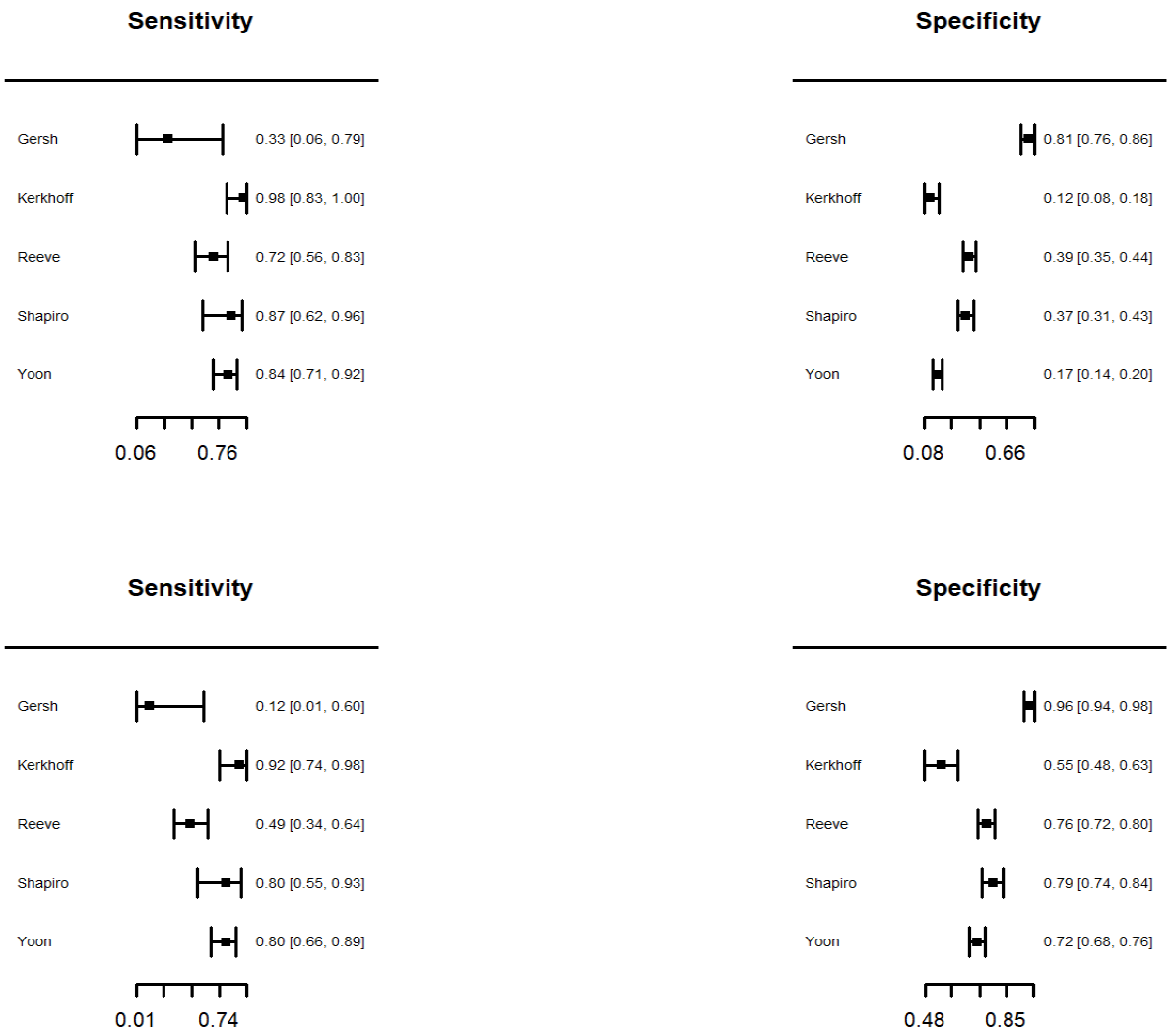
CD4 >200 cells/μL - Forest plot for
BMI (<18.5 kg/m²)



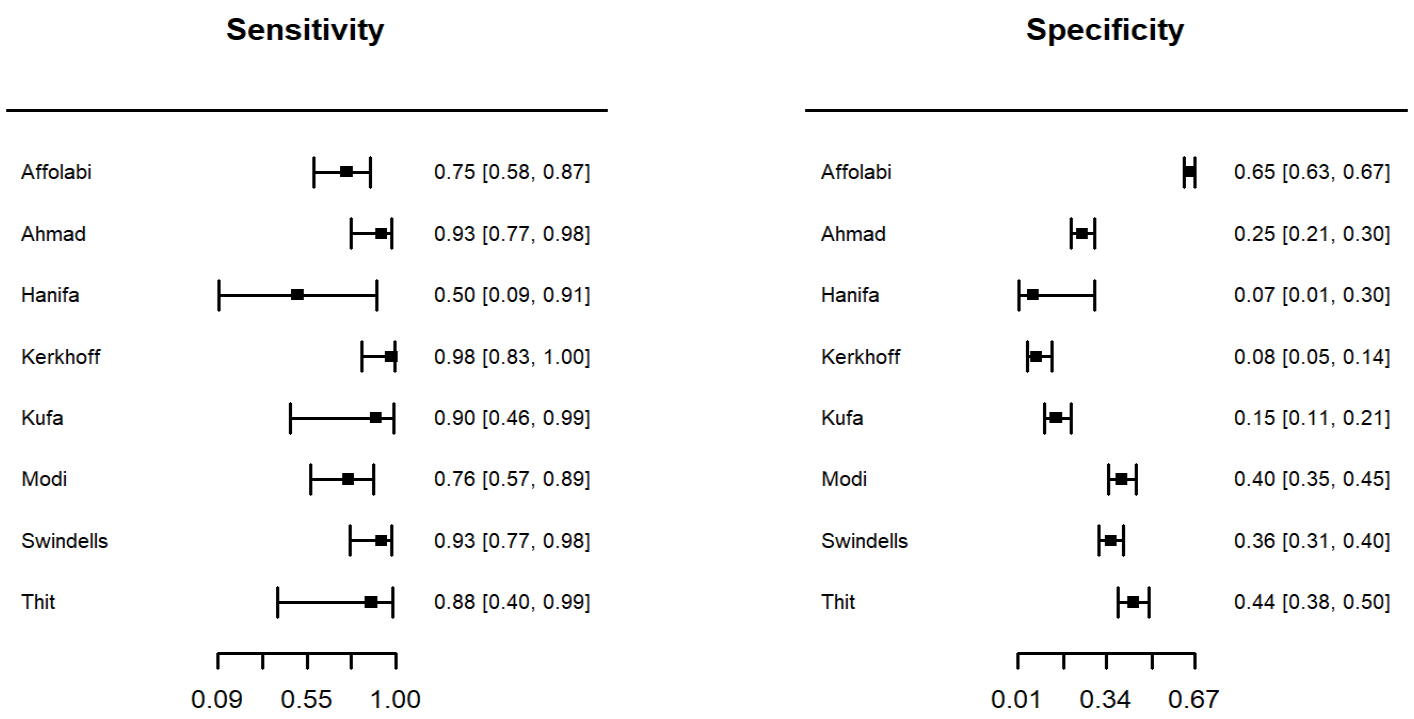
CD4 >200 cells/μL - Forest plot for Lymphadenopathy



CD4 >200 cells/μL - Forest plot for
Top: W4SS with CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)

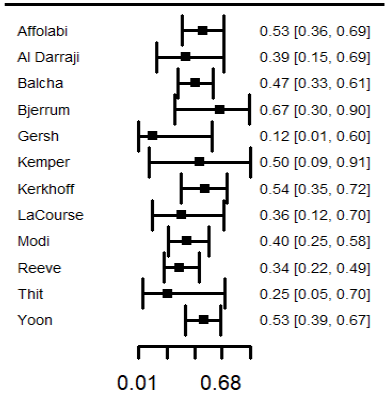


CD4 >200 cells/μL - Forest plot for
W4SS with CXR (abnormal)

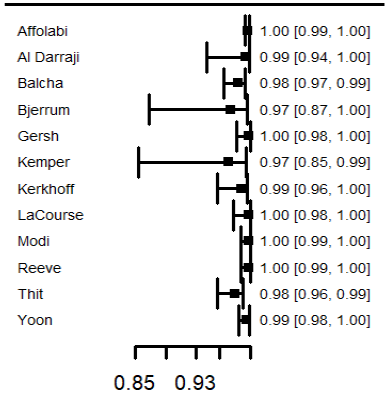


CD4 >200 cells/μL - Forest plot for
Top: W4SS then Xpert and Bottom: Xpert alone

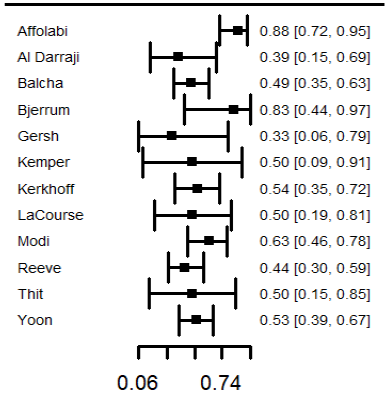
Sensitivity



Specificity



Sensitivity



Specificity

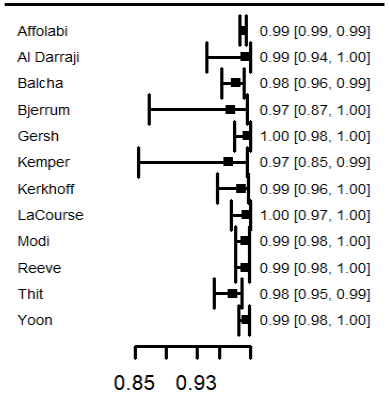
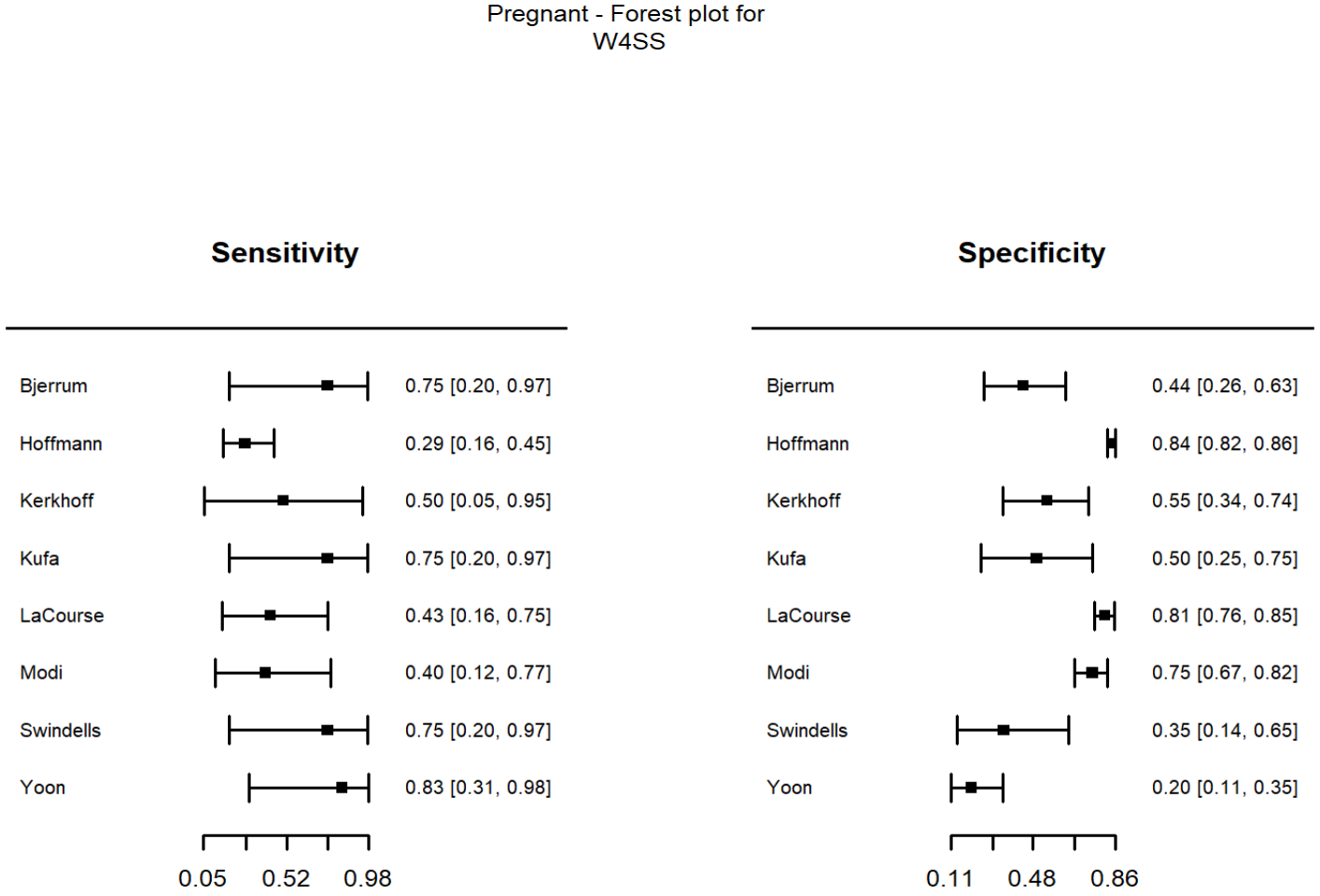
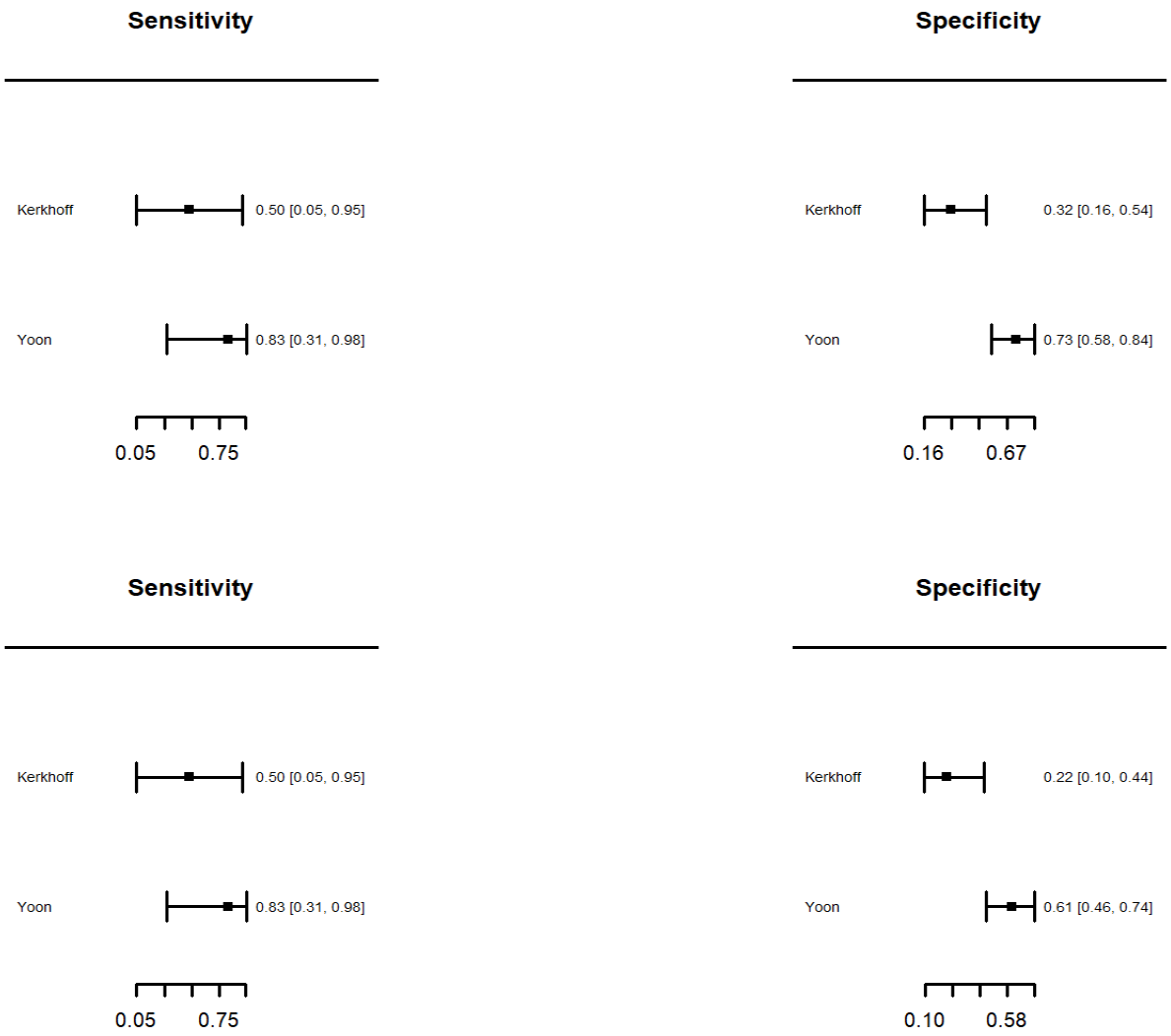


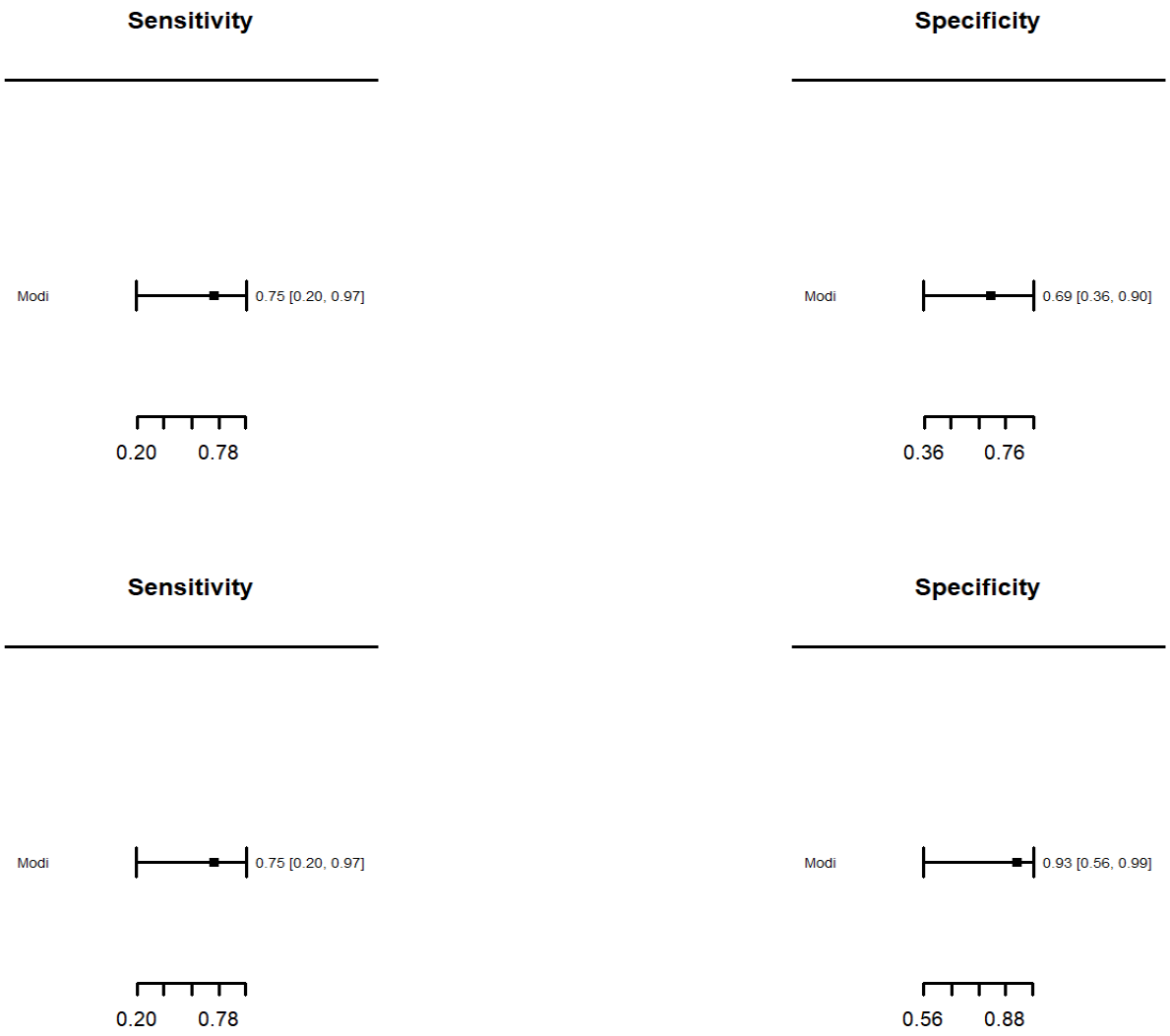
Figure 8-3F - Forest plots of sensitivity and specificity estimates in pregnant participants



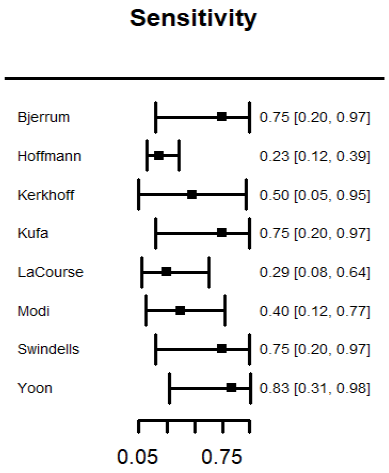
Pregnant - Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)

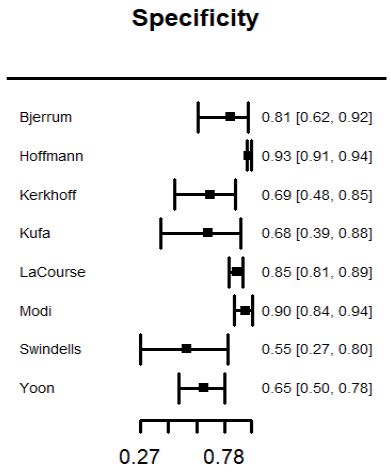


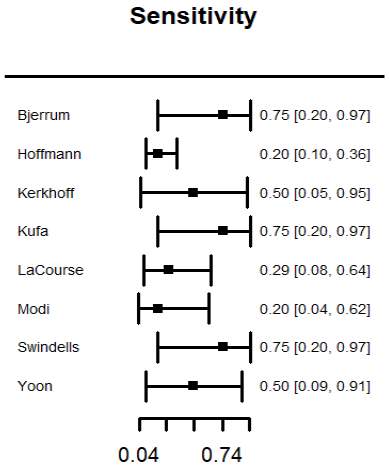
Pregnant - Forest plot for
Top: CXR (abnormal) and Bottom: CXR (suggests tuberculosis)

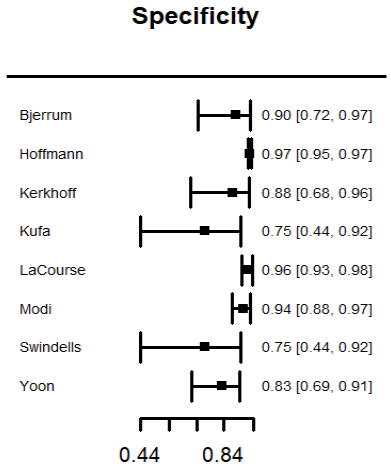


Pregnant - Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

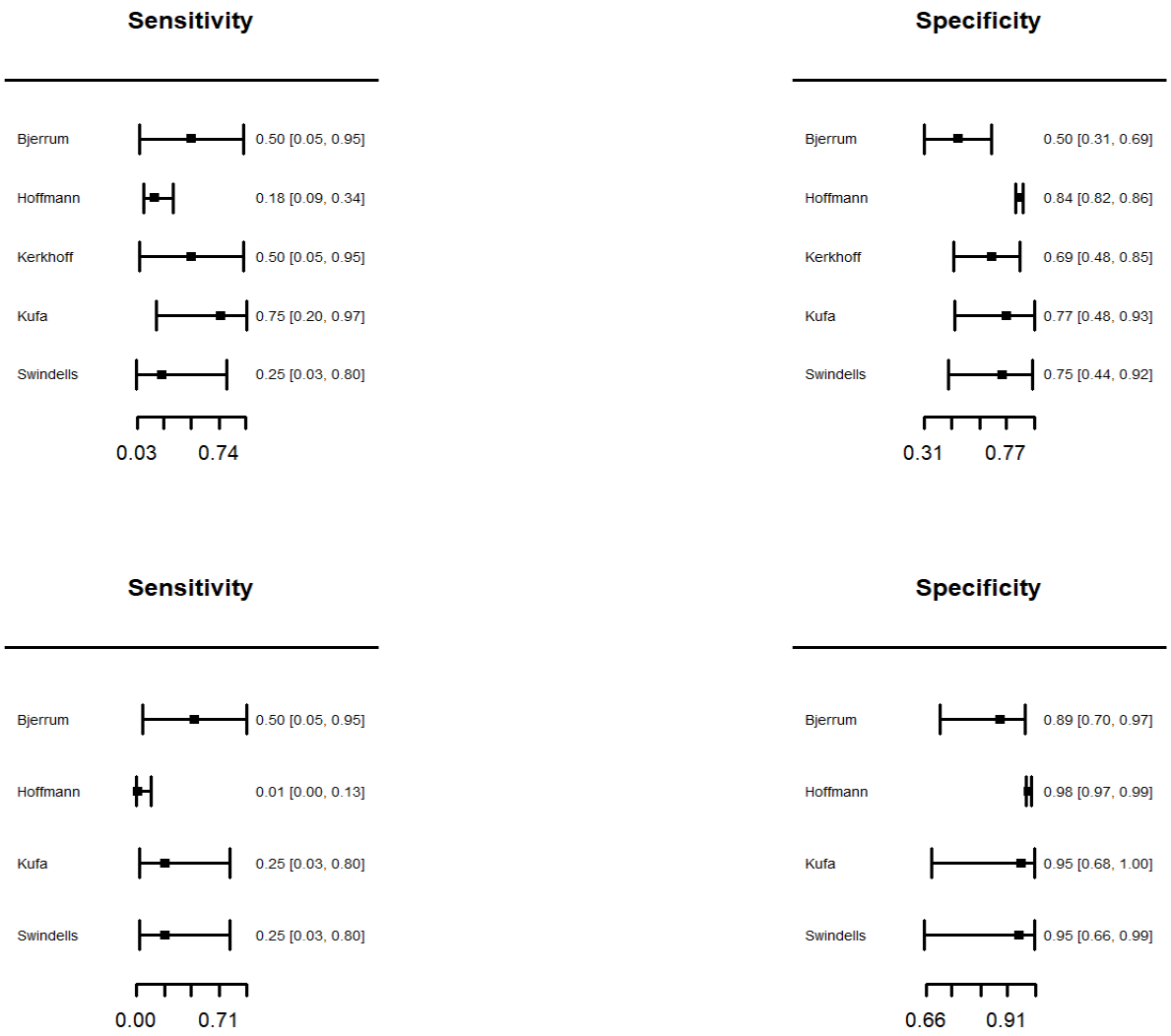




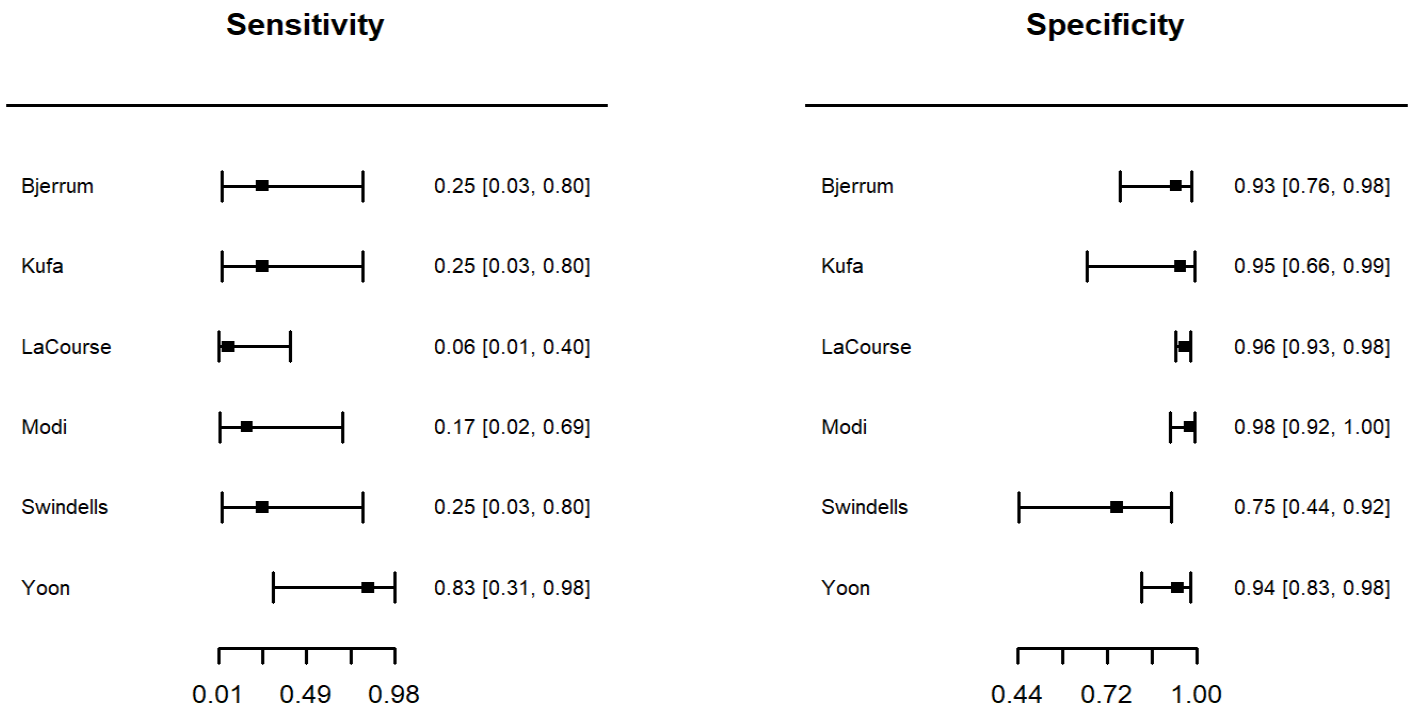




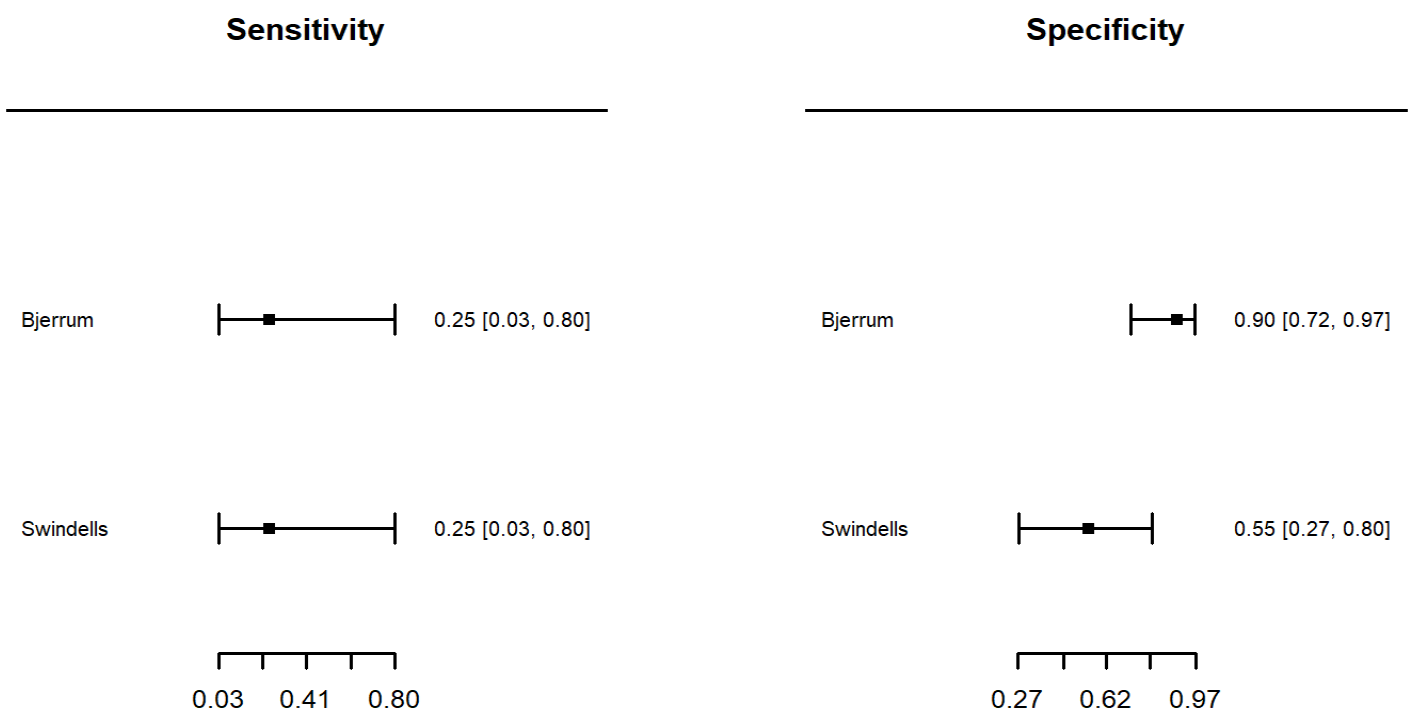
Pregnant - Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)



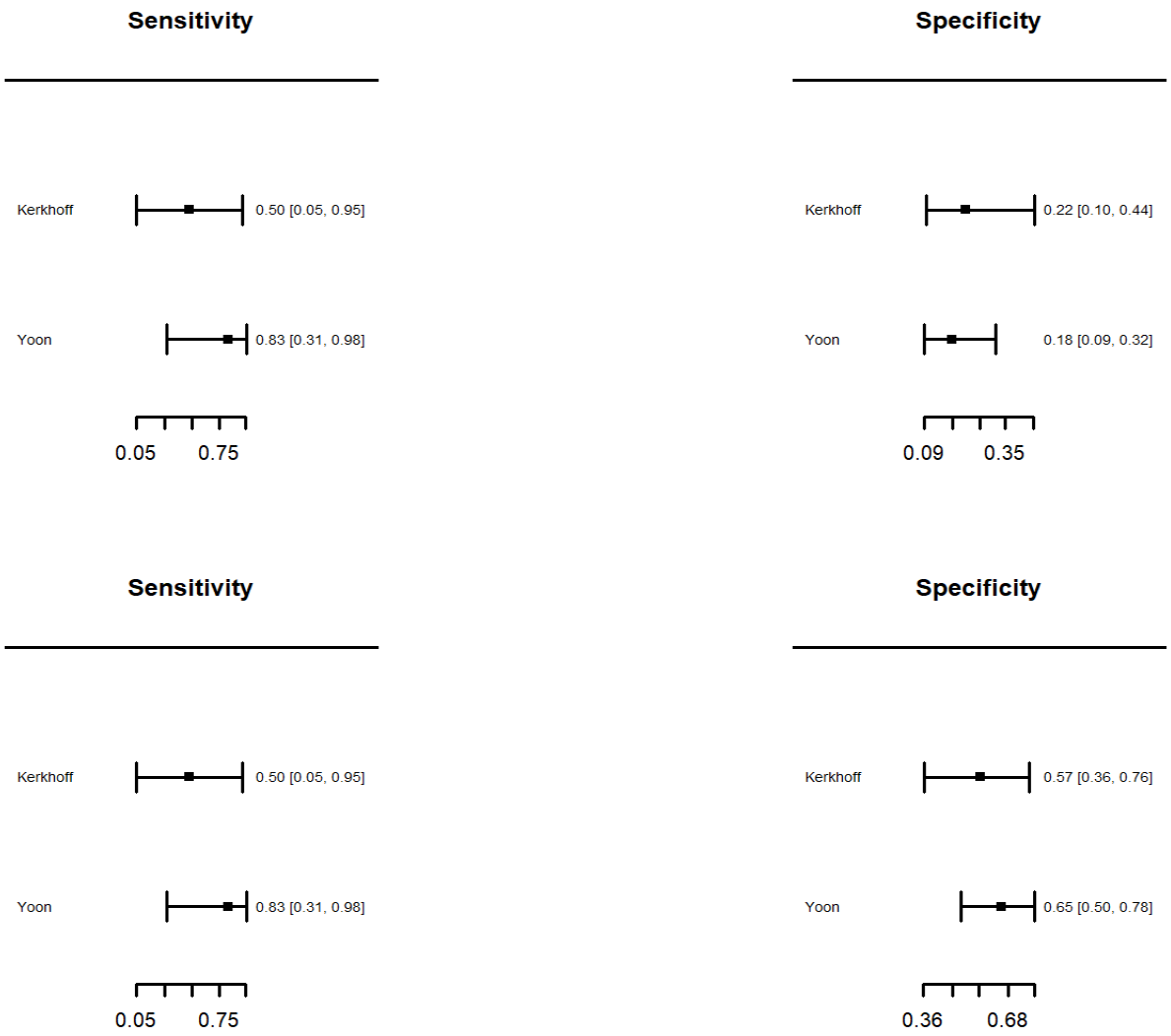
Pregnant - Forest plot for
BMI (<18.5 kg/m²)



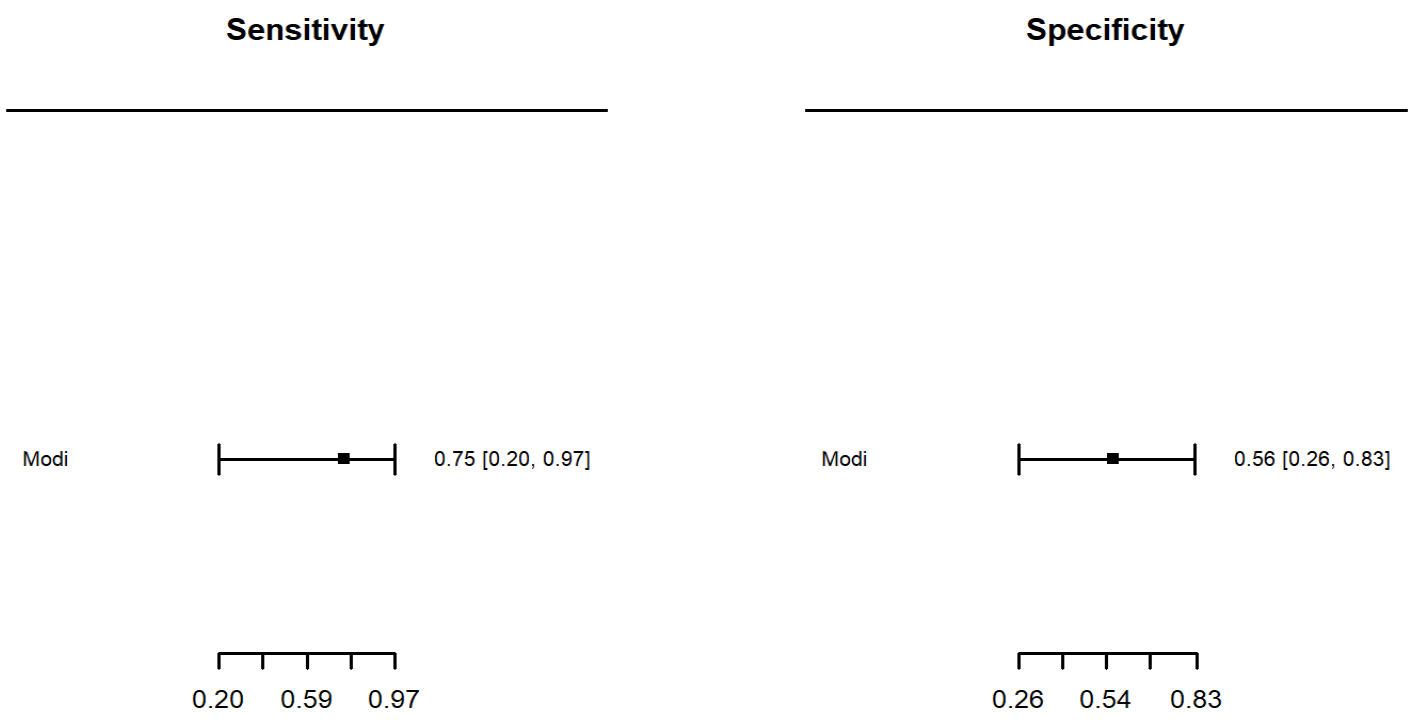
Pregnant - Forest plot for
Lymphadenopathy



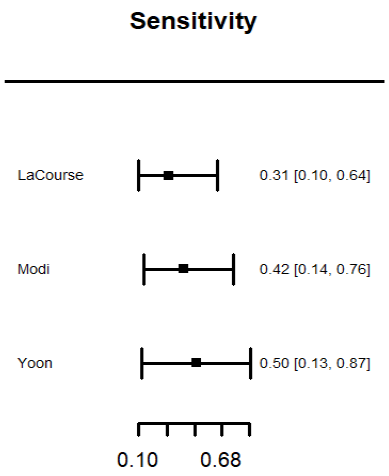
Pregnant - Forest plot for
Top: W4SS with CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)

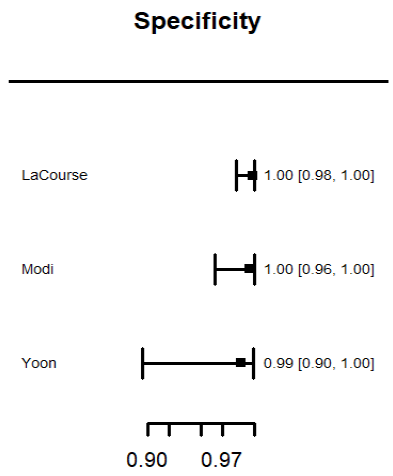


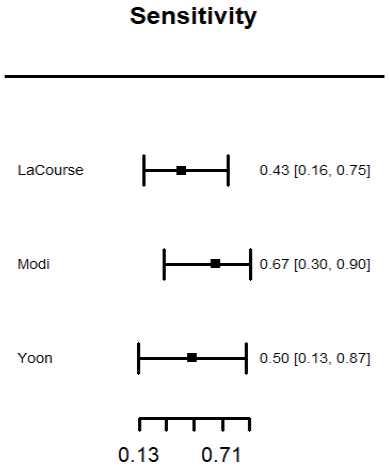
Pregnant - Forest plot for
W4SS with CXR (abnormal)



Pregnant - Forest plot for
Top: W4SS then Xpert and Bottom: Xpert alone







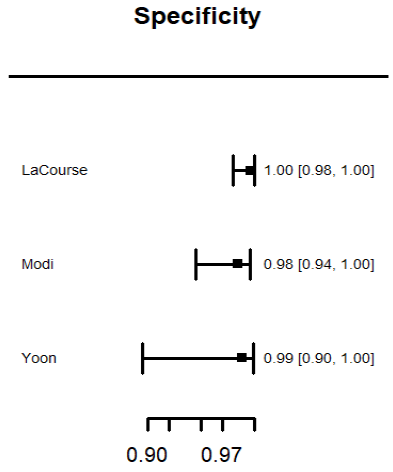
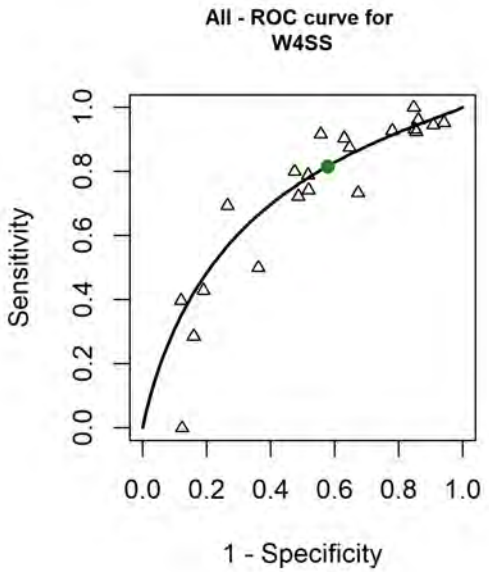
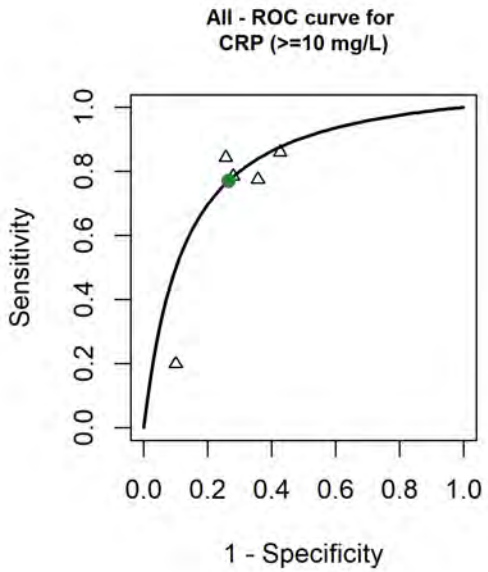
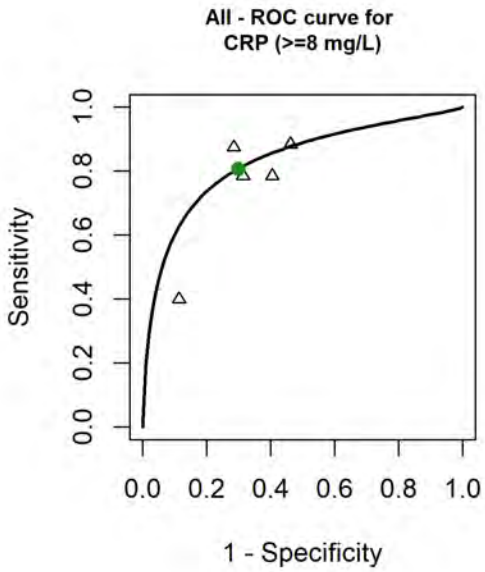


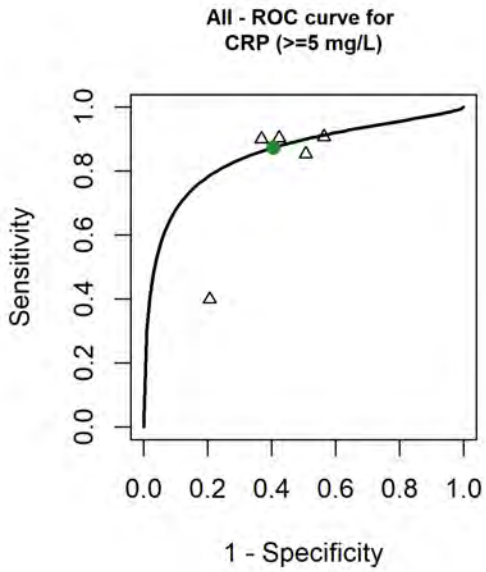
Figure 8-4: Summary receiver operating characteristics curves in all participants and subgroups (for tests/strategies with >=2 studies available)

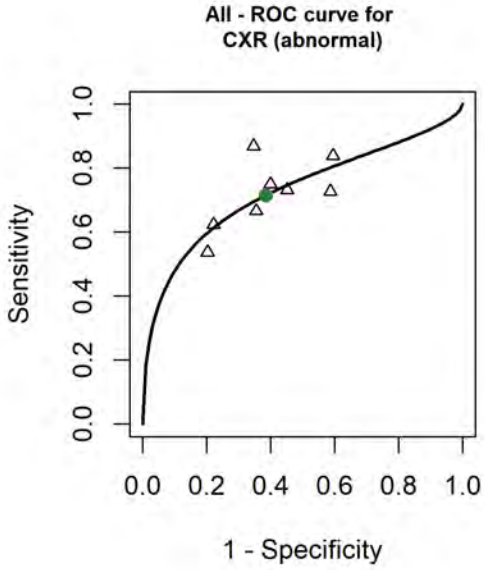
Figure 8-4A - Summary receiver operating characteristics curves comparing each test and WHO four-symptom screen for the detection of tuberculosis in all participants

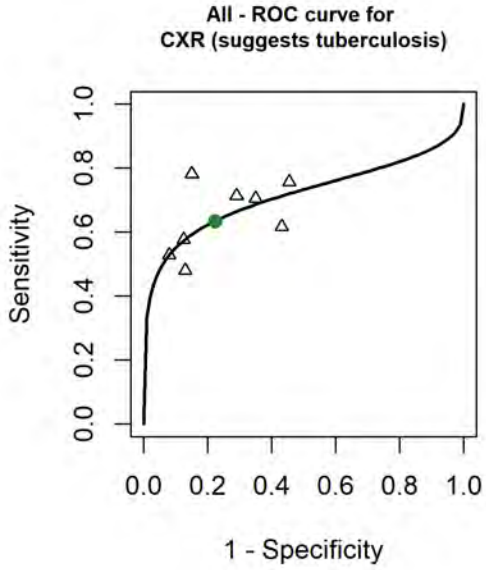


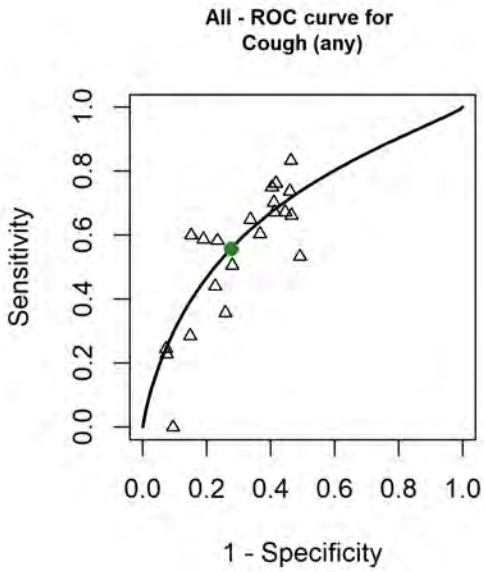


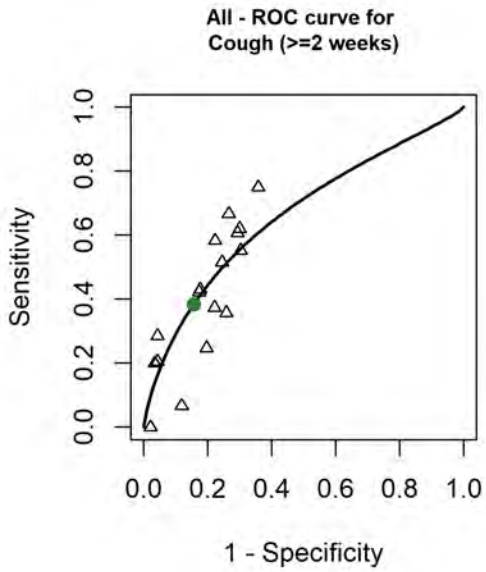

All - ROC curve for
CRP (>=8 mg/L)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

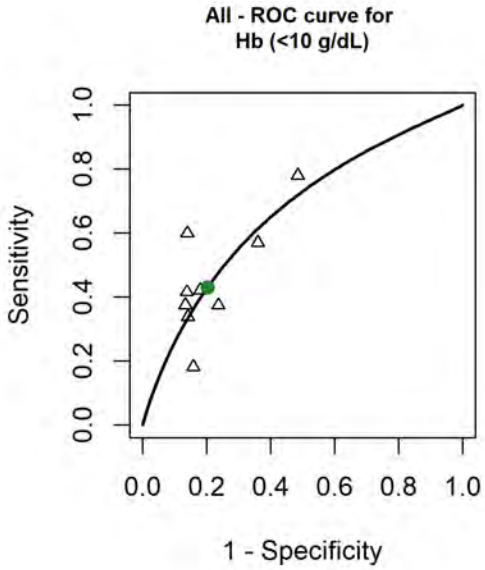

All - ROC curve for
CRP (>=5 mg/L)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

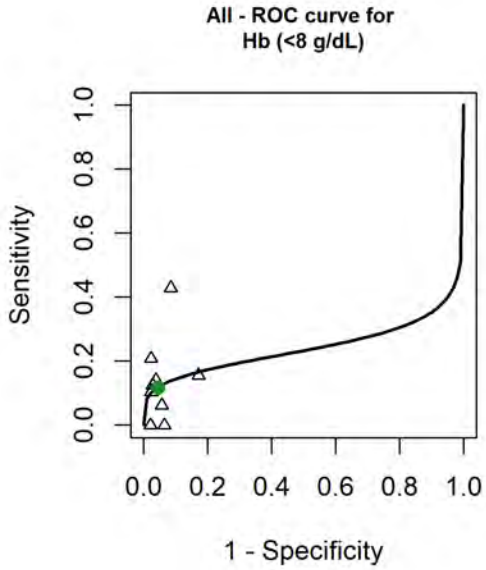

All - ROC curve for
CXR (abnormal)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

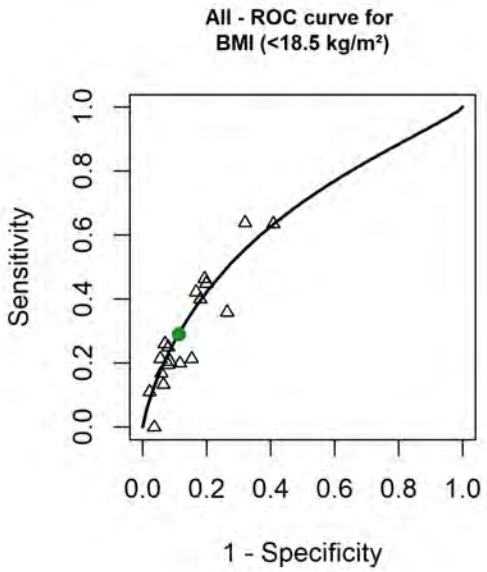

All - ROC curve for
CXR (suggests tuberculosis)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

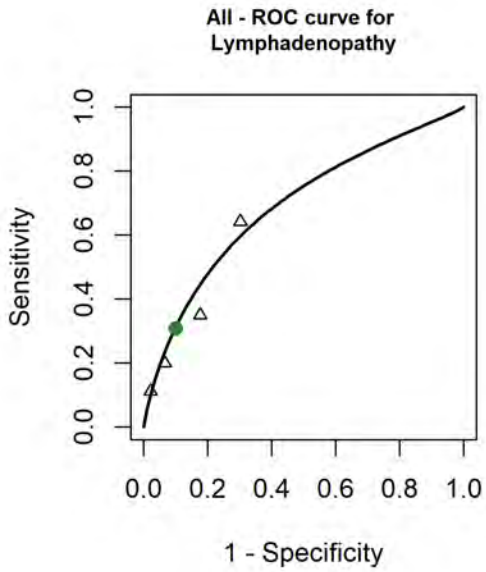

All - ROC curve for
Cough (any)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

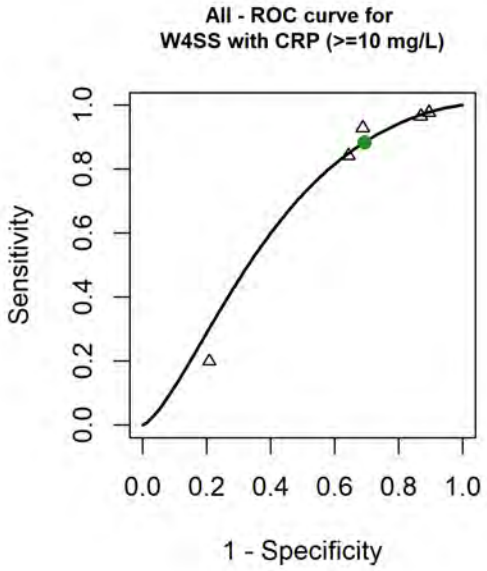

All - ROC curve for
Cough (>=2 weeks)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

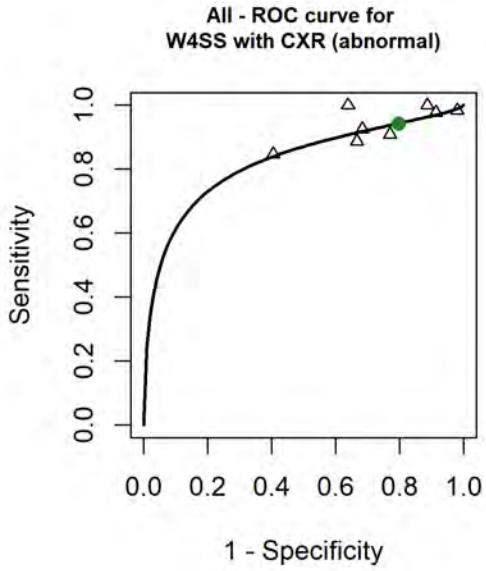

All - ROC curve for
Hb (<10 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

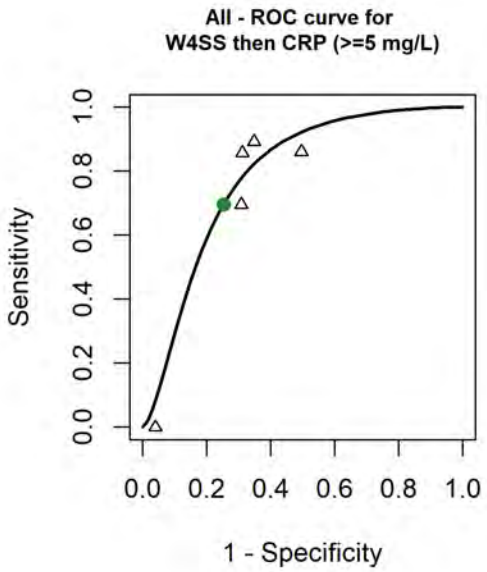

All - ROC curve for
Hb (<8 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

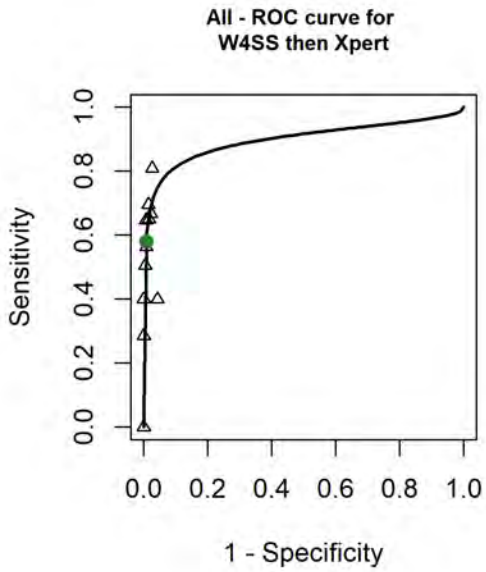

All - ROC curve for
BMI (<18.5 kg/m²)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


All - ROC curve for
Lymphadenopathy
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


All - ROC curve for
W4SS with CRP (>=10 mg/L)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


All - ROC curve for
W4SS with CXR (abnormal)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


All - ROC curve for
W4SS then CRP (>=5 mg/L)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


All - ROC curve for
W4SS then Xpert
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

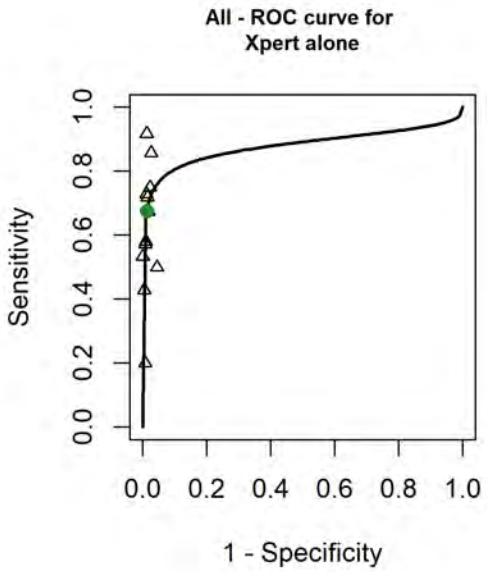
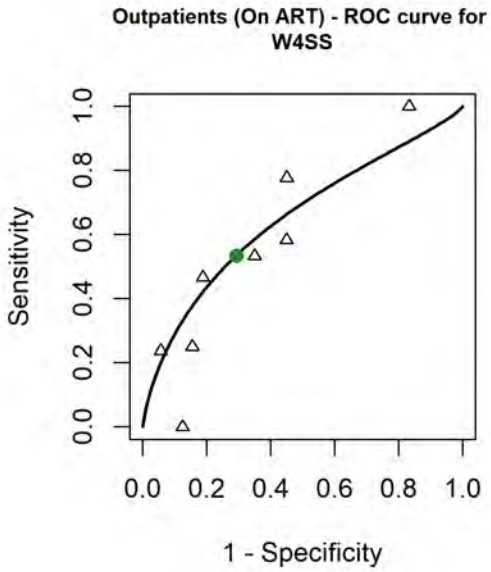
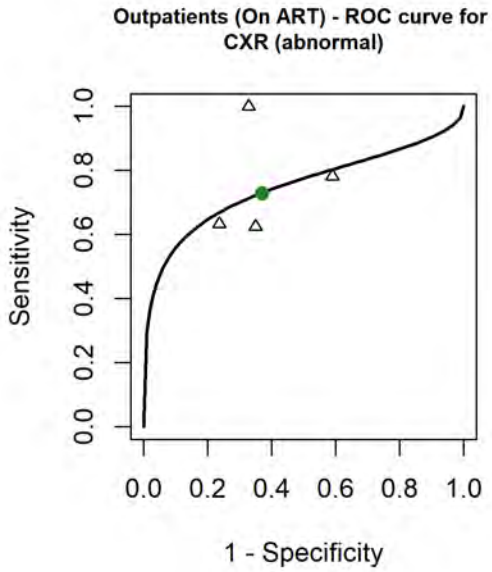
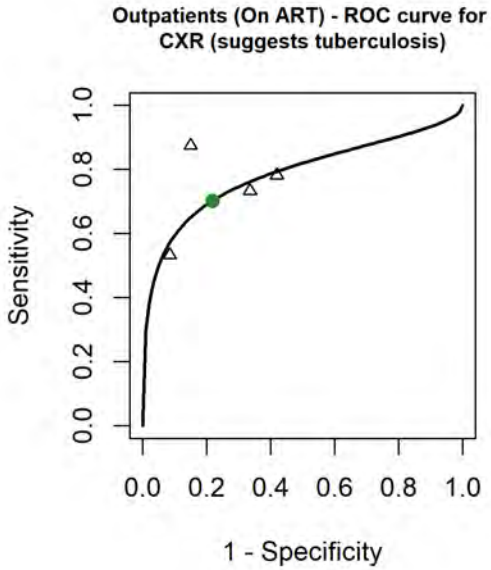
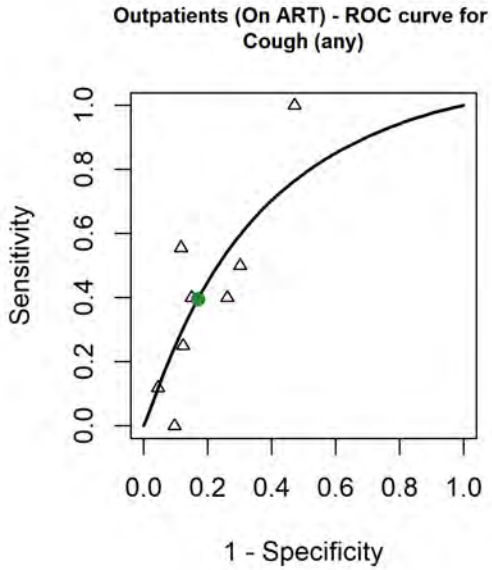

All - ROC curve for
Xpert alone
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

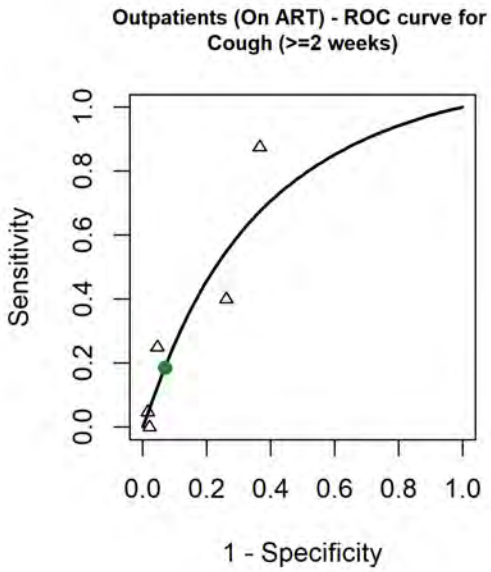
Figure 8-4B - Summary receiver operating characteristics curves comparing each test and WHO four-symptom screen for the detection of tuberculosis in outpatients (on ART)

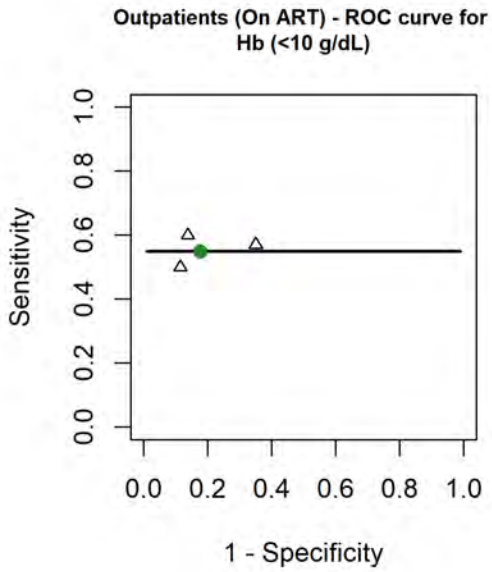


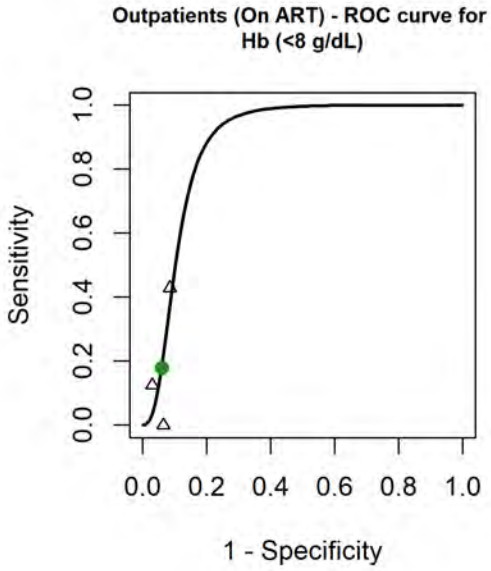


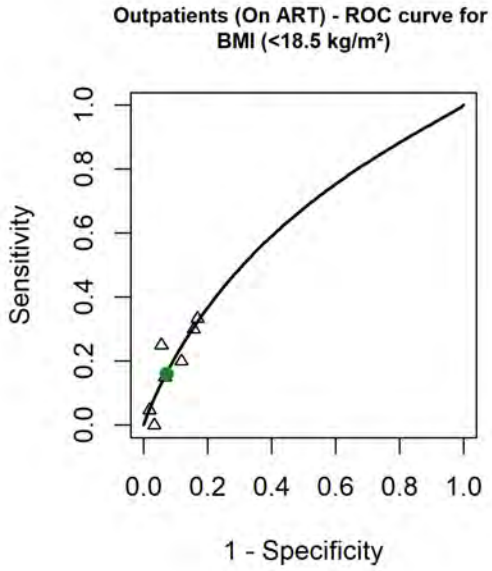


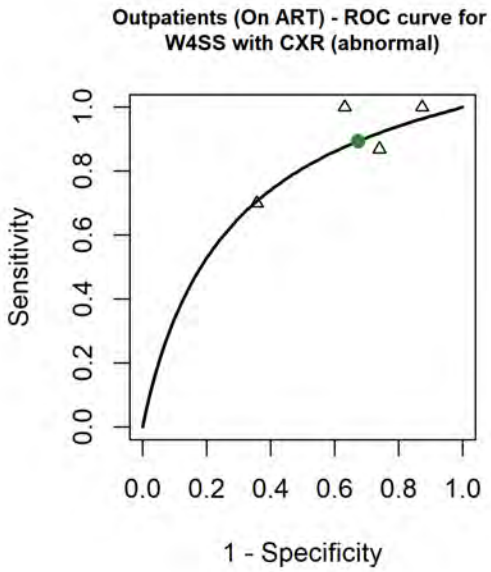


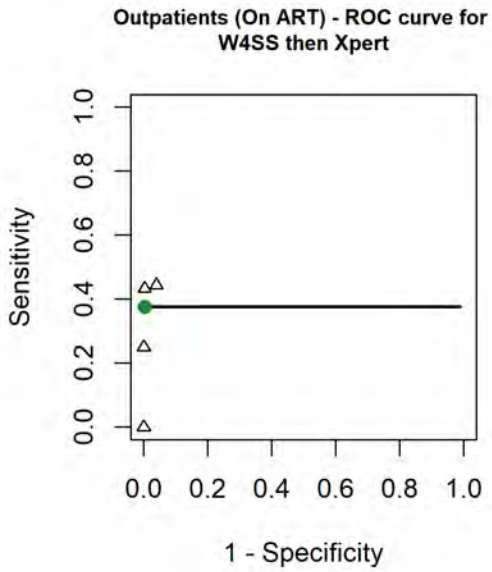

Outpatients (On ART) - ROC curve for
Cough (>=2 weeks)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


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


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


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


Outpatients (On ART) - ROC curve for W4SS with CXR (abnormal)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Outpatients (On ART) - ROC curve for W4SS then Xpert
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

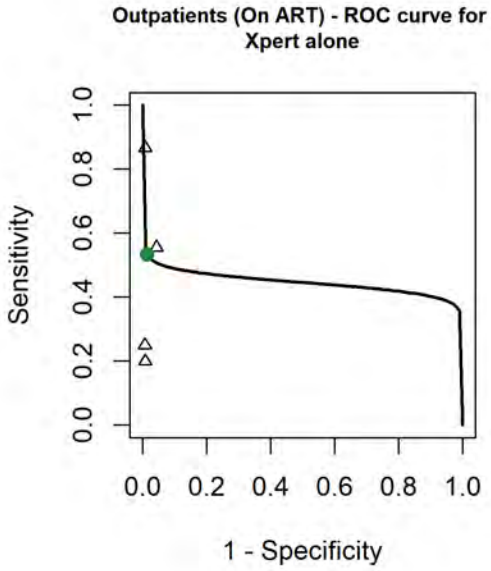
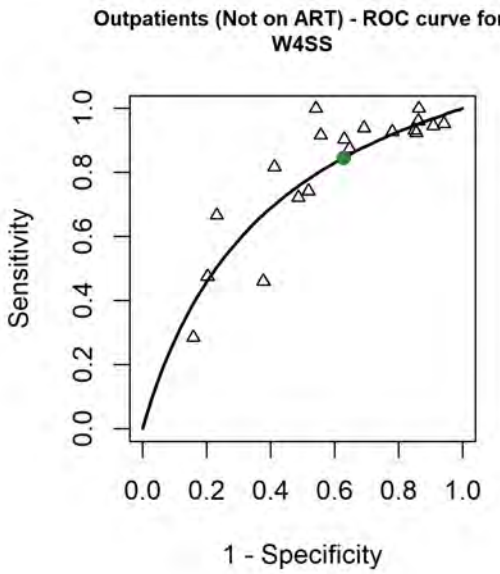
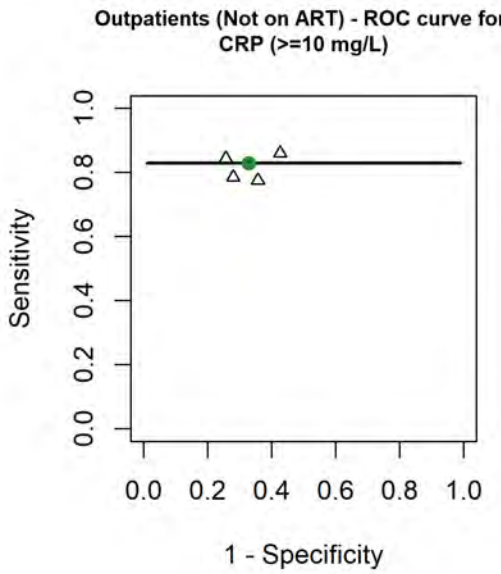
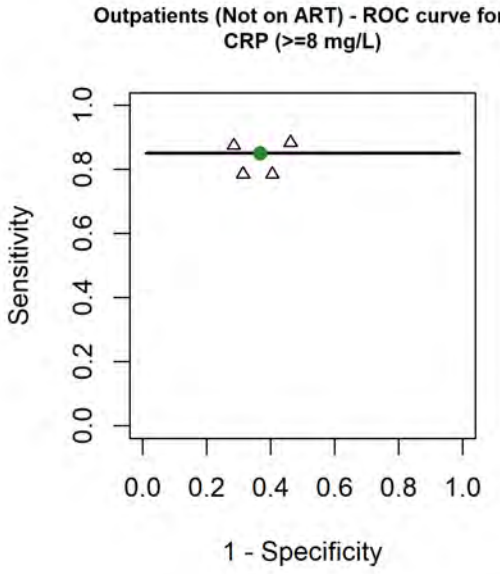
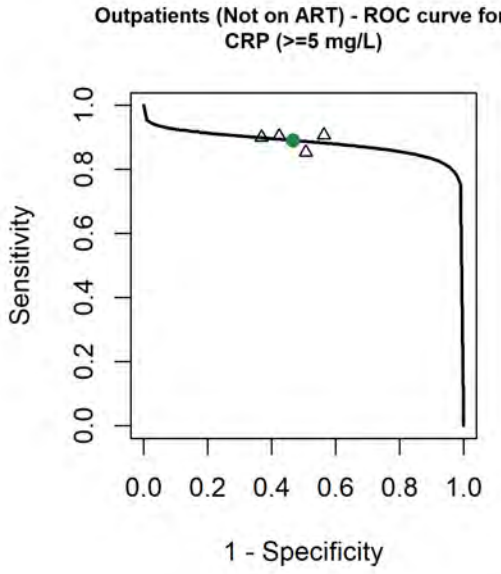

Outpatients (On ART) - ROC curve for Xpert alone
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

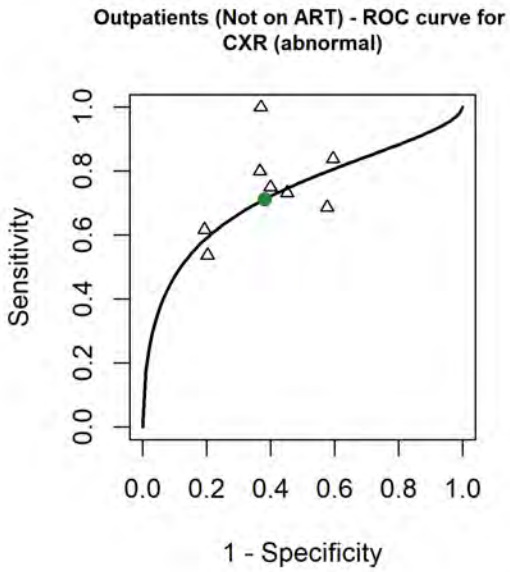
Figure 8-4C - Summary receiver operating characteristics curves comparing each test and WHO four-symptom screen for the detection of tuberculosis in outpatients (not on ART)

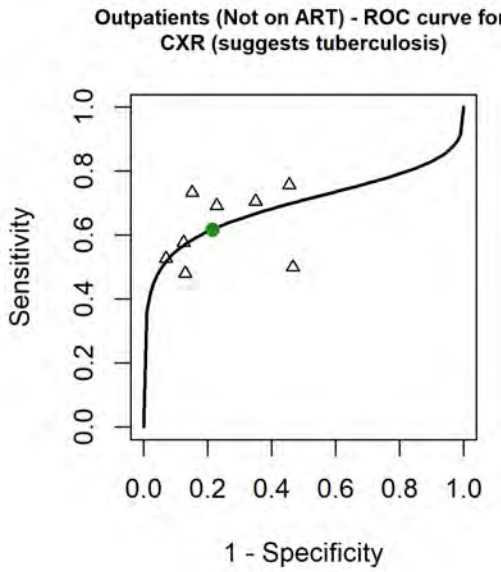


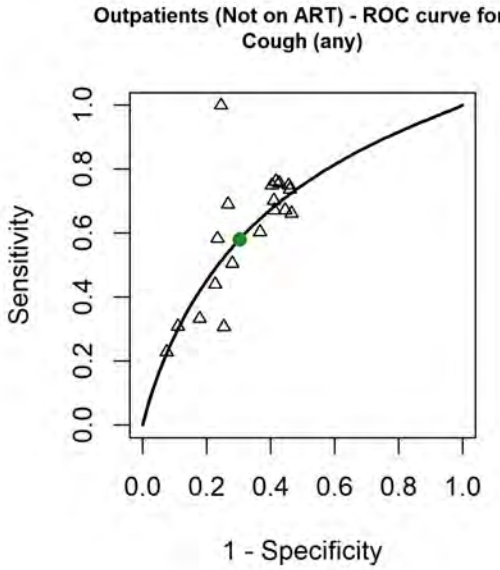


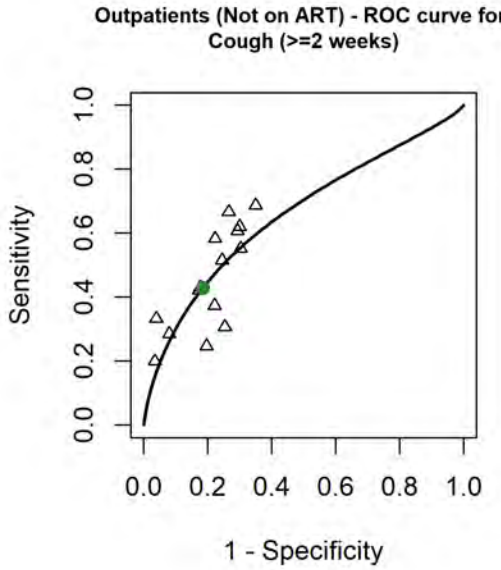


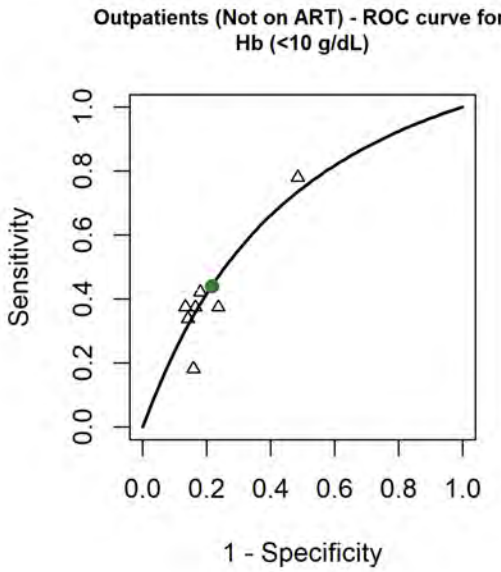


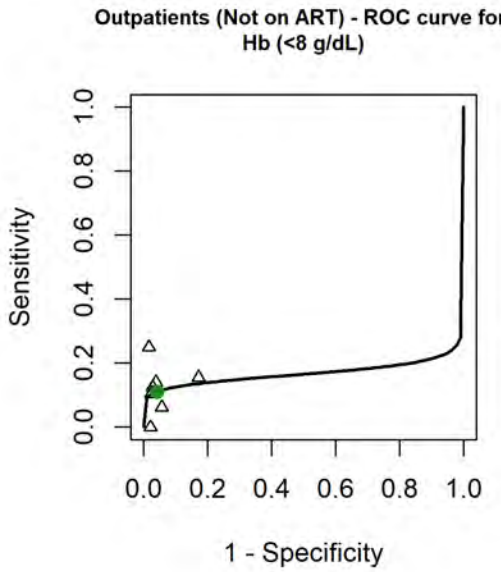

Outpatients (Not on ART) - ROC curve for CXR (abnormal)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

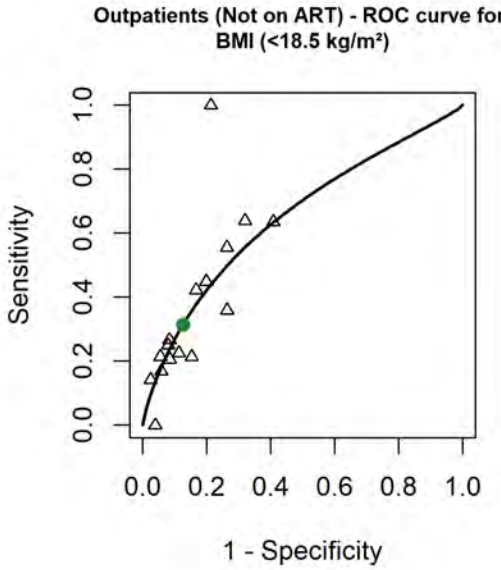

Outpatients (Not on ART) - ROC curve for CXR (suggests tuberculosis)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

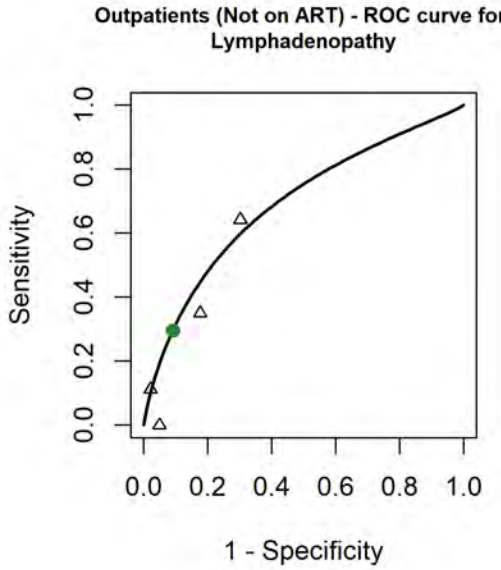

Outpatients (Not on ART) - ROC curve for Cough (any)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

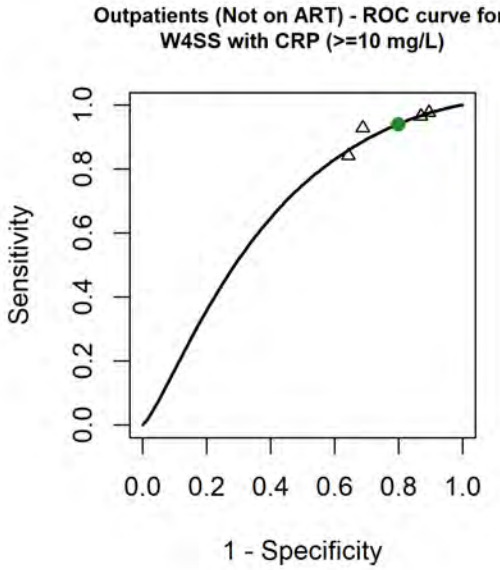

Outpatients (Not on ART) - ROC curve for Cough (>=2 weeks)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

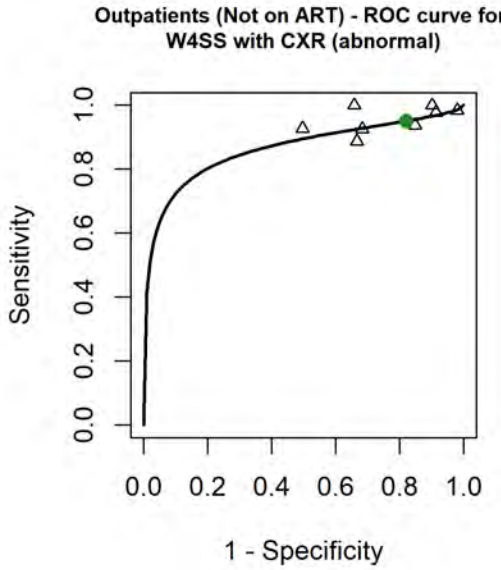

Outpatients (Not on ART) - ROC curve for Hb (<10 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

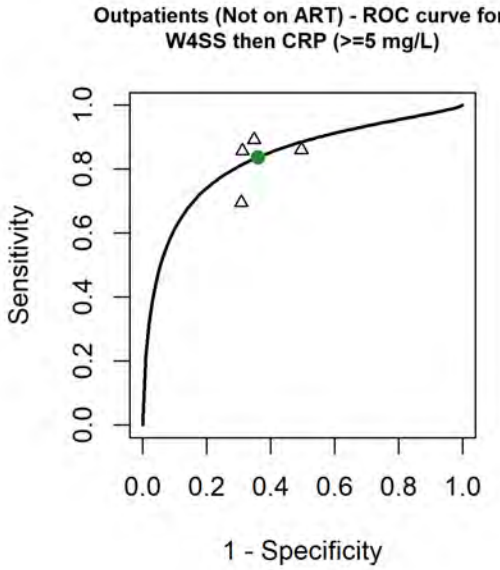

Outpatients (Not on ART) - ROC curve for Hb (<8 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

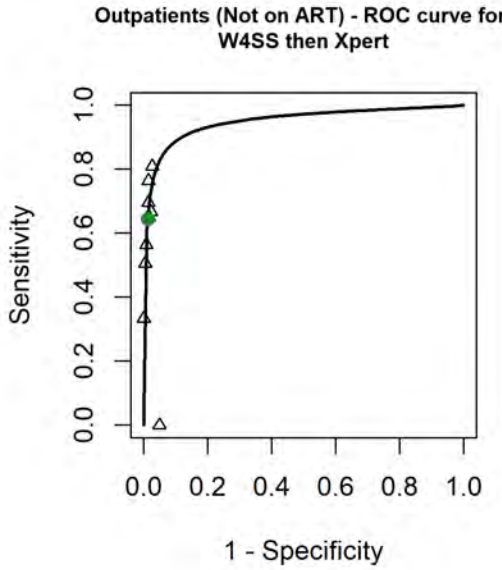

Outpatients (Not on ART) - ROC curve for BMI (<18.5 kg/m²)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Outpatients (Not on ART) - ROC curve for Lymphadenopathy
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Outpatients (Not on ART) - ROC curve for
W4SS with CRP (>=10 mg/L)
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0


Outpatients (Not on ART) - ROC curve for
W4SS with CXR (abnormal)
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0


Outpatients (Not on ART) - ROC curve for
W4SS then CRP (>=5 mg/L)
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0


Outpatients (Not on ART) - ROC curve for
W4SS then Xpert
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0

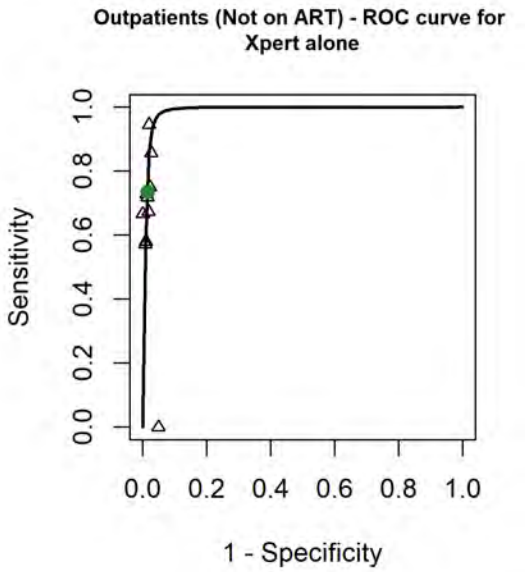
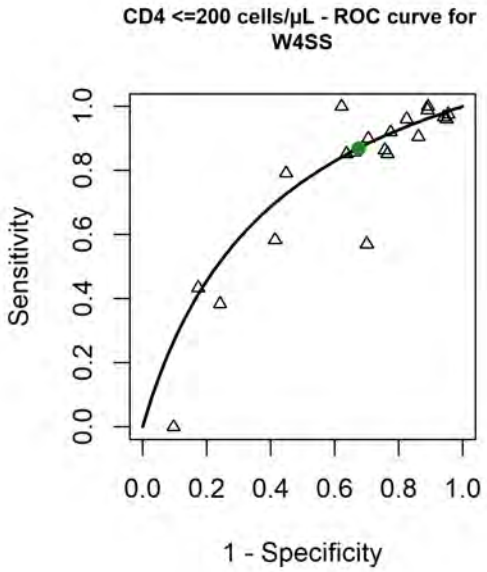
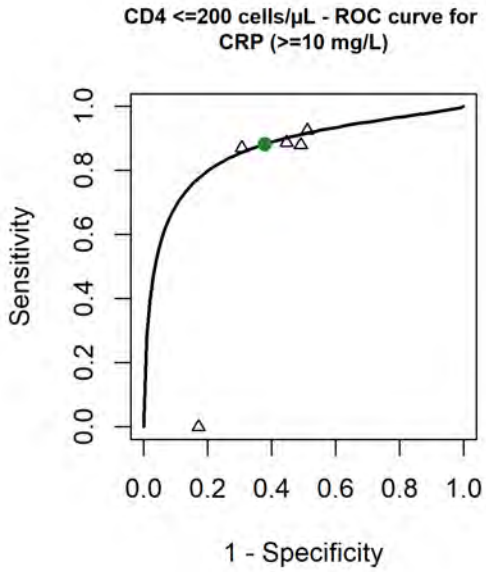
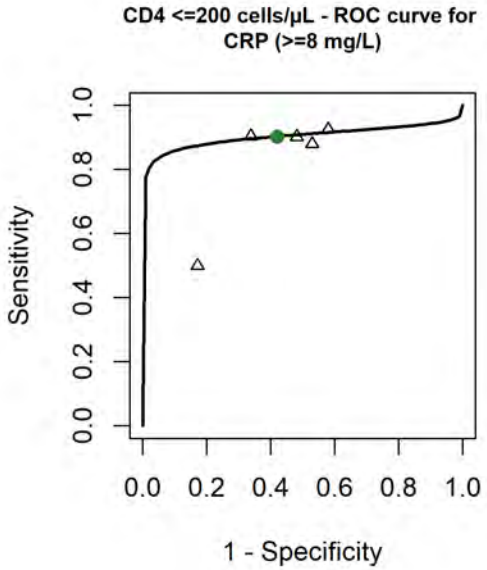
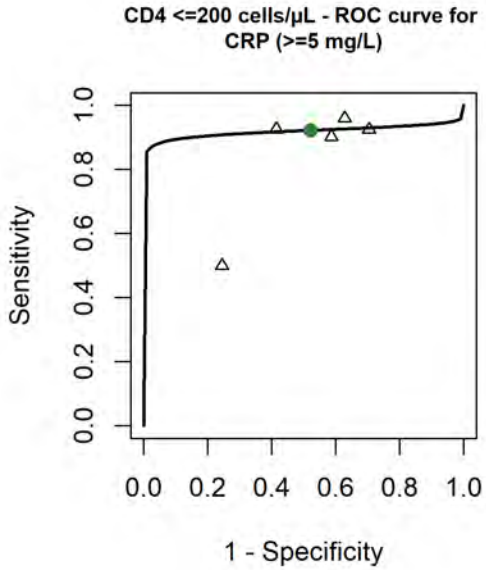

Outpatients (Not on ART) - ROC curve for Xpert alone
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity

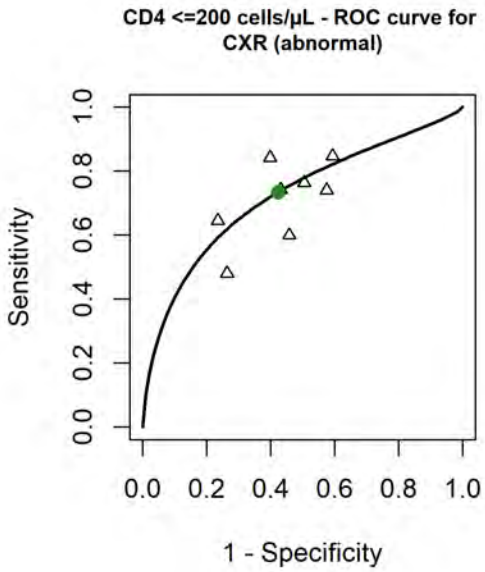
Figure 8-4D - Summary receiver operating characteristics curves comparing each test and WHO four-symptom screen for the detection of tuberculosis in participants with a CD4 cell count <=200 cells/µL

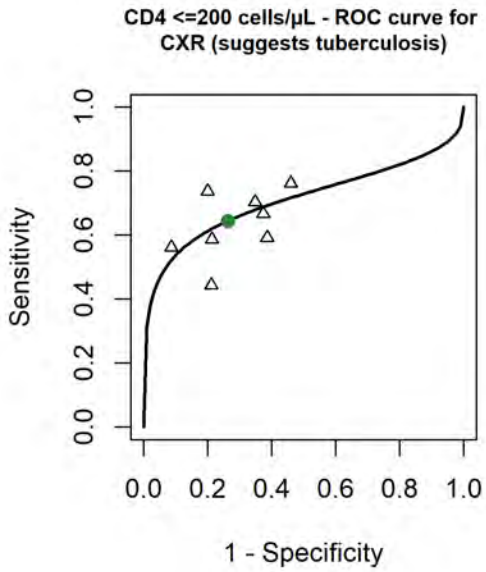


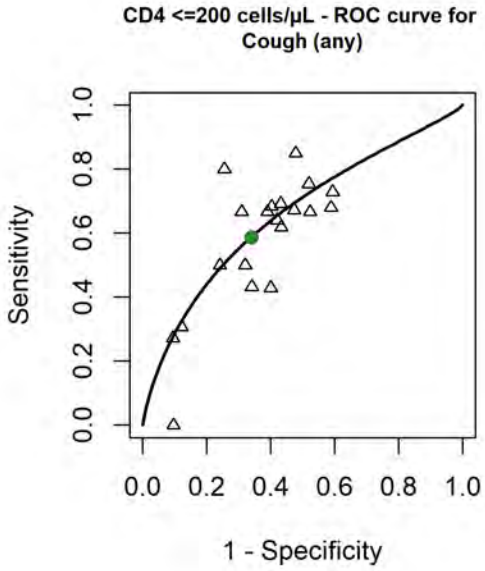


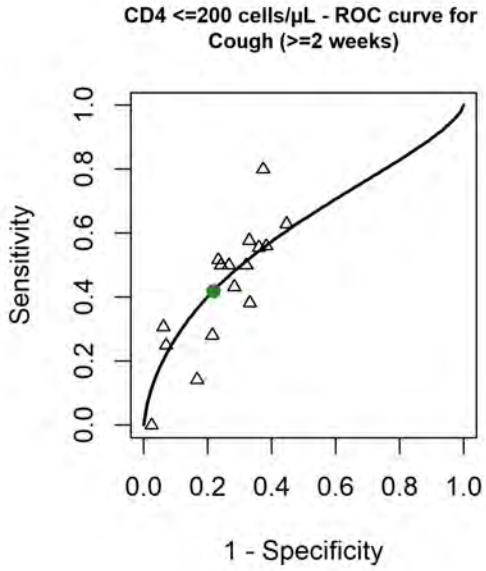


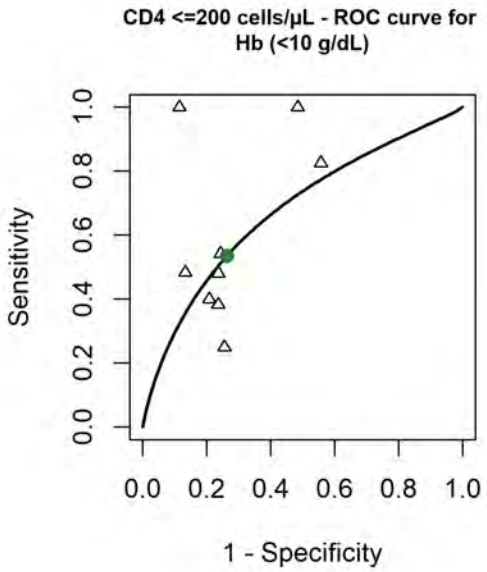


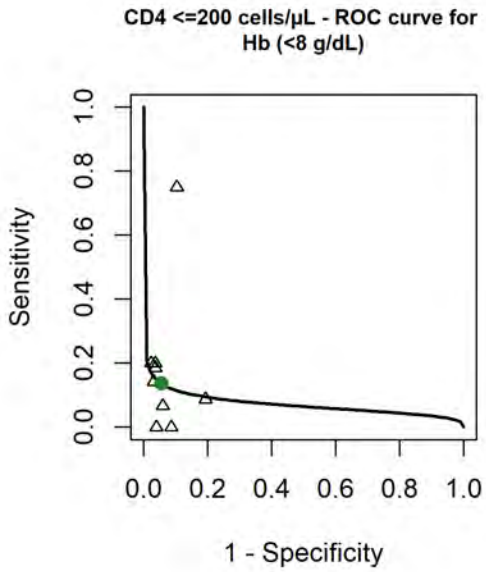

CD4 <=200 cells/µL - ROC curve for
CXR (abnormal)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

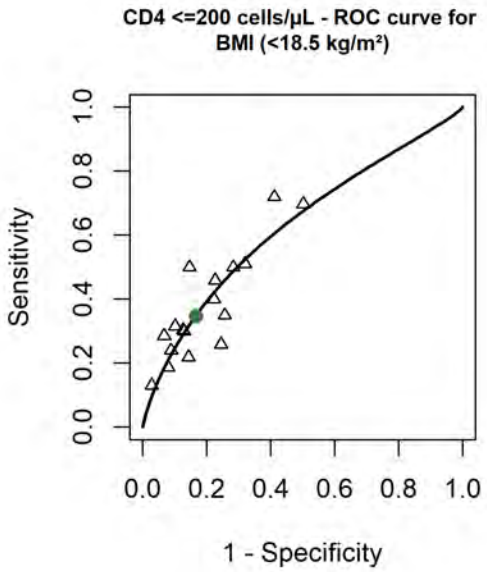

CD4 <=200 cells/µL - ROC curve for
CXR (suggests tuberculosis)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

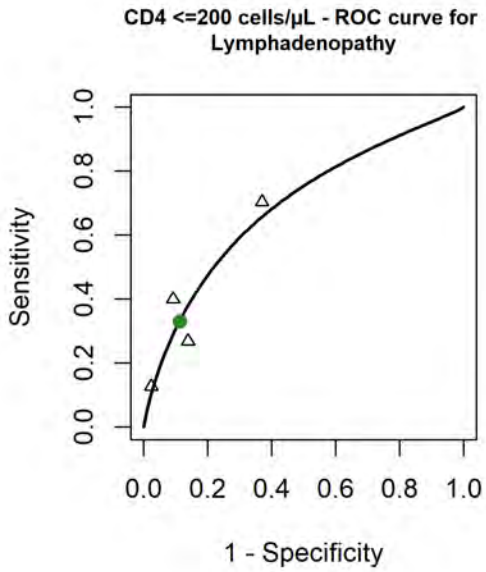

CD4 <=200 cells/µL - ROC curve for
Cough (any)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

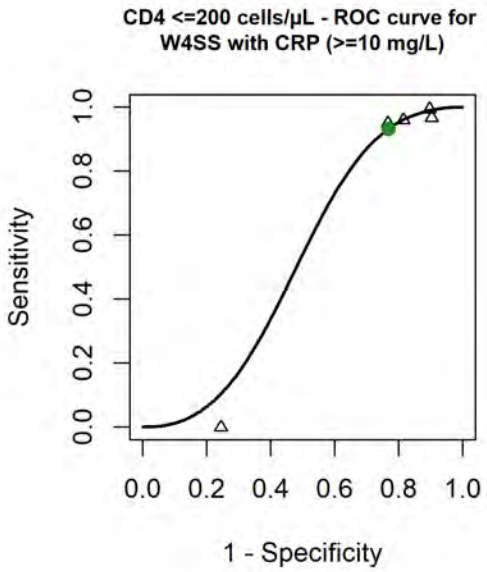

CD4 <=200 cells/µL - ROC curve for
Cough (>=2 weeks)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

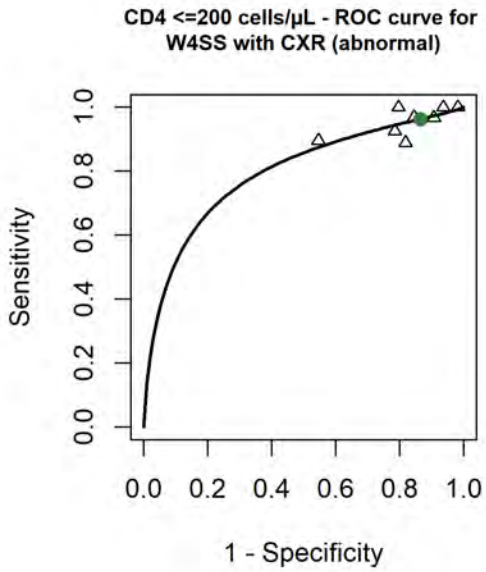

CD4 <=200 cells/µL - ROC curve for Hb (<10 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

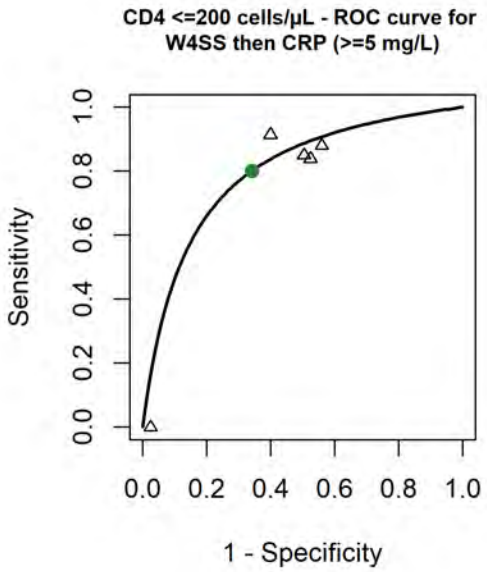

CD4 <=200 cells/µL - ROC curve for Hb (<8 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

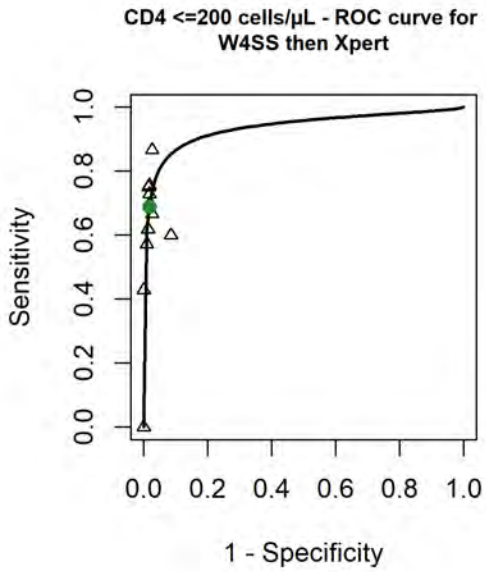

CD4 <=200 cells/µL - ROC curve for BMI (<18.5 kg/m²)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


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


CD4 <=200 cells/µL - ROC curve for W4SS with CRP (>=10 mg/L)
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0


CD4 <=200 cells/µL - ROC curve for W4SS with CXR (abnormal)
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0


CD4 <=200 cells/µL - ROC curve for W4SS then CRP (>=5 mg/L)
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0


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

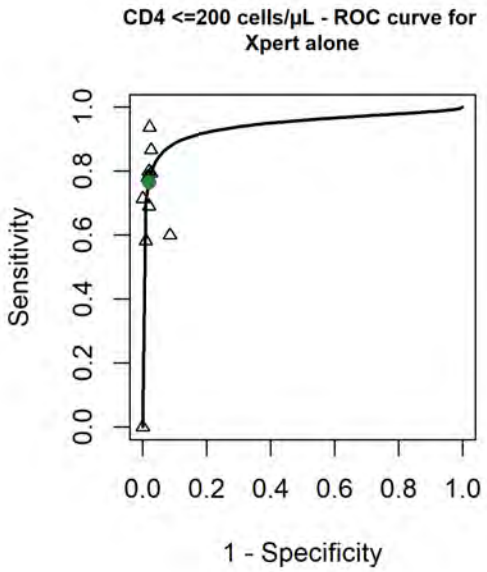
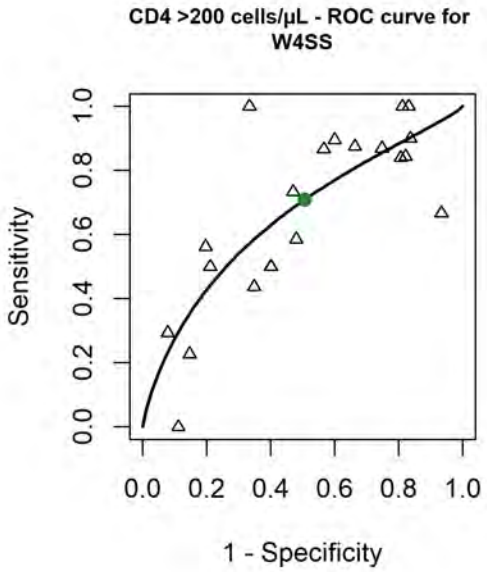
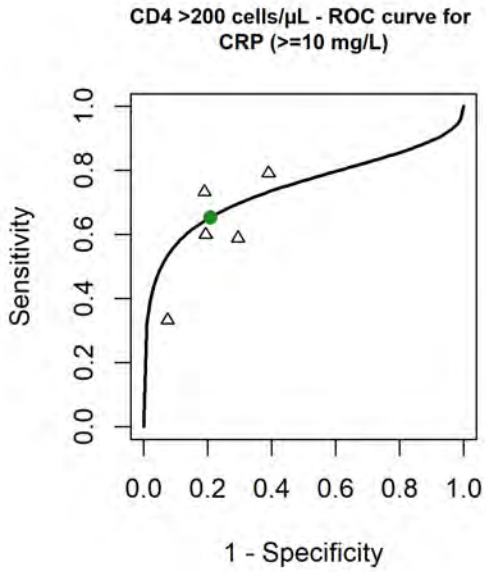
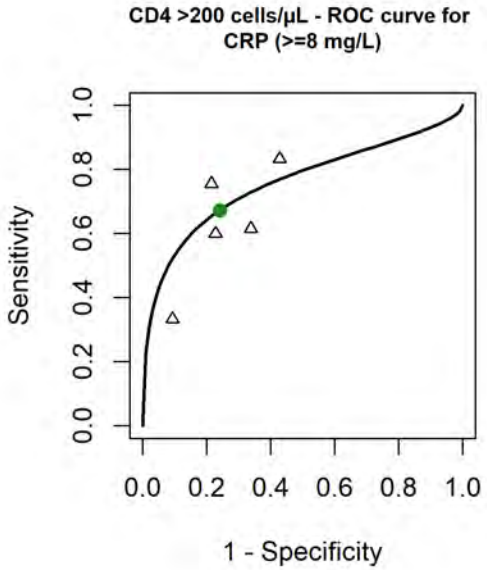
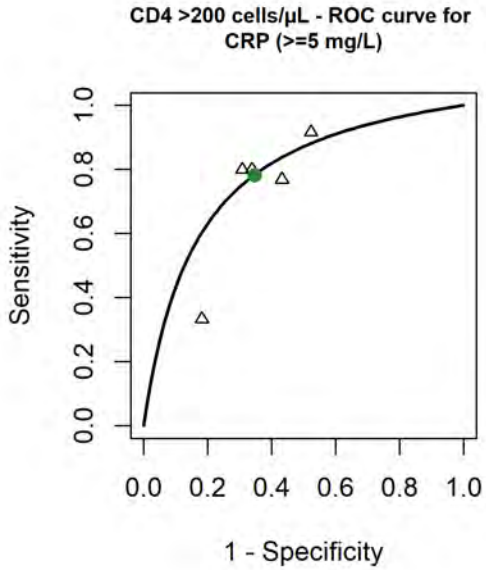

CD4 <=200 cells/µL - ROC curve for Xpert alone
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity

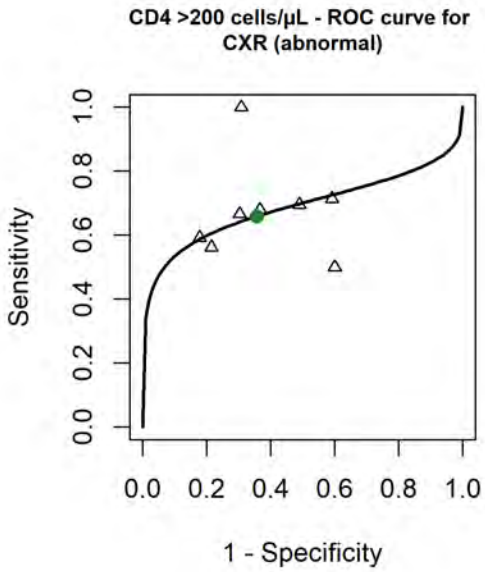
Figure 8-4E - Summary receiver operating characteristics curves comparing each test and WHO four-symptom screen for the detection of tuberculosis in in participants with a CD4 cell count >200 cells/µL

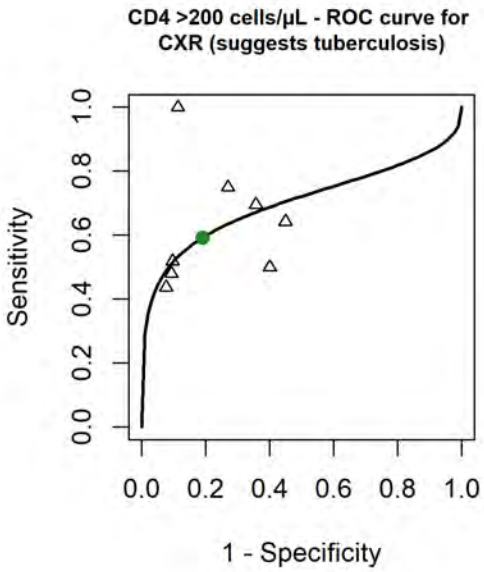


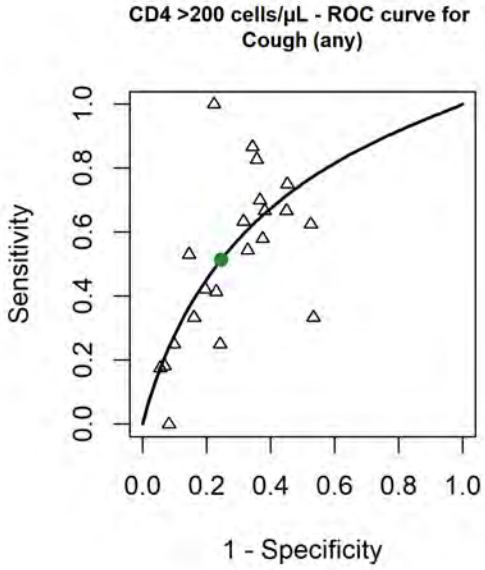


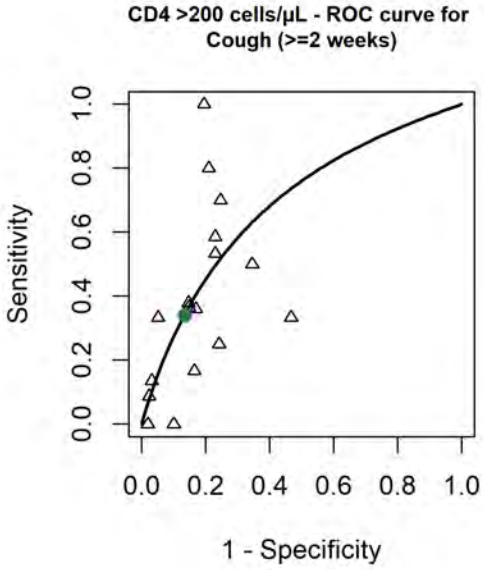


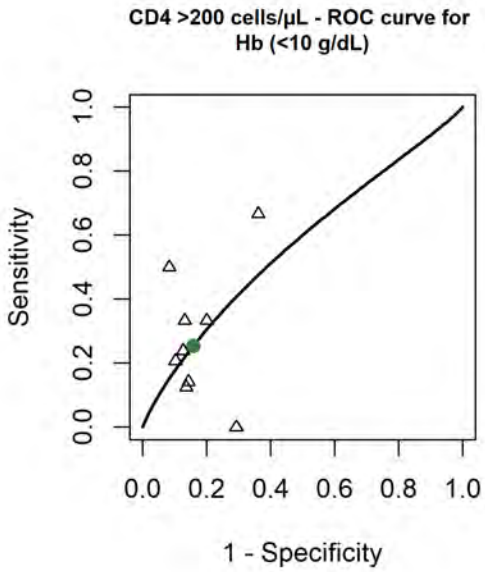


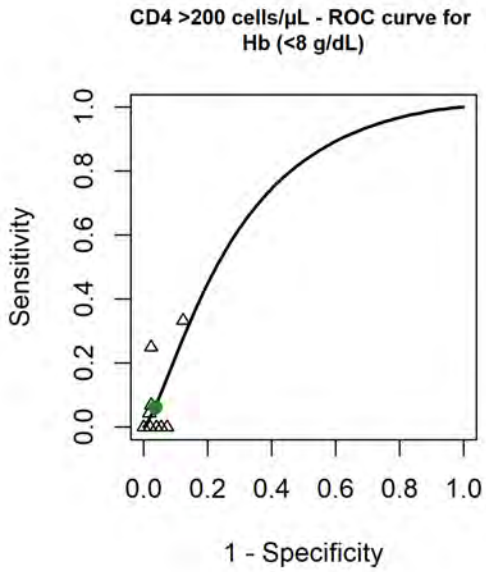

CD4 >200 cells/µL - ROC curve for CXR (abnormal)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

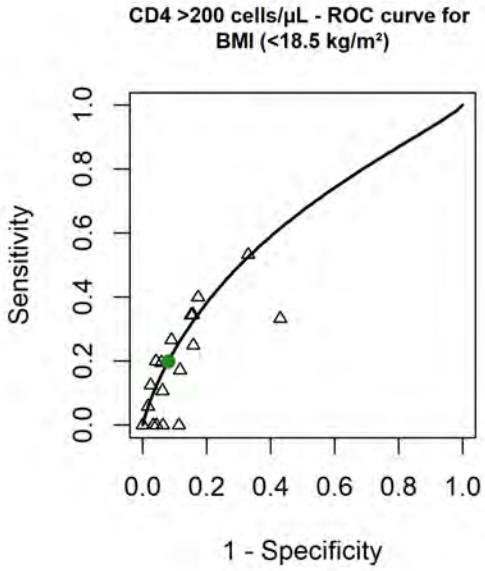

CD4 >200 cells/µL - ROC curve for CXR (suggests tuberculosis)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

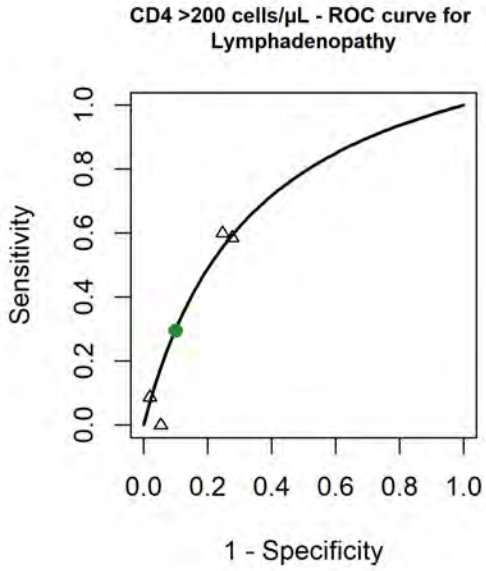

CD4 >200 cells/µL - ROC curve for Cough (any)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

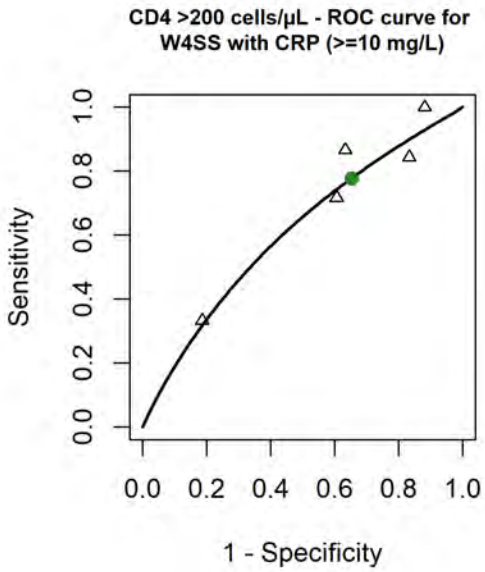

CD4 >200 cells/µL - ROC curve for Cough (>=2 weeks)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

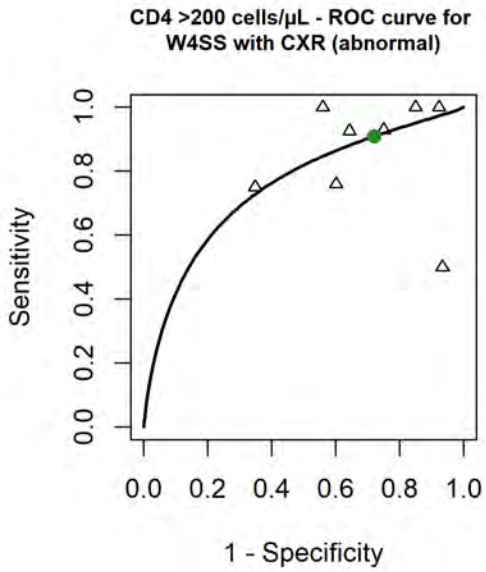

CD4 >200 cells/µL - ROC curve for
Hb (<10 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

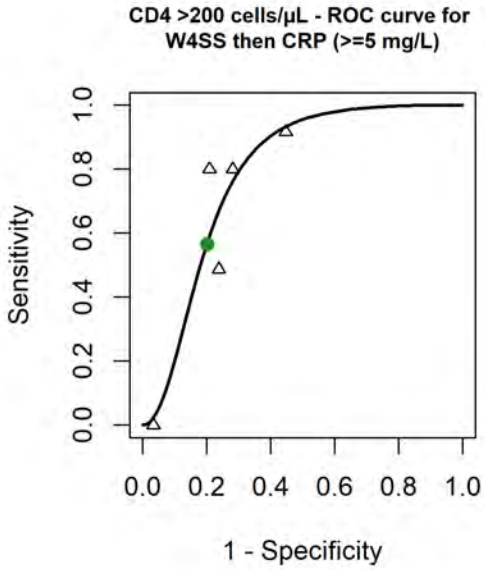

CD4 >200 cells/µL - ROC curve for
Hb (<8 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

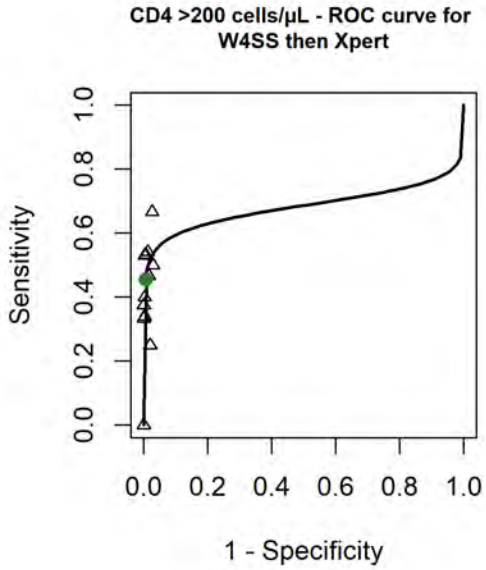

CD4 >200 cells/µL - ROC curve for
BMI (<18.5 kg/m²)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


CD4 >200 cells/µL - ROC curve for
Lymphadenopathy
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


CD4 >200 cells/µL - ROC curve for
W4SS with CRP (>=10 mg/L)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


CD4 >200 cells/µL - ROC curve for
W4SS with CXR (abnormal)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


CD4 >200 cells/µL - ROC curve for
W4SS then CRP (>=5 mg/L)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


CD4 >200 cells/µL - ROC curve for
W4SS then Xpert
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

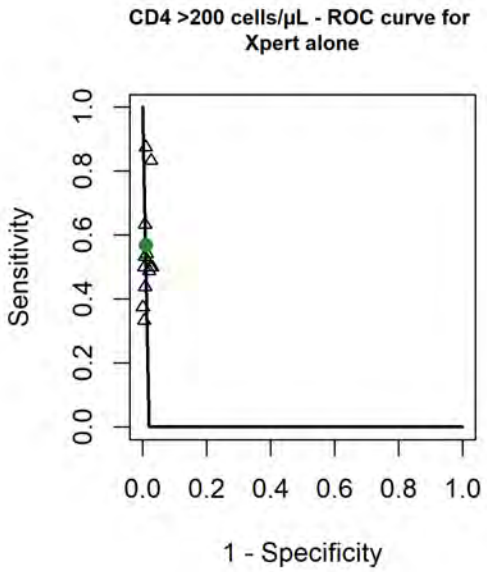
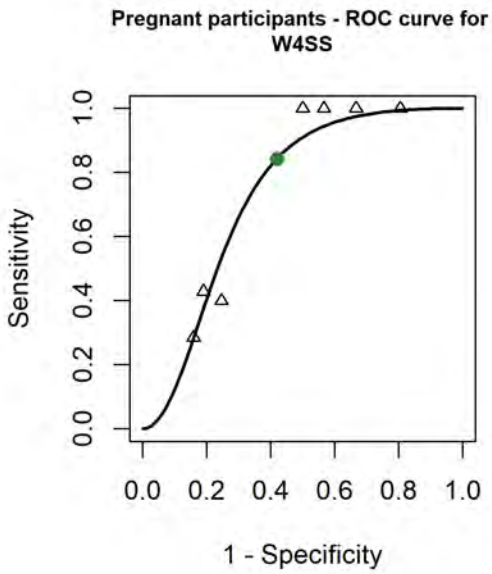
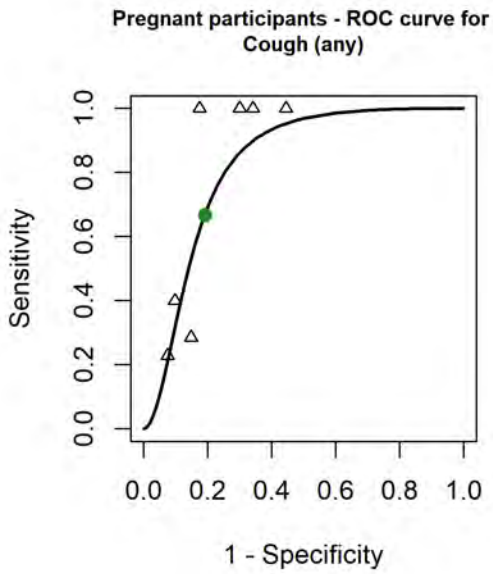
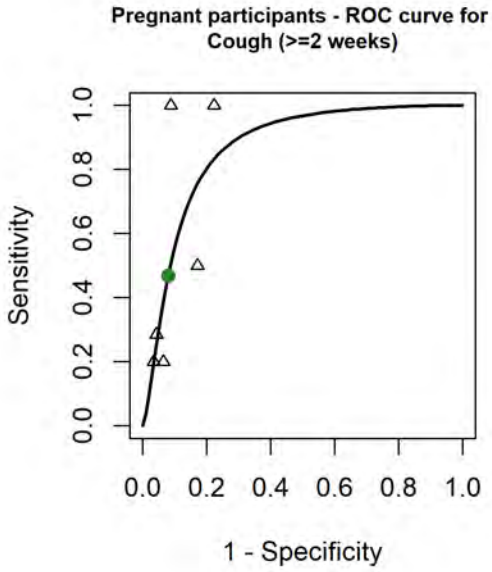
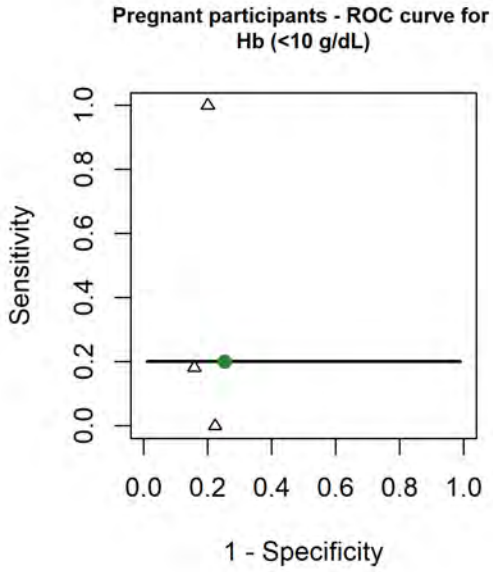

CD4 >200 cells/µL - ROC curve for Xpert alone
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity

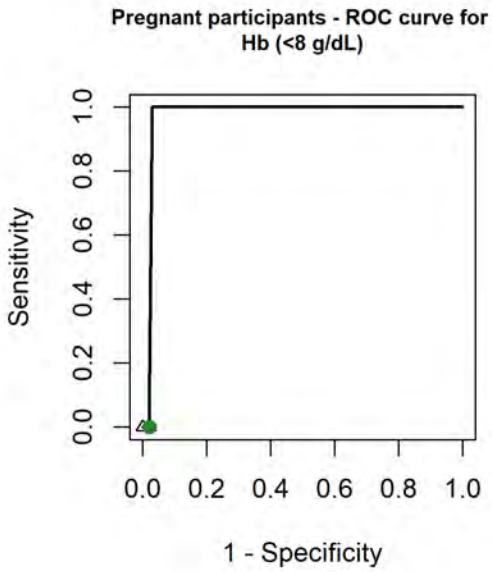
Figure 8-4F - Summary receiver operating characteristics curves comparing each test and WHO four-symptom screen for the detection of tuberculosis in pregnant participants

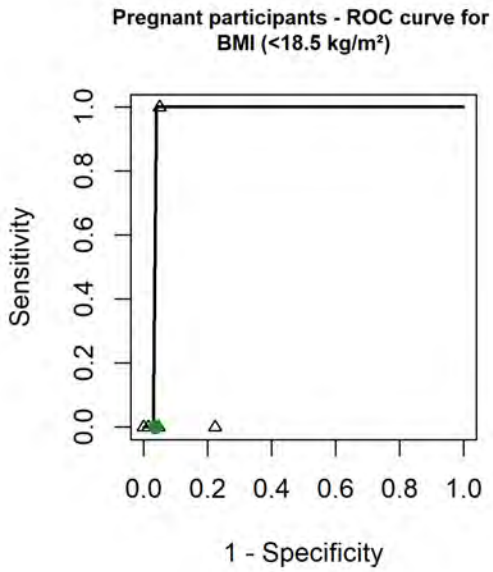


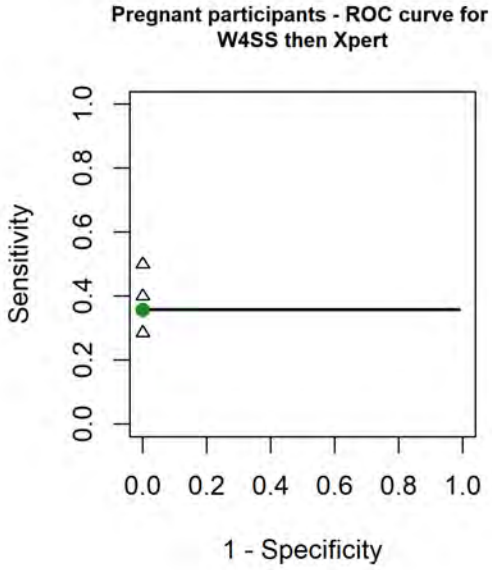







Pregnant participants - ROC curve for Hb (<8 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


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


Pregnant participants - ROC curve for W4SS then Xpert
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

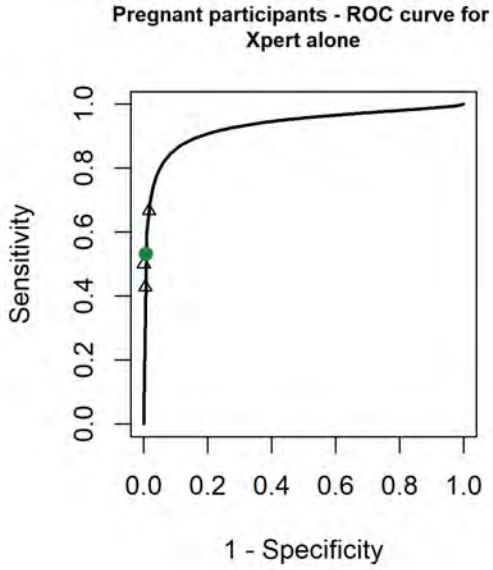
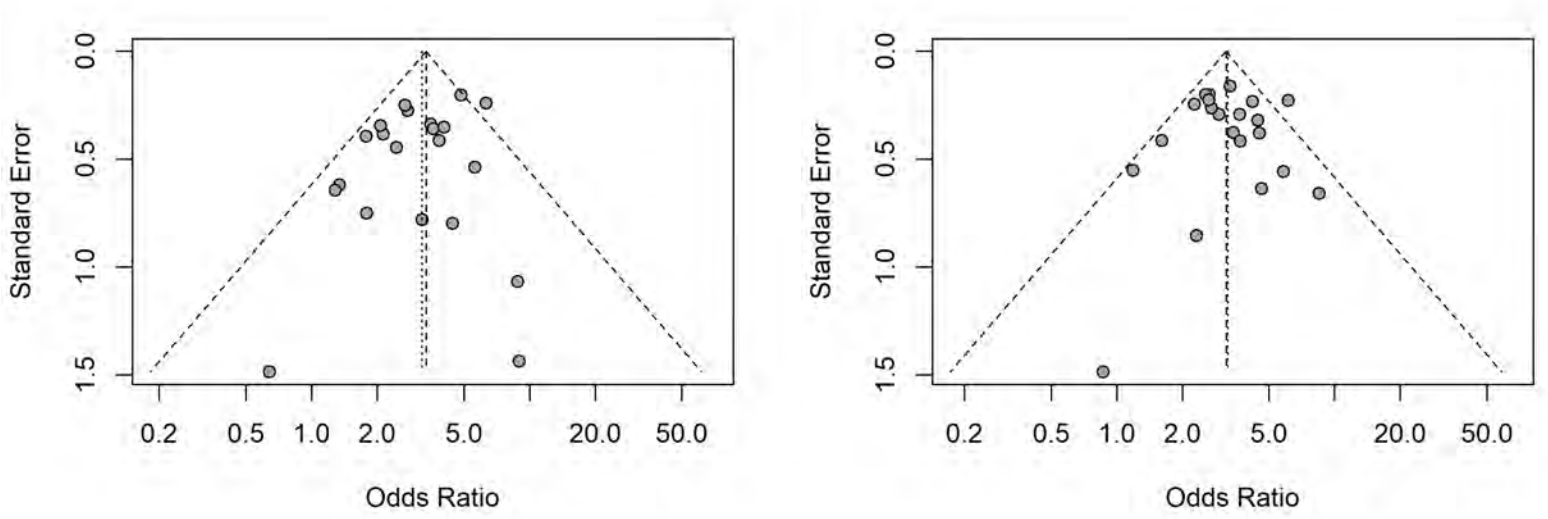
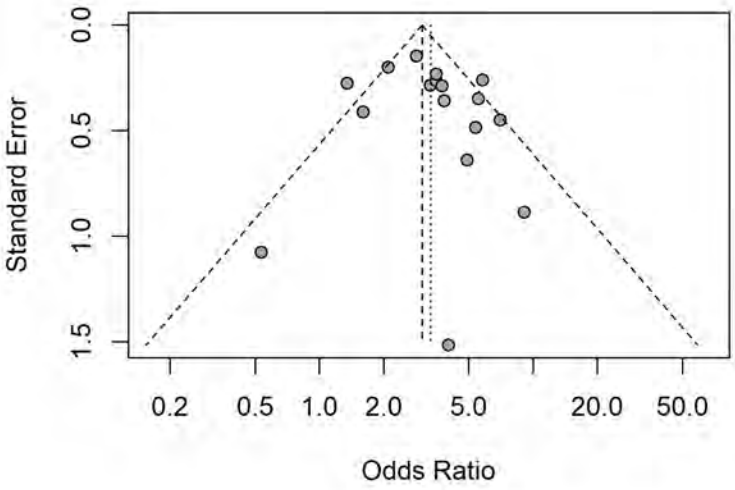
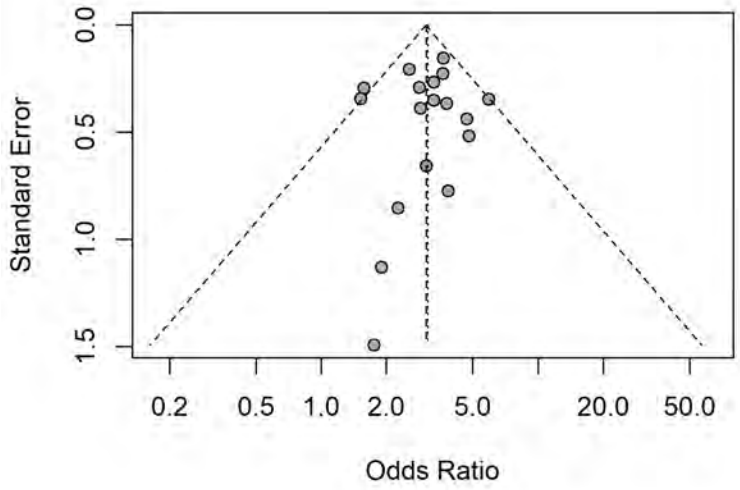

Pregnant participants - ROC curve for Xpert alone
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

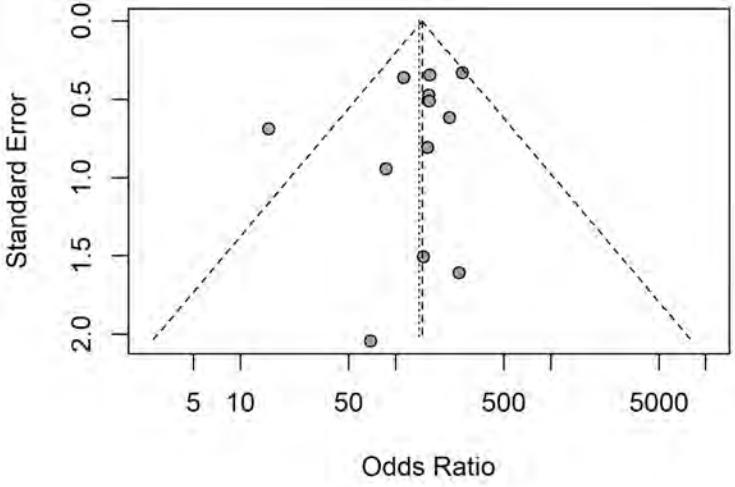
Figure 8-5: Funnel plots (for tests/strategies with >=10 studies available)

Figure 8-5A - Funnel plots in all participants




Standard Error
0.0
0.5
1.0
1.5
0.2
0.5
1.0
2.0
5.0
20.0
50.0
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
0.2
0.5
1.0
2.0
5.0
20.0
50.0
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
2.0
5
10
50
500
5000
Odds Ratio

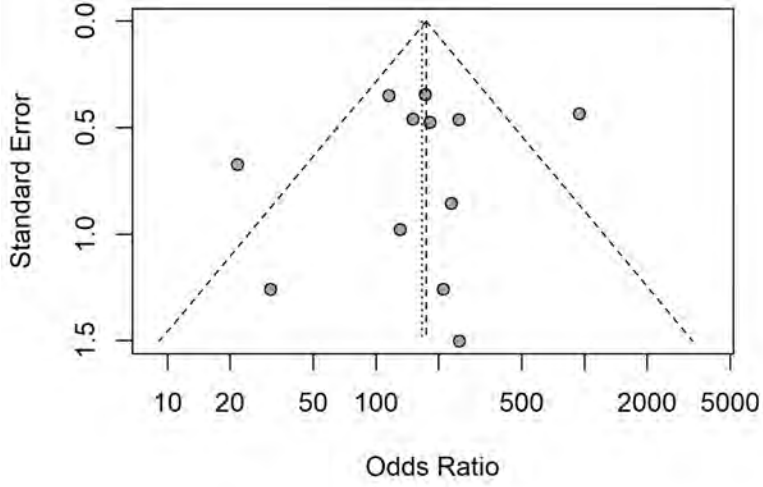
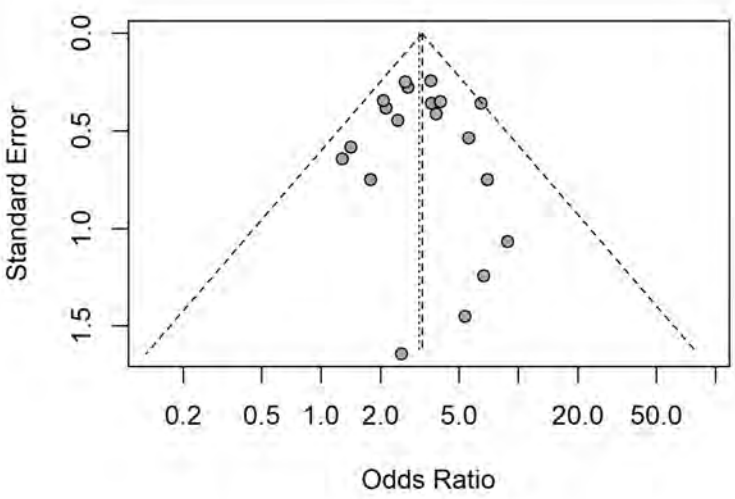
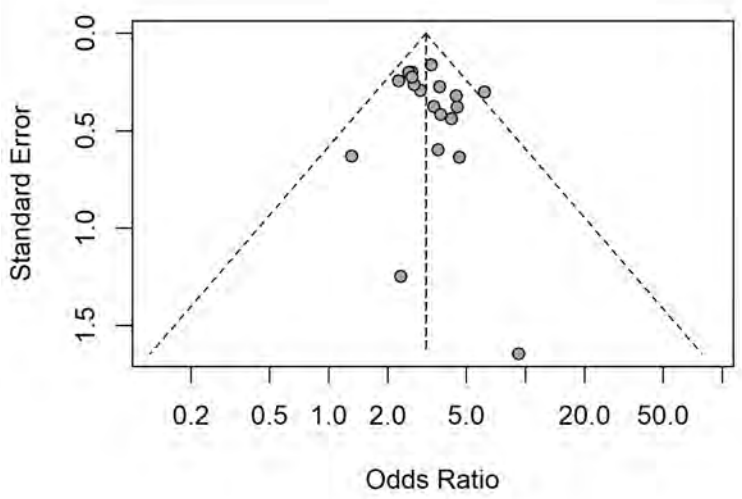
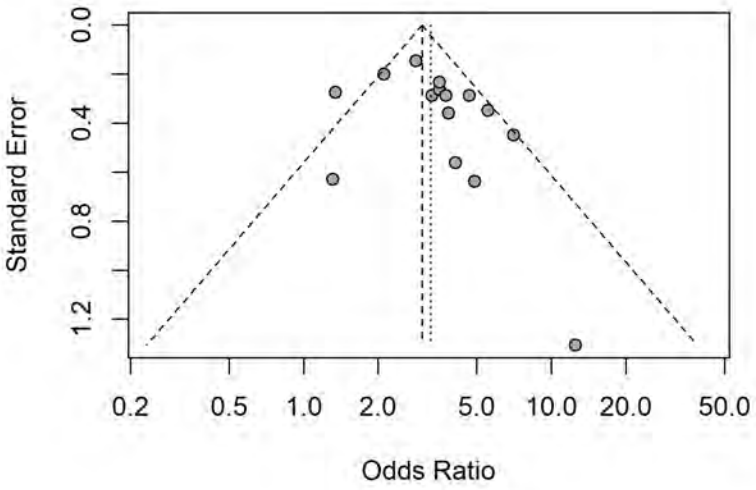
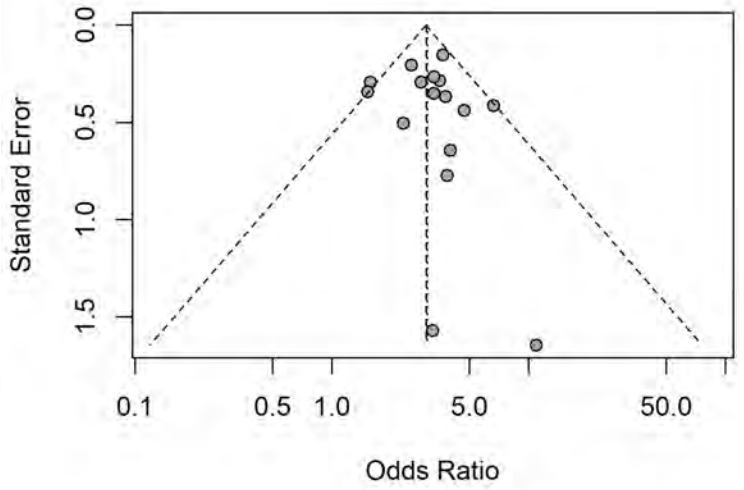

Standard Error
0.0
0.5
1.0
1.5
10
20
50
100
500
2000
5000
Odds Ratio

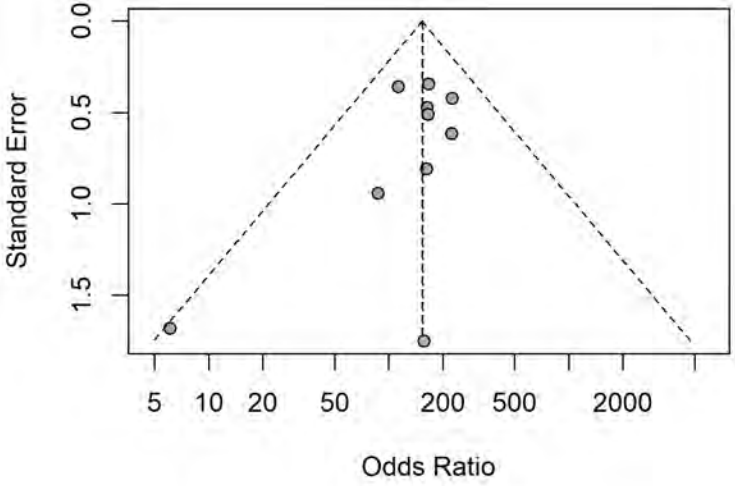
Figure 8-5B - Funnel plots in outpatients (on ART)
Figure 8-5C - Funnel plots in outpatients (not on ART)






Standard Error
0.0
0.4
0.8
1.2
0.2
0.5
1.0
2.0
5.0
10.0
20.0
50.0
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
0.1
0.5
1.0
5.0
50.0
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
5
10
20
50
200
500
2000
Odds Ratio

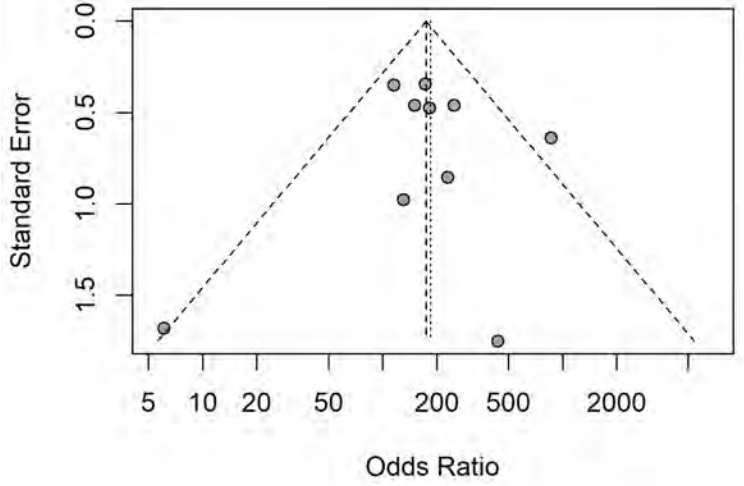
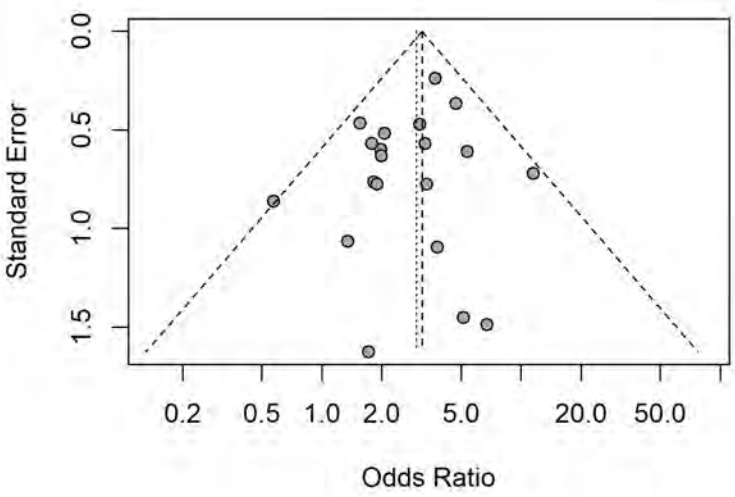
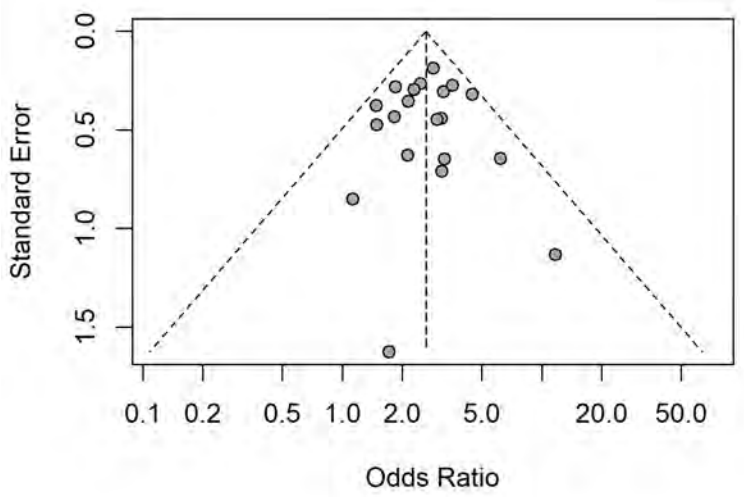
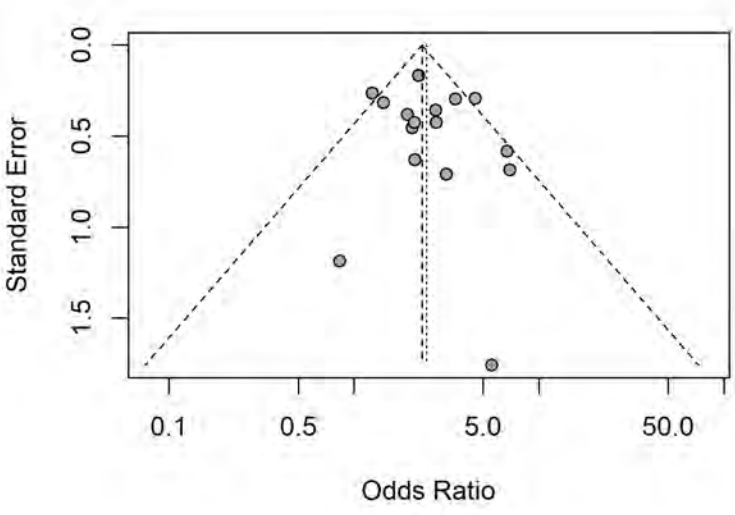
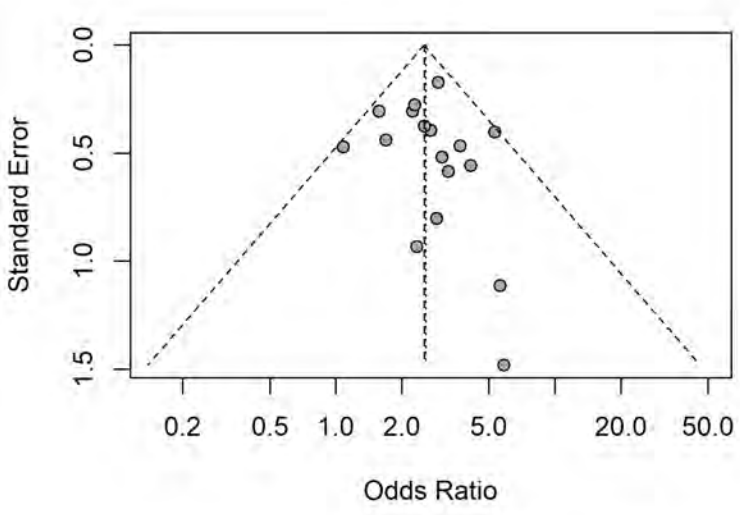

Standard Error
0.0
0.5
1.0
1.5
5
10
20
50
200
500
2000
Odds Ratio

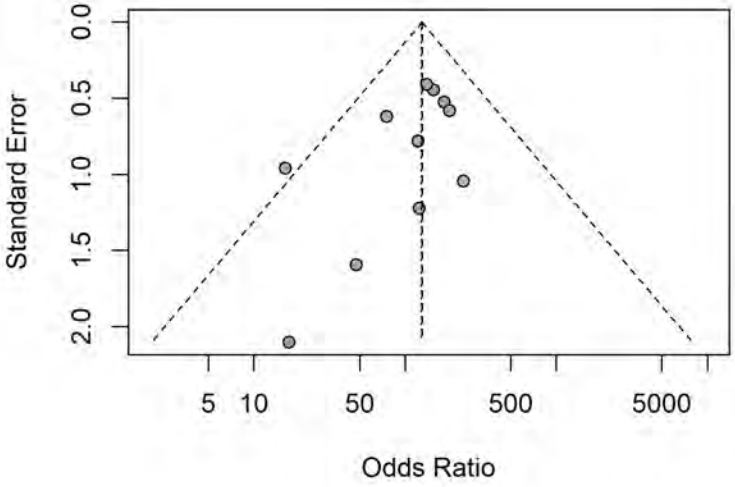
Figure 8-5D - Funnel plots in participants with a CD4 cell count <=200 cells/µL






Standard Error
0.0
0.5
1.0
1.5
0.1
0.5
5.0
50.0
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
0.2
0.5
1.0
2.0
5.0
20.0
50.0
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
2.0
5
10
50
500
5000
Odds Ratio

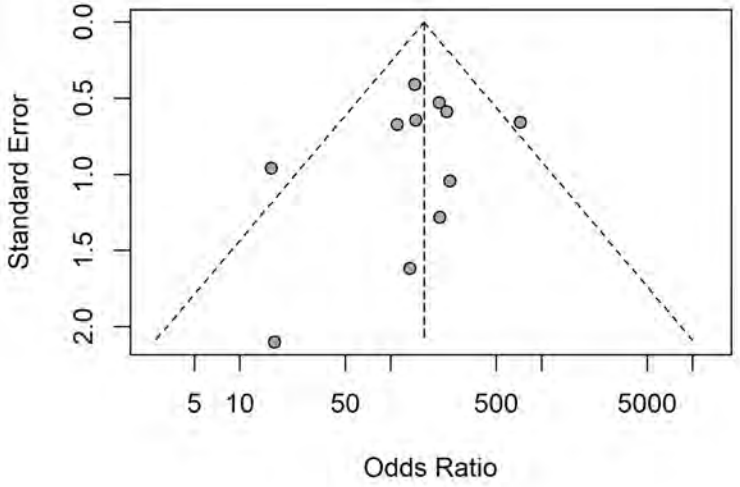
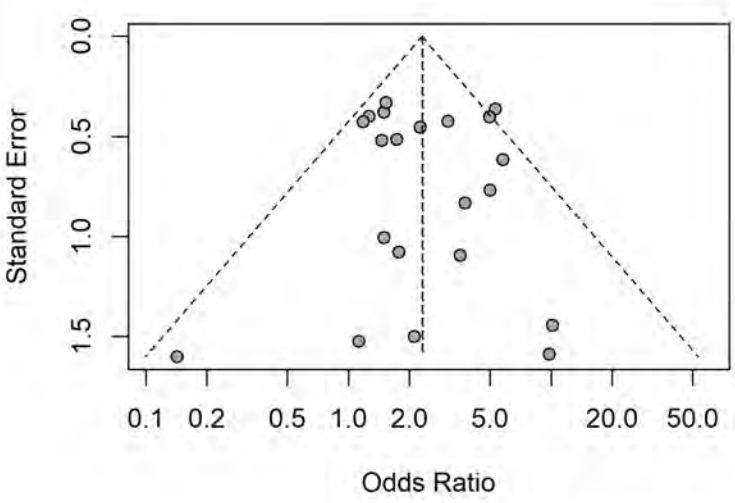
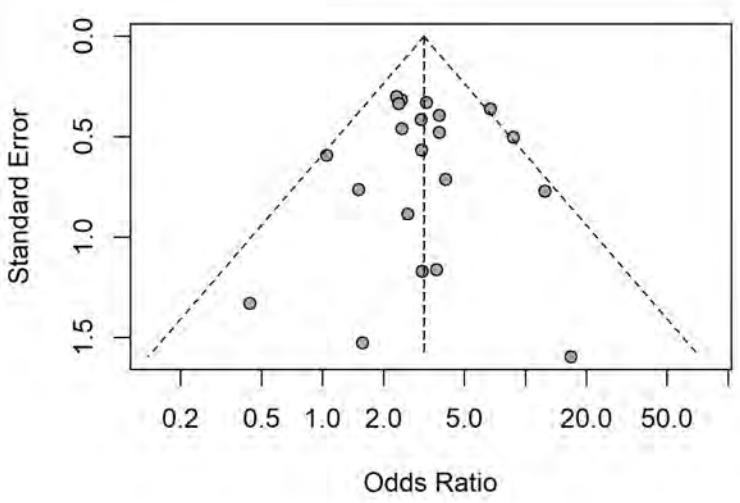
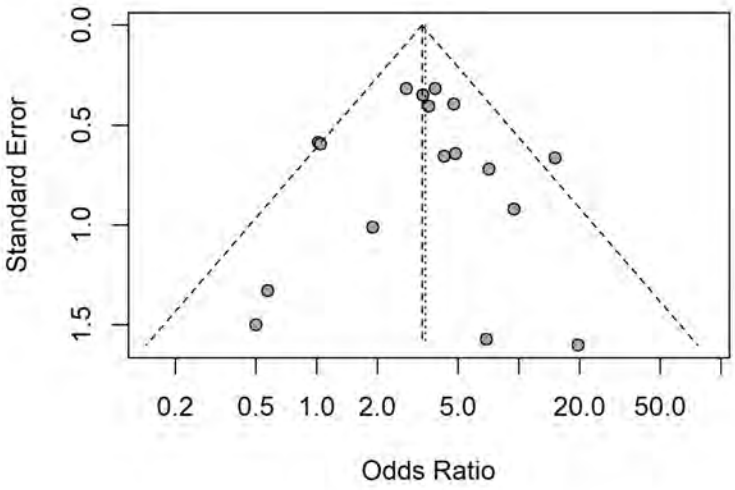
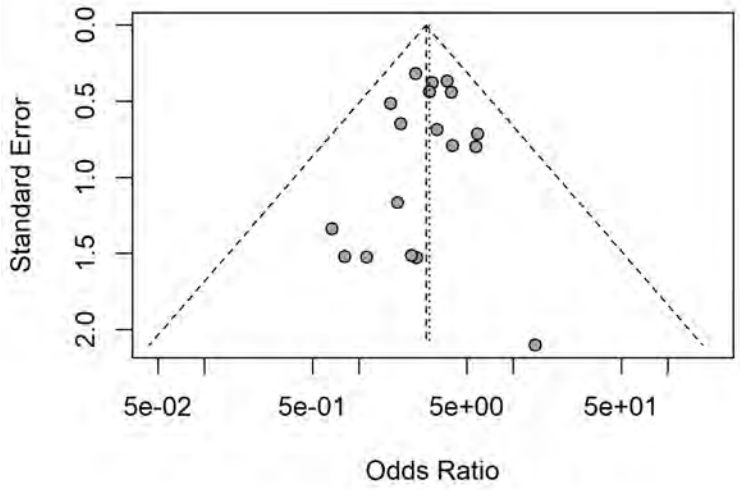

Standard Error
0.0
0.5
1.0
1.5
2.0
5
10
50
500
5000
Odds Ratio

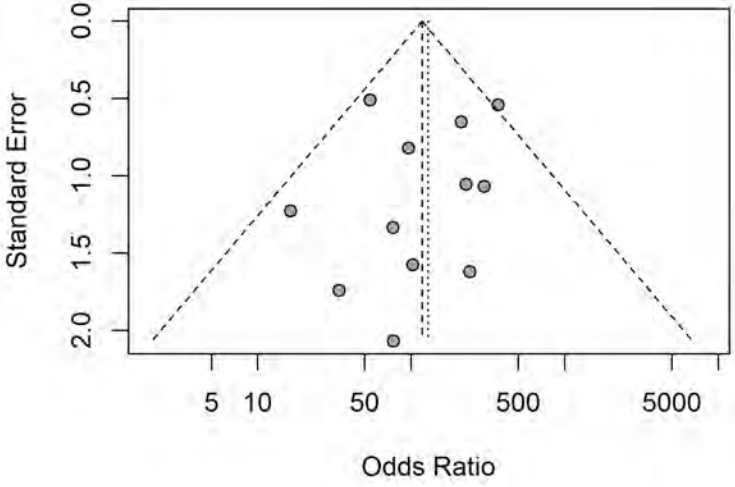
Figure 8-5E - Funnel plots in participants with a CD4 cell count >200 cells/µL

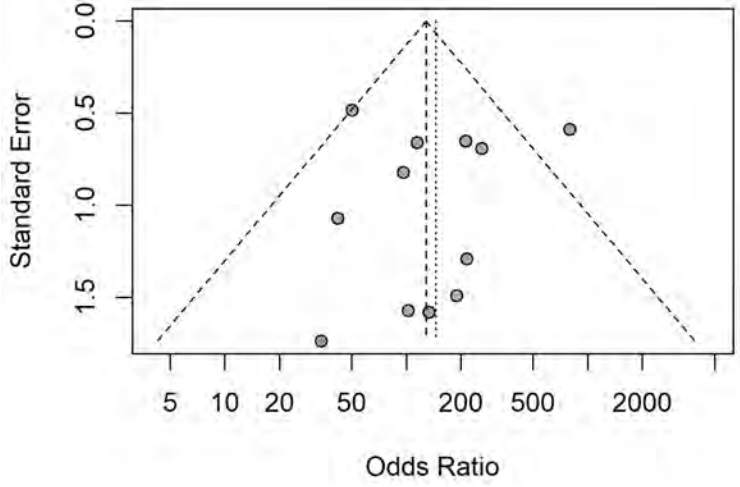





Standard Error
0.0
0.5
1.0
1.5
0.2
0.5
1.0
2.0
5.0
20.0
50.0
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
2.0
5e-02
5e-01
5e+00
5e+01
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
2.0
5
10
50
500
5000
Odds Ratio


Standard Error
0.0
0.5
1.0
1.5
5
10
20
50
200
500
2000
Odds Ratio

8.2 Appendix for Chapter 3

Table 8-13: 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 8-14: 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}
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}
sputum xpert result	{positive, negative, NA}, indeterminate = negative
sputum culture result	{positive, negative, NA}, contaminated culture = negative
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 8-15: 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
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	No
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
Heidebrecht, 2016[3]	South Africa	2013-2013	Inpatient PLHIV admitted to medical wards	1 hospital	≥3 doses of ATT	1 spot sample, induced if physician requested	1 spot sample, induced if physician requested	Both	No
Huerga, 2021[4]	Malawi	2015-2017	Inpatient PLHIV aged ≥15 years admitted to medical wards	1 hospital	On ATT	-	1 spot sample	-	-
Lawn, 2015[5]	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
Thit, 2017[6]	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

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

Table 8-16: Risk of bias results of studies that assessed proportion of HIV-positive inpatients eligible for Xpert

Domain	Gupta-Wright, 2018	Huerga, 2021	Lawn, 2015	Heidebrecht, 2016	Bjerrum, 2015	Thit, 2017
1. Was the sample frame appropriate to address the target population?¶	Yes	Yes	Yes	Yes	No¶¶	Yes
2. Were study participants recruited in an appropriate way?§	Yes	Yes	Yes	Yes	Yes	Yes
3. Were the study subjects and setting described in detail?*	Yes	Yes	Yes	Yes	Yes	Yes
4. Were valid methods used for the identification of eligibility criteria?#	Yes	Yes	Yes	Yes	Yes	Yes
5. Was the response rate adequate (>80%)?	No	Yes	Yes	No	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, WHO stage 3/4 or 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 (i.e., WHO four-symptom screen) assessed based on existing definitions or diagnostic criteria?

Definition of abbreviations: ART = antiretroviral therapy

Table 8-17: Risk of bias and applicability results on the QUADAS-2 criteria tool among studies with culture-based reference standard*

Domain	Bjerrum, 2015	Heidebrecht, 2016	Lawn, 2015	Thit, 2017
Patient selection (Risk of Bias)¶	Low	Low	Low	Low
Index test (Risk of Bias)¶¶	Low	Low	Low	Low
Reference test (Risk of Bias)§	High	High	Low	High
Flow and timing (Risk of Bias)#	Low	Low	High	Low
Patient selection (Applicability)†	High	Low	Low	Low
Index test (Applicability)†	Low	Low	Low	Low
Reference test (Applicability)†	Low	Low	Low	Low

*Assessment done for all index tests

¶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? 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 8-18: Percentage of missing data for each variable by study†§

Variable†	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Heidebrecht	Huerga	Lawn	Thit
Clinical setting	0	0	0	0	0	0	0
Age	0	0	0	0	0	0	0
Sex	0	0	0	1	0	0	0
ART status	0	0	0	2	4	0	0
History of tuberculosis	0	0	0	2	100	0	0
W4SS*	0	0	0	0	0	0	0
Cough	1	0	0	1	0	1	0
Fever	4	0	0	0	1	1	0
Weight loss	1	0	0	0	1	0	0
Night sweats	0	0	0	2	0	1	0
Cough >=2 weeks	1	1	1	6	100	1	100
BMI	14	0	0	100	28	100	2
Lymphadenopathy	0	100	100	100	1	100	0
CD4 count	10	0	1	100	2	0	0
CRP	100	100	100	100	100	5	100
Haemoglobin	6	0	0	100	1	1	11
CXR (abnormal)**	100	100	100	100	57	100	4
Sputum Xpert***	28	35	39	6	39	54	0
Non-sputum Xpert	100	1	100	100	100	2	100
Total Xpert***	28	1	39	6	39	1	0
Sputum culture	0	100	100	16	100	50	0
Non-sputum culture	100	100	100	100	100	0	100
Total culture	0	100	100	16	100	0	0
Total (culture or Xpert)	0	1	39	0	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 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

**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

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 8-19: Summary of main characteristics for participants overall and by each study

Variable†	All	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Heidebrecht	Huerga	Lawn	Thit
Frequency	3660 (100)	69 (2)	1287 (35)	1287 (35)	156 (4)	387 (11)	420 (11)	54 (1)
Age (years)	37 (31-45)	37 (32-43)	38 (31-46)	38 (31-46)	36 (28-44)	38 (32-45)	36 (29-42)	33 (30-44)
N	3660	69	1287	1287	156	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	2104 (58)	33 (48)	727 (56)	734 (57)	112 (72)	216 (56)	255 (61)	27 (50)
N	3659	69	1287	1287	155	387	420	54
On ART	2445 (67)	0 (0)	926 (72)	935 (73)	82 (54)	305 (82)	175 (42)	22 (41)
N	3642	69	1287	1287	153	372	420	54
History of TB	902 (28)	5 (7)	335 (26)	309 (24)	46 (30)	-	190 (45)	17 (31)
N	3268	69	1287	1287	153	-	418	54
Current Smoker	293 (11)	1 (2)	151 (12)	128 (10)	-	-	-	13 (24)
N	2693	65	1287	1287	-	-	-	54
W4SS*	3306 (90)	69 (100)	1152 (90)	1164 (90)	144 (92)	349 (90)	382 (91)	46 (85)
N	3658	69	1287	1287	156	387	418	54
Cough	1945 (53)	48 (71)	651 (51)	681 (53)	111 (72)	230 (59)	199 (48)	25 (46)
N	3655	68	1287	1287	155	387	417	54
Fever	1969 (54)	46 (70)	753 (59)	747 (58)	98 (63)	228 (59)	62 (15)	35 (65)
N	3652	66	1287	1287	156	385	417	54
Weight loss	2638 (72)	65 (96)	906 (70)	875 (68)	117 (75)	277 (73)	356 (85)	42 (78)
N	3651	68	1287	1286	156	382	418	54
Night sweats	1490 (41)	29 (42)	497 (39)	540 (42)	76 (50)	154 (40)	171 (41)	23 (43)
N	3652	69	1287	1286	153	386	417	54
Cough >= 2 weeks	765 (24)	35 (51)	342 (27)	321 (25)	34 (23)	-	33 (8)	-
N	3172	68	1271	1270	147	-	416	-
Lymphadenopathy	58 (11)	8 (12)	-	-	-	42 (11)	-	8 (15)
N	508	69	-	-	-	385	-	54
CXR (abnormal)	130 (59)	-	-	-	-	100 (60)	-	30 (58)
N	220	-	-	-	-	168	-	52
Sputum Xpert +	305 (13)	9 (18)	85 (10)	82 (11)	35 (24)	33 (14)	57 (29)	4 (7)
N	2291	50	832	779	146	235	195	54
Non-sputum Xpert +	163 (10)	-	74 (6)	-	-	-	89 (22)	-
N	1681	-	1270	-	-	-	411	-
Total Xpert +§	401 (14)	9 (18)	122 (10)	82 (11)	35 (24)	33 (14)	116 (28)	4 (7)
N	2957	50	1279	779	146	235	414	54
Sputum culture +	106 (23)	13 (19)	-	-	31 (24)	-	58 (28)	4 (7)
N	463	69	-	-	131	-	209	54
Non-sputum culture +	70 (17)	-	-	-	-	-	70 (17)	-
N	420	-	-	-	-	-	420	-
Total culture +¶	157 (23)	13 (19)	-	-	31 (24)	-	109 (26)	4 (7)

Variable†	All	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Heidebrecht	Huerga	Lawn	Thit
N	674	69	-	-	131	-	420	54
Total Xpert & culture +	439 (15)	15 (22)	122 (10)	82 (11)	41 (26)	33 (14)	139 (33)	7 (13)
N	2992	69	1279	779	156	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 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

§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 8-20: Direct comparisons of the diagnostic accuracy (pooled sensitivity and specificity) between each screening test/strategy and WHO four-symptom screen for the detection of tuberculosis

Table 8-20A: Direct comparisons of the diagnostic accuracy (pooled sensitivity and specificity) between each screening test/strategy and WHO four-symptom screen for the detection of tuberculosis in all participants using culture as reference standard†

	Index Test				W4SS			
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)
CRP (>=10 mg/L)	1	399	97 (91-99)	21 (17-26)	1	399	97 (91-99)	10 (7-14)
CRP (>=8 mg/L)	1	399	97 (91-99)	18 (14-23)	1	399	97 (91-99)	10 (7-14)
CRP (>=5 mg/L)	1	399	98 (92-100)	12 (9-17)	1	399	97 (91-99)	10 (7-14)
CXR (abnormal)	1	52	75 (24-97)	44 (31-58)	1	52	90 (33-99)	17 (9-31)
Cough (any)	4	669	79 (58-91)	43 (31-57)	4	669	98 (93-100)	8 (4-14)
Cough (>=2 weeks)	3	608	28 (16-46)	80 (52-93)	3	608	98 (92-99)	5 (1-17)
Hb (<10 g/dL)	3	525	78 (70-84)	40 (26-57)	3	525	98 (93-99)	8 (4-17)
Hb (<8 g/dL)	3	525	55 (46-64)	67 (50-81)	3	525	98 (93-99)	8 (4-17)
BMI (<18.5 kg/m²)§	2	112	57 (32-79)	62 (44-78)	2	112	100 (85-100)	6 (2-18)
Lymphadenopathy§	2	123	12 (3-37)	87 (68-96)	2	123	100 (83-100)	6 (2-18)
W4SS or CRP (>=10 mg/L)¶	1	399	100 (93-100)	5 (3-8)	1	399	97 (91-99)	10 (7-14)
W4SS or CXR (abnormal)¶	1	52	90 (33-99)	7 (3-19)	1	52	90 (33-99)	17 (9-31)
W4SS then CRP (>=5 mg/L)¶	1	399	95 (89-98)	20 (16-25)	1	399	97 (91-99)	10 (7-14)

†Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

§We computed binomial 95% CIs for W4SS sensitivity as all studies had 100% sensitivity

¶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

Table 8-20B: Direct comparisons of the diagnostic accuracy (pooled sensitivity and specificity) between each screening test/strategy and WHO four-symptom screen for the detection of tuberculosis in all participants using Xpert as reference standard†

	Index Test				W4SS			
	No of studies	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**No of studies**	**N**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
CRP (>=10 mg/L)	1	394	94 (87-97)	20 (16-25)	1	394	97 (92-99)	11 (8-15)
CRP (>=8 mg/L)	1	394	94 (87-97)	17 (14-22)	1	394	97 (92-99)	11 (8-15)
CRP (>=5 mg/L)	1	394	96 (91-99)	12 (9-16)	1	394	97 (92-99)	11 (8-15)
CXR (abnormal)	2	176	69 (41-88)	40 (33-48)	2	176	92 (61-99)	9 (5-14)
Cough (any)	6	2,173	84 (70-92)	46 (39-54)	6	2,173	99 (96-100)	10 (7-13)
Cough (>=2 weeks)	4	1,860	42 (22-65)	81 (67-90)	4	1,860	99 (96-100)	8 (4-17)
Hb (<10 g/dL)	5	2,013	72 (64-79)	48 (40-57)	5	2,013	99 (97-100)	9 (7-13)
Hb (<8 g/dL)	5	2,013	49 (41-58)	73 (67-79)	5	2,013	99 (97-100)	10 (7-13)
BMI (<18.5 kg/m²)§	4	1,553	50 (42-57)	61 (49-71)	4	1,553	100 (98-100)	8 (5-13)
Lymphadenopathy	3	337	24 (14-38)	90 (86-93)	3	337	98 (86-100)	7 (4-10)
W4SS or CRP (>=10 mg/L)¶	1	394	100 (93-100)	5 (3-8)	1	394	97 (92-99)	11 (8-15)
W4SS or CXR (abnormal)¶	2	176	93 (54-99)	4 (2-9)	2	176	93 (54-99)	9 (5-14)
W4SS then CRP (>=5 mg/L)¶	1	394	94 (87-97)	20 (16-25)	1	394	97 (92-99)	11 (8-15)

†Direct comparisons are based on all studies that evaluated both the W4SS and relevant screening test

§Bivariate model did not converge; results from a univariate random-effects model. We computed binomial 95% CIs for W4SS sensitivity as all studies had 100% sensitivity

¶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

Table 8-21: Translation of pooled sensitivity and specificity estimates of different screening tests/strategies and diagnostic algorithms to a population of 1000 persons

Table 8-21A: Translation of pooled sensitivity and specificity estimates of different screening tests/strategies and diagnostic algorithms to a population of 1000 persons using culture as a reference standard§

		Outcome of screening§§							Outcome of screening then diagnosis§§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
	5% prevalence														
W4SS	50	932	49	883	67	1	5.3	98.5	-	-	-	-	-	-	-
CRP (>=10 mg/L)	50	799	48	750	200	2	6.1	99.3	-	-	-	-	-	-	-
CRP (>=8 mg/L)	50	828	48	779	171	2	5.9	99.1	-	-	-	-	-	-	-
CRP (>=5 mg/L)	50	885	49	836	114	1	5.5	99.1	-	-	-	-	-	-	-
CXR (abnormal)	50	570	38	532	418	12	6.6	97.1	-	-	-	-	-	-	-
Cough (any)	50	581	40	542	408	10	6.8	97.5	-	-	-	-	-	-	-
Cough (>=2 weeks)	50	204	14	190	760	36	7.1	95.5	-	-	-	-	-	-	-
Hb (<10 g/dL)	50	599	38	560	390	12	6.4	97.1	-	-	-	-	-	-	-
Hb (<8 g/dL)	50	341	28	313	637	22	8.1	96.6	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	50	390	29	361	589	21	7.3	96.5	-	-	-	-	-	-	-
Lymphadenopathy	50	130	6	124	826	44	4.6	94.9	-	-	-	-	-	-	-
W4SS or CRP (>=10 mg/L)¶	50	952	50	902	48	0	5.2	100	-	-	-	-	-	-	-
W4SS or CXR (abnormal)¶	50	928	45	883	67	5	4.8	93	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	50	808	48	760	190	2	5.9	98.7	-	-	-	-	-	-	-
WHO Xpert algorithm*†	50	-	-	-	-	-	-	-	38	884	66	12	36.4	98.7	26
Xpert alone*†	50	-	-	-	-	-	-	-	39	884	66	11	37	98.8	26
	10% prevalence														
W4SS	100	935	98	837	63	2	10.5	96.9	-	-	-	-	-	-	-
CRP (>=10 mg/L)	100	808	97	711	189	3	12	98.4	-	-	-	-	-	-	-
CRP (>=8 mg/L)	100	835	97	738	162	3	11.6	98.2	-	-	-	-	-	-	-
CRP (>=5 mg/L)	100	890	98	792	108	2	11	98.2	-	-	-	-	-	-	-
CXR (abnormal)	100	579	75	504	396	25	13	94.1	-	-	-	-	-	-	-
Cough (any)	100	592	79	513	387	21	13.3	94.9	-	-	-	-	-	-	-
Cough (>=2 weeks)	100	209	29	180	720	71	13.9	91	-	-	-	-	-	-	-
Hb (<10 g/dL)	100	608	77	531	369	23	12.7	94.1	-	-	-	-	-	-	-
Hb (<8 g/dL)	100	352	55	297	603	45	15.6	93.1	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	100	399	57	342	558	43	14.3	92.8	-	-	-	-	-	-	-
Lymphadenopathy	100	129	12	117	783	88	9.3	89.9	-	-	-	-	-	-	-

		Outcome of screening§§							Outcome of screening then diagnosis§§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
W4SS or CRP (>=10 mg/L)¶	100	955	100	855	45	0	10.5	100	-	-	-	-	-	-	-
W4SS or CXR (abnormal)¶	100	927	90	837	63	10	9.7	86.3	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	100	815	95	720	180	5	11.7	97.3	-	-	-	-	-	-	-
WHO Xpert algorithm*†	100	-	-	-	-	-	-	-	76	837	63	24	54.7	97.2	13
Xpert alone*†	100	-	-	-	-	-	-	-	78	837	63	22	55.3	97.4	13
	20% prevalence														
W4SS	200	940	196	744	56	4	20.9	93.3	-	-	-	-	-	-	-
CRP (>=10 mg/L)	200	826	194	632	168	6	23.5	96.6	-	-	-	-	-	-	-
CRP (>=8 mg/L)	200	850	194	656	144	6	22.8	96	-	-	-	-	-	-	-
CRP (>=5 mg/L)	200	900	196	704	96	4	21.8	96	-	-	-	-	-	-	-
CXR (abnormal)	200	598	150	448	352	50	25.1	87.6	-	-	-	-	-	-	-
Cough (any)	200	614	158	456	344	42	25.7	89.1	-	-	-	-	-	-	-
Cough (>=2 weeks)	200	218	58	160	640	142	26.6	81.8	-	-	-	-	-	-	-
Hb (<10 g/dL)	200	626	154	472	328	46	24.6	87.7	-	-	-	-	-	-	-
Hb (<8 g/dL)	200	374	110	264	536	90	29.4	85.6	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	200	418	114	304	496	86	27.3	85.2	-	-	-	-	-	-	-
Lymphadenopathy	200	128	24	104	696	176	18.8	79.8	-	-	-	-	-	-	-
W4SS or CRP (>=10 mg/L)¶	200	960	200	760	40	0	20.8	100	-	-	-	-	-	-	-
W4SS or CXR (abnormal)¶	200	924	180	744	56	20	19.5	73.7	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	200	830	190	640	160	10	22.9	94.1	-	-	-	-	-	-	-
WHO Xpert algorithm*†	200	-	-	-	-	-	-	-	152	744	56	48	73.1	93.9	7
Xpert alone*†	200	-	-	-	-	-	-	-	156	744	56	44	73.6	94.4	6
	30% prevalence														
W4SS	300	945	294	651	49	6	31.1	89.1	-	-	-	-	-	-	-
CRP (>=10 mg/L)	300	844	291	553	147	9	34.5	94.2	-	-	-	-	-	-	-
CRP (>=8 mg/L)	300	865	291	574	126	9	33.6	93.3	-	-	-	-	-	-	-
CRP (>=5 mg/L)	300	910	294	616	84	6	32.3	93.3	-	-	-	-	-	-	-
CXR (abnormal)	300	617	225	392	308	75	36.5	80.4	-	-	-	-	-	-	-
Cough (any)	300	636	237	399	301	63	37.3	82.7	-	-	-	-	-	-	-
Cough (>=2 weeks)	300	227	87	140	560	213	38.3	72.4	-	-	-	-	-	-	-
Hb (<10 g/dL)	300	644	231	413	287	69	35.9	80.6	-	-	-	-	-	-	-
Hb (<8 g/dL)	300	396	165	231	469	135	41.7	77.6	-	-	-	-	-	-	-
BMI (<18.5 kg/m²)	300	437	171	266	434	129	39.1	77.1	-	-	-	-	-	-	-
Lymphadenopathy	300	127	36	91	609	264	28.3	69.8	-	-	-	-	-	-	-

		Outcome of screening§§							Outcome of screening then diagnosis§§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**	**TP**	**TN**	**FP**	**FN**	**PPV**	**NPV**	**NNS**
W4SS or CRP (>=10 mg/L)¶	300	965	300	665	35	0	31.1	100	-	-	-	-	-	-	-
W4SS or CXR (abnormal)¶	300	921	270	651	49	30	29.3	62	-	-	-	-	-	-	-
W4SS then CRP (>=5 mg/L)¶	300	845	285	560	140	15	33.7	90.3	-	-	-	-	-	-	-
WHO Xpert algorithm*†	300	-	-	-	-	-	-	-	228	651	49	72	82.3	90	4
Xpert alone*†	300	-	-	-	-	-	-	-	234	651	49	66	82.7	90.8	4

§According to WHO screening & diagnostic algorithm, Xpert testing 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)

§§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e., the number who need subsequent diagnostic testing)

¶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

*Accuracy measures for entire algorithm using sputum and/or urine Xpert result

†The test accuracy of Xpert in those who were W4SS positive was: 4 studies; 586 participants; sensitivity 0.78 (0.68-0.86), specificity 0.93 (0.87-0.96).

Definition of abbreviations: BMI = body mass index, CRP = C-reactive protein, CXR = chest X-ray, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-21B: Translation of pooled sensitivity and specificity estimates of different screening tests/strategies to a population of 1000 persons using Xpert as a reference standard§

		Outcome of screening§§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**
	5% prevalence							
W4SS	50	904	49	855	95	1	5.4	99
CRP (>=10 mg/L)	50	807	47	760	190	3	5.8	98.4
CRP (>=8 mg/L)	50	836	47	788	162	3	5.6	98.2
CRP (>=5 mg/L)	50	884	48	836	114	2	5.4	98.3
CXR (abnormal)	50	604	34	570	380	16	5.7	96.1
Cough (any)	50	555	42	513	437	8	7.6	98.2
Cough (>=2 weeks)	50	201	21	180	770	29	10.4	96.4
Hb (<10 g/dL)	50	530	36	494	456	14	6.7	96.9
Hb (<8 g/dL)	50	271	24	247	703	26	8.9	96.4
BMI (<18.5 kg/m²)	50	396	25	370	580	25	6.3	95.9
Lymphadenopathy	50	107	12	95	855	38	11.2	95.7
W4SS or CRP (>=10 mg/L)¶	50	952	50	902	48	0	5.2	100
W4SS or CXR (abnormal)¶	50	958	46	912	38	4	4.9	91.6
W4SS then CRP (>=5 mg/L)¶	50	807	47	760	190	3	5.8	98.4
	10% prevalence							
W4SS	100	908	98	810	90	2	10.8	97.8
CRP (>=10 mg/L)	100	814	94	720	180	6	11.5	96.8
CRP (>=8 mg/L)	100	841	94	747	153	6	11.2	96.2
CRP (>=5 mg/L)	100	888	96	792	108	4	10.8	96.4
CXR (abnormal)	100	609	69	540	360	31	11.3	92.1
Cough (any)	100	570	84	486	414	16	14.7	96.3
Cough (>=2 weeks)	100	213	42	171	729	58	19.7	92.6
Hb (<10 g/dL)	100	539	71	468	432	29	13.2	93.7
Hb (<8 g/dL)	100	282	48	234	666	52	17	92.8
BMI (<18.5 kg/m²)	100	401	50	351	549	50	12.5	91.7
Lymphadenopathy	100	114	24	90	810	76	21.1	91.4
W4SS or CRP (>=10 mg/L)¶	100	955	100	855	45	0	10.5	100
W4SS or CXR (abnormal)¶	100	957	93	864	36	7	9.7	83.7
W4SS then CRP (>=5 mg/L)¶	100	814	94	720	180	6	11.5	96.8
	20% prevalence							
W4SS	200	916	196	720	80	4	21.4	95.2
CRP (>=10 mg/L)	200	828	188	640	160	12	22.7	93
CRP (>=8 mg/L)	200	852	188	664	136	12	22.1	91.9

		Outcome of screening§§						
Test	**Total TB**	**TP+FP‡**	**TP**	**FP**	**TN**	**FN**	**PPV**	**NPV**
CRP (>=5 mg/L)	200	896	192	704	96	8	21.4	92.3
CXR (abnormal)	200	618	138	480	320	62	22.3	83.8
Cough (any)	200	600	168	432	368	32	28	92
Cough (>=2 weeks)	200	236	84	152	648	116	35.6	84.8
Hb (<10 g/dL)	200	558	142	416	384	58	25.4	86.9
Hb (<8 g/dL)	200	304	96	208	592	104	31.6	85.1
BMI (<18.5 kg/m²)	200	412	100	312	488	100	24.3	83
Lymphadenopathy	200	128	48	80	720	152	37.5	82.6
W4SS or CRP (>=10 mg/L)¶	200	960	200	760	40	0	20.8	100
W4SS or CXR (abnormal)¶	200	954	186	768	32	14	19.5	69.6
W4SS then CRP (>=5 mg/L)¶	200	828	188	640	160	12	22.7	93
	30% prevalence							
W4SS	300	924	294	630	70	6	31.8	92.1
CRP (>=10 mg/L)	300	842	282	560	140	18	33.5	88.6
CRP (>=8 mg/L)	300	863	282	581	119	18	32.7	86.9
CRP (>=5 mg/L)	300	904	288	616	84	12	31.9	87.5
CXR (abnormal)	300	627	207	420	280	93	33	75.1
Cough (any)	300	630	252	378	322	48	40	87
Cough (>=2 weeks)	300	259	126	133	567	174	48.6	76.5
Hb (<10 g/dL)	300	577	213	364	336	87	36.9	79.4
Hb (<8 g/dL)	300	326	144	182	518	156	44.2	76.9
BMI (<18.5 kg/m²)	300	423	150	273	427	150	35.5	74
Lymphadenopathy	300	142	72	70	630	228	50.7	73.4
W4SS or CRP (>=10 mg/L)¶	300	965	300	665	35	0	31.1	100
W4SS or CXR (abnormal)¶	300	951	279	672	28	21	29.3	57.1
W4SS then CRP (>=5 mg/L)¶	300	842	282	560	140	18	33.5	88.6

§§Estimated using the pooled point estimates for sensitivity and specificity for different tests/strategies

‡TP+FP is the number of participants who screen positive (i.e., the number who need subsequent diagnostic testing)

¶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, FN= false negative, FP = false positive, Hb = haemoglobin, NNS = number needed to screen, NPV = negative predictive value, PPV = positive predictive value, TB = tuberculosis, TN = true negative, TP = true positive, W4SS = WHO four-symptom screen

Table 8-22: Sensitivity analyses of diagnostic accuracy (pooled sensitivity and specificity) for each screening test/strategy for the detection of tuberculosis using an alternative reference standard of culture or Xpert*

	Sensitivity analyses 1*				Sensitivity analyses 2**			
	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)	No of studies	N	Sensitivity (95% CI)	Specificity (95% CI)
W4SS	6	2,211	99 (93-100)	9 (6-13)	4	697	97 (91-99)	8 (4-17)
CRP (>=10 mg/L)	1	400	94 (88-97)	22 (17-27)	1	400	94 (88-97)	22 (17-27)
CRP (>=8 mg/L)	1	400	94 (88-97)	19 (14-24)	1	400	94 (88-97)	19 (14-24)
CRP (>=5 mg/L)	1	400	96 (91-98)	13 (9-17)	1	400	96 (91-98)	13 (9-17)
CXR (abnormal)	2	176	69 (43-86)	41 (33-48)	1	52	57 (23-86)	42 (29-57)
Cough (any)	6	2,208	83 (69-92)	46 (37-55)	4	694	75 (59-86)	45 (33-58)
Cough (>=2 weeks)	4	1,895	41 (20-65)	81 (62-92)	3	631	30 (14-53)	81 (53-94)
Hb (<10 g/dL)	5	2,037	73 (66-80)	47 (37-58)	3	527	73 (66-80)	41 (29-56)
Hb (<8 g/dL)	5	2,037	53 (42-64)	72 (63-80)	3	527	56 (34-77)	68 (54-80)
BMI (<18.5 kg/m²)	4	1,571	50 (41-59)	61 (49-71)	2	112	56 (33-76)	63 (53-72)
Lymphadenopathy	3	356	22 (13-35)	90 (86-93)	2	123	14 (4-35)	87 (79-92)
W4SS or CRP (>=10 mg/L)¶	1	399	100 (94-100)	5 (3-8)	1	399	100 (94-100)	5 (3-8)
W4SS or CXR (abnormal)¶	2	176	94 (66-99)	4 (2-9)	1	52	86 (42-98)	4 (1-16)
W4SS then CRP (>=5 mg/L)¶	1	399	93 (87-96)	21 (17-26)	1	399	93 (87-96)	21 (17-26)

*Reference standard of culture or Xpert of sputum and/or other specimens

*Reference standard of culture or Xpert of sputum and/or other specimens among datasets that collected sputum for culture

¶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

Table 8-23: Diagnostic yield of different Xpert tests and sample types as a proportion of total microbiologically confirmed tuberculosis cases†

Study	Bjerrum‡	Gupta-Wright intervention*	Gupta-Wright control*	Heidebrecht	Huerga	Lawn**	Thit
Total sample size	69	1287	1287	156	387	420	54
Microbiological sample available	69	1279	779	156	235	420	54
Microbiologically confirmed¶	15	122	82	41	33	139	7
Sputum culture + (%)	13 (87%)	-	-	31 (76%)	-	58 (42%)	4 (57%)
N	69	-	-	131	-	209	54
Non-sputum culture + (%)	-	-	-	-	-	70 (50%)	-
N	-	-	-	-	-	420	-
Total culture + (%)	13 (87%)	-	-	31 (76%)	-	109 (78%)	4 (57%)
N	69	-	-	131	-	420	54
Sputum Xpert + (%)	9 (60%)	85 (70%)	82 (100%)	35 (85%)	33 (100%)	57 (41%)	4 (57%)
N	50	832	779	146	235	195	54
Urine Xpert + (%)	-	74 (61%)	-	-	-	89 (64%)	-
N	-	1270	-	-	-	411	-
Total Xpert + (%)	9 (60%)	122 (100%)	82 (100%)	35 (85%)	33 (100%)	116 (83%)	4 (57%)
N	50	1279	779	146	235	414	54

†Denominator for % is microbiologically confirmed

‡Study by Bjerrum et al has fewer samples collected for Xpert testing because Xpert only became available after study enrollment began

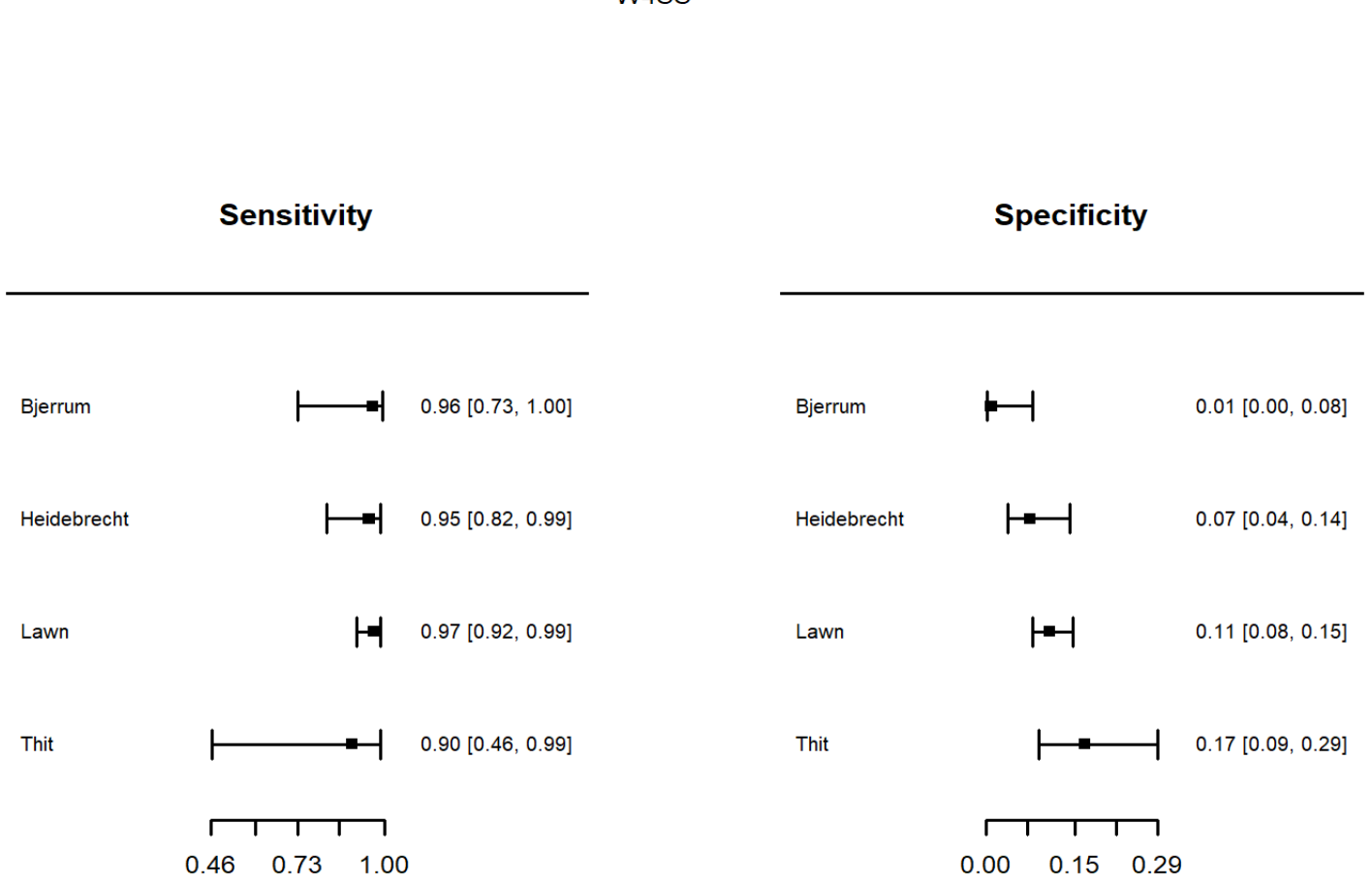
*Study by Gupta-Wright et al (2018) involved an intervention arm (systematically collected concentrated urine Xpert and sputum Xpert) and control arm (systematically collected sputum Xpert only)

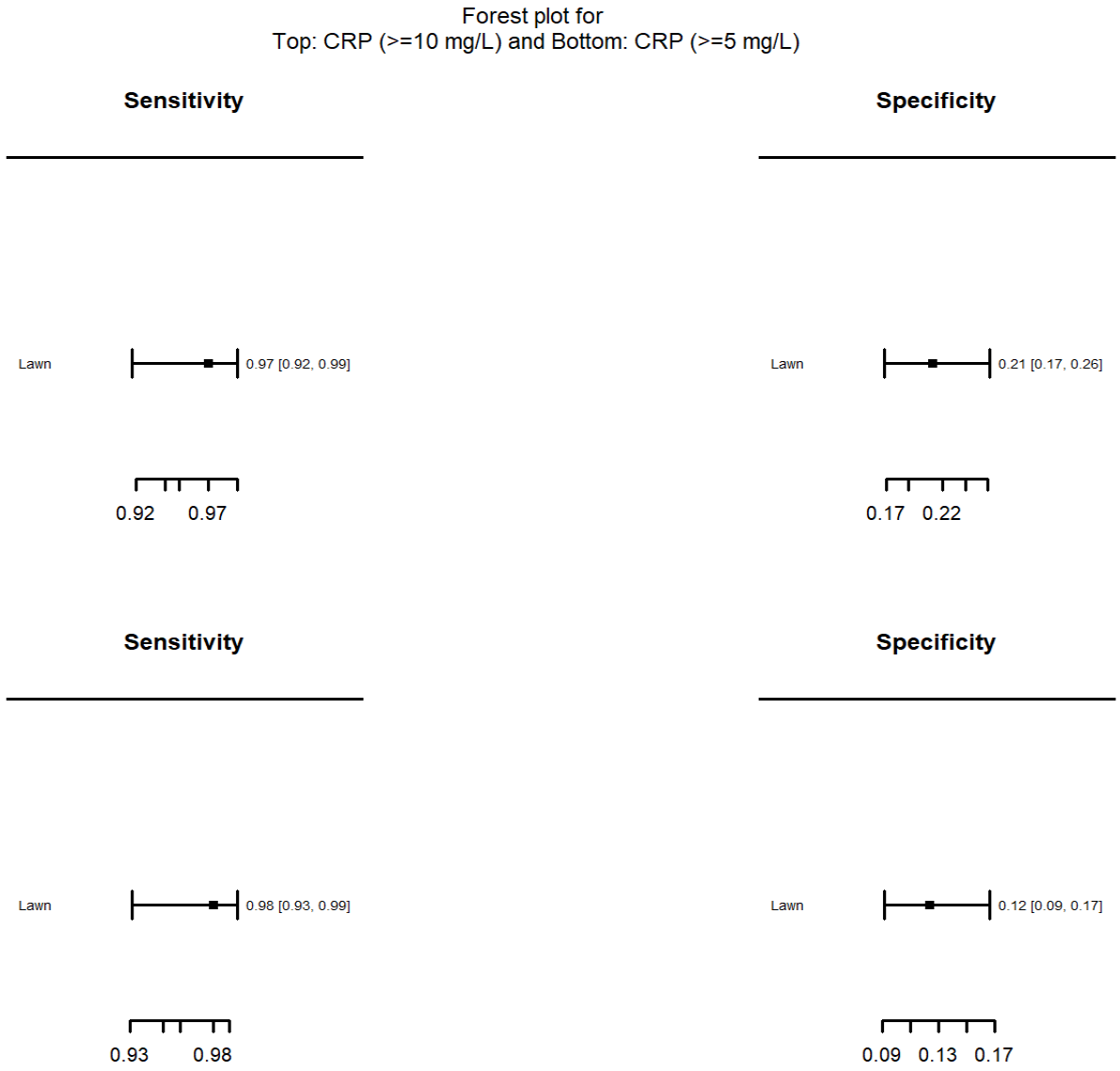
**The number (%) of all microbiologically cases diagnosed with concentrated urine Xpert was 82 (59%; 402 participants) and with unconcentrated urine Xpert was 59 (42%; 405 participants).

¶Defined as any Xpert, or culture positive.

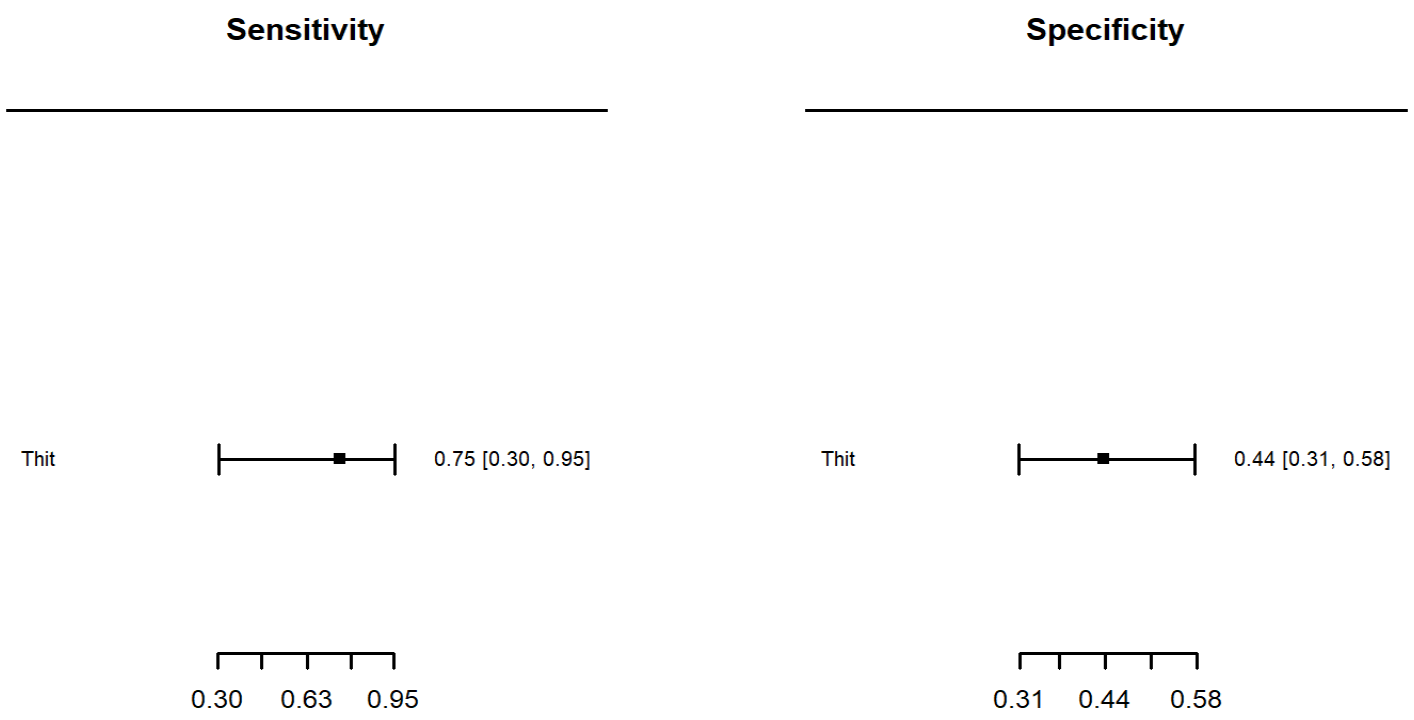
Figure 8-6: Forest plots of sensitivity and specificity estimates for each screening test/strategy (C-reactive protein >=8 mg/L omitted)

Figure 8-6A: Forest plots of sensitivity and specificity estimates for each screening test/strategy using culture as a reference standard

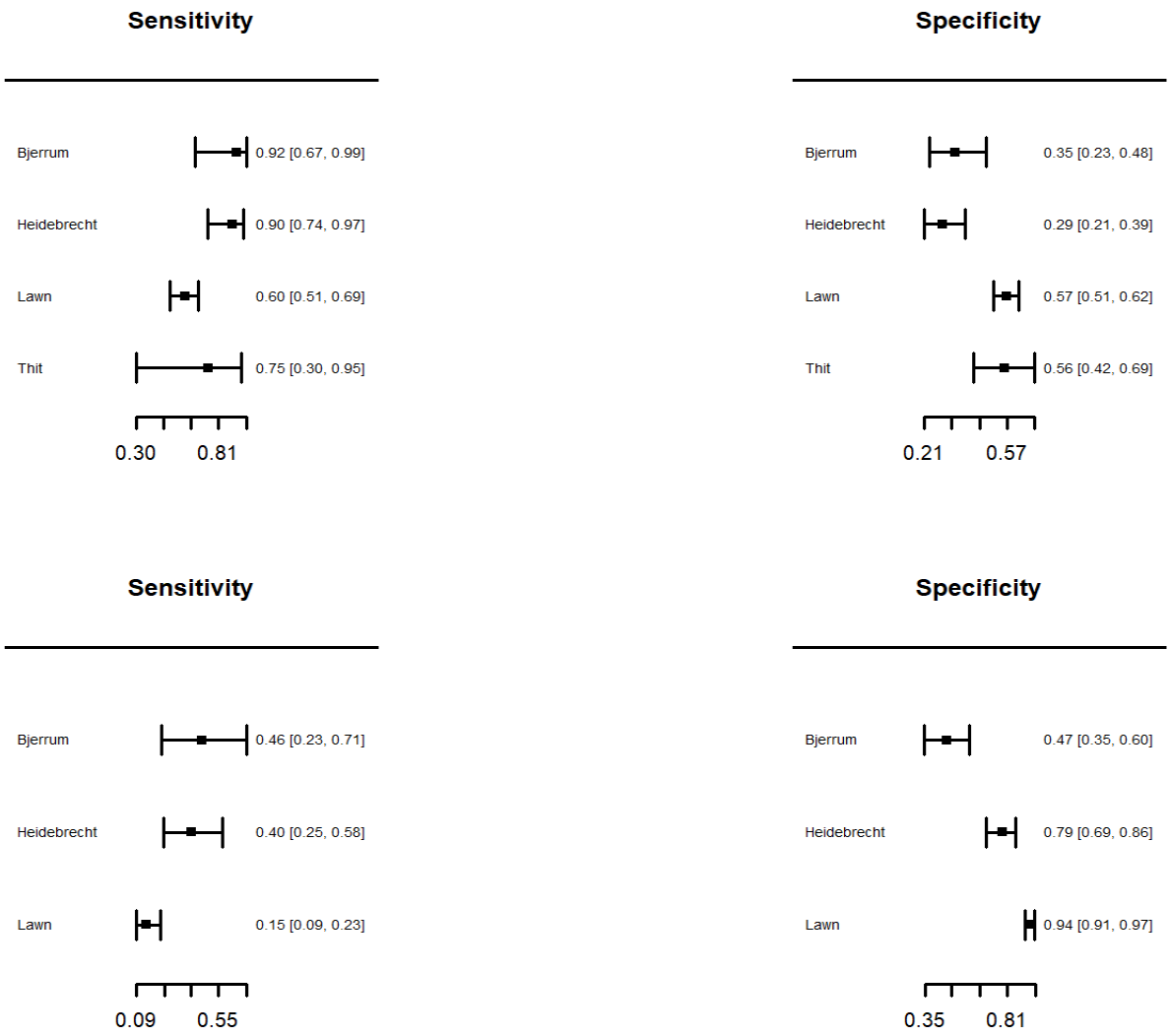



Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
0.97 [0.92, 0.99]
0.92
0.97
Lawn
0.21 [0.17, 0.26]
0.17
0.22
Sensitivity
Specificity
Lawn
0.98 [0.93, 0.99]
0.93
0.98
Lawn
0.12 [0.09, 0.17]
0.09
0.13
0.17

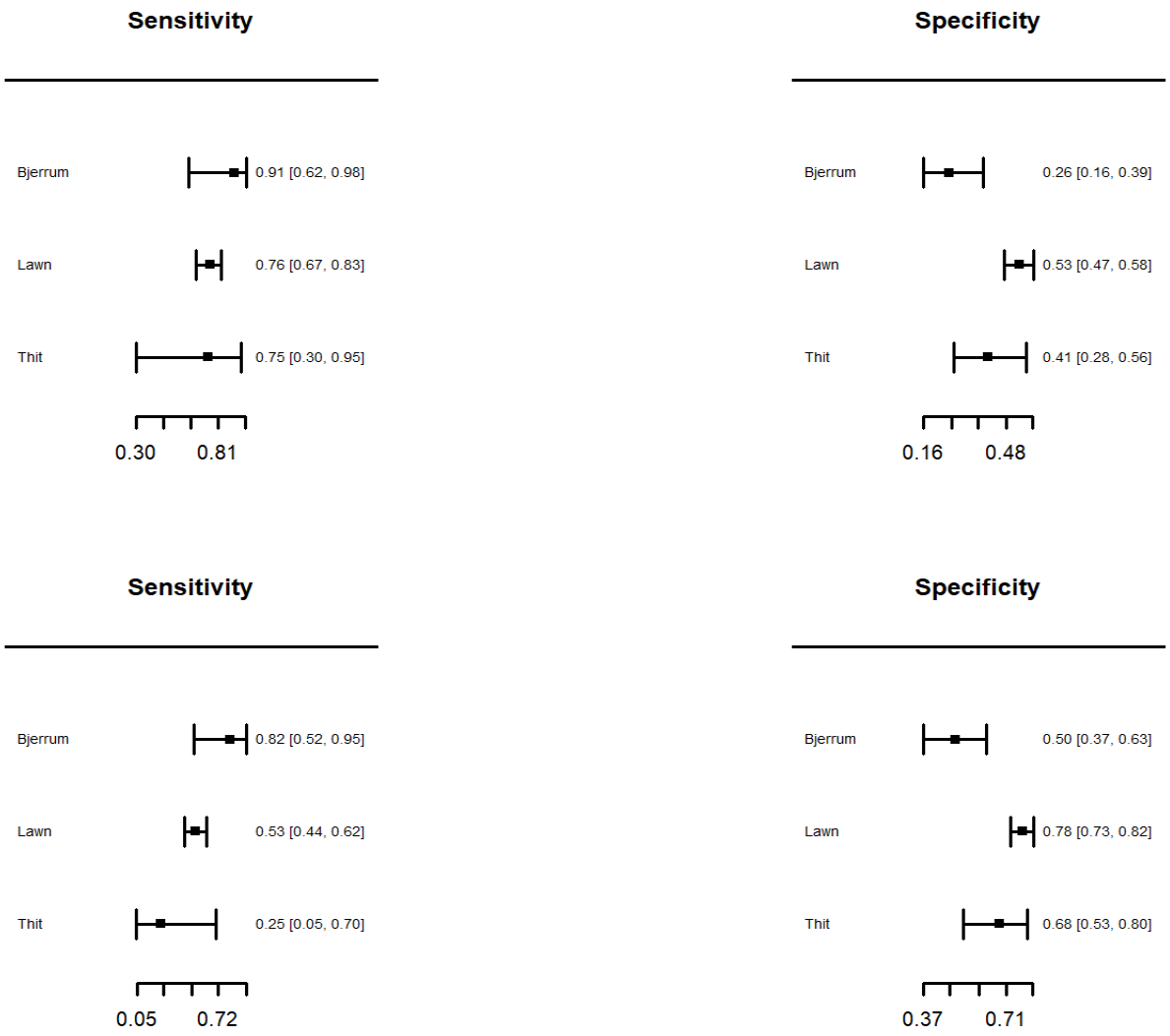
Forest plot for
CXR (abnormal)

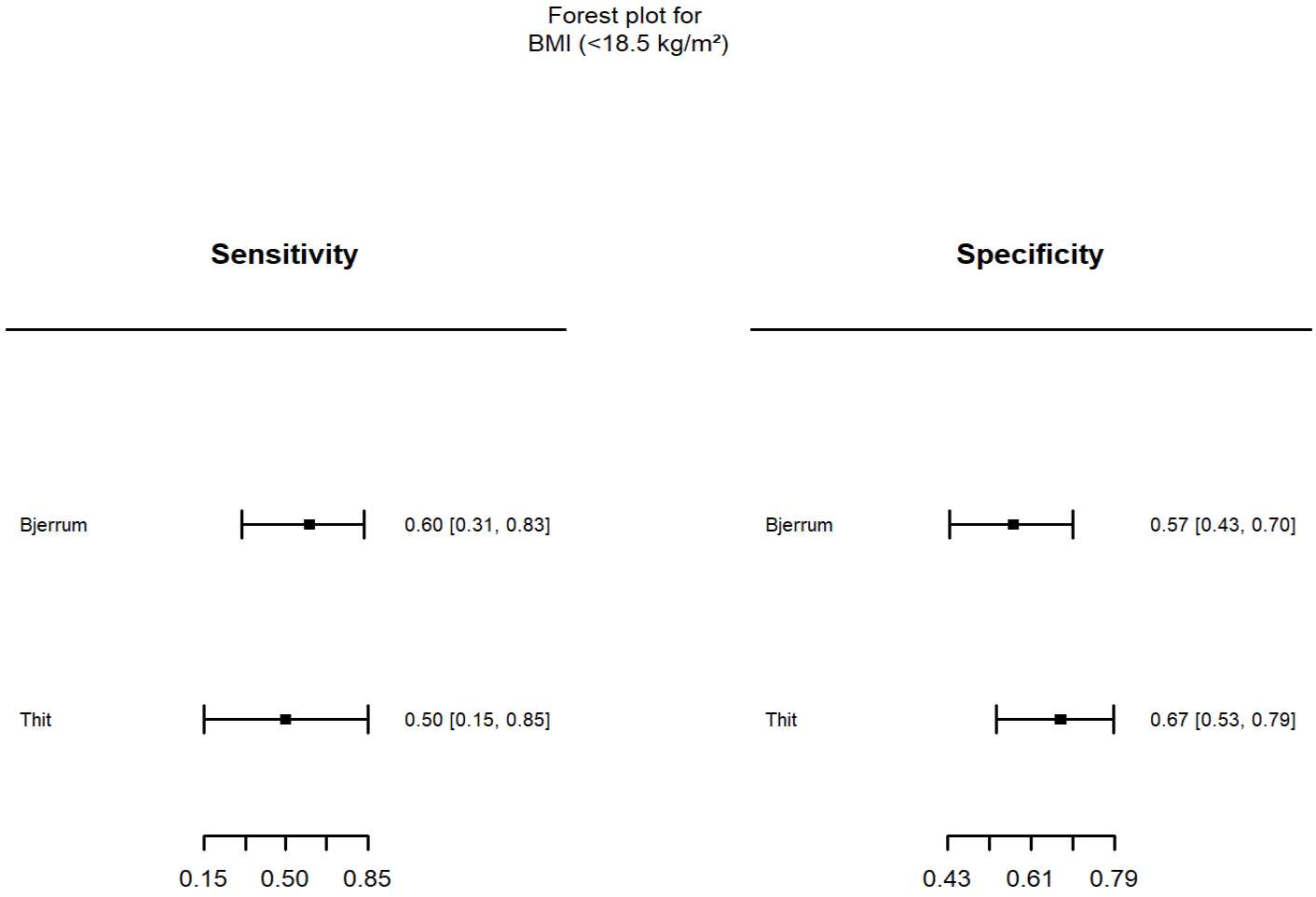


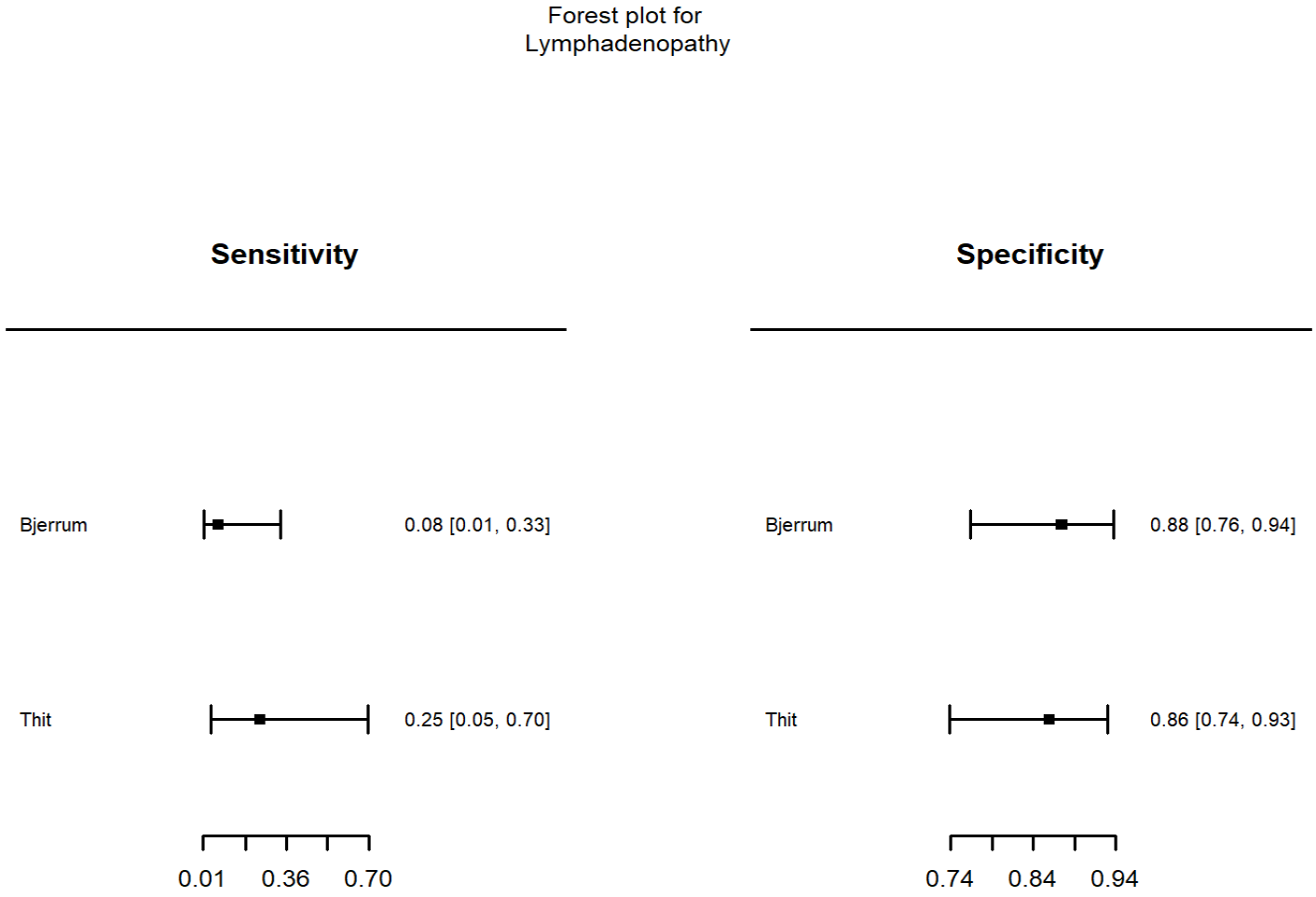
Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)



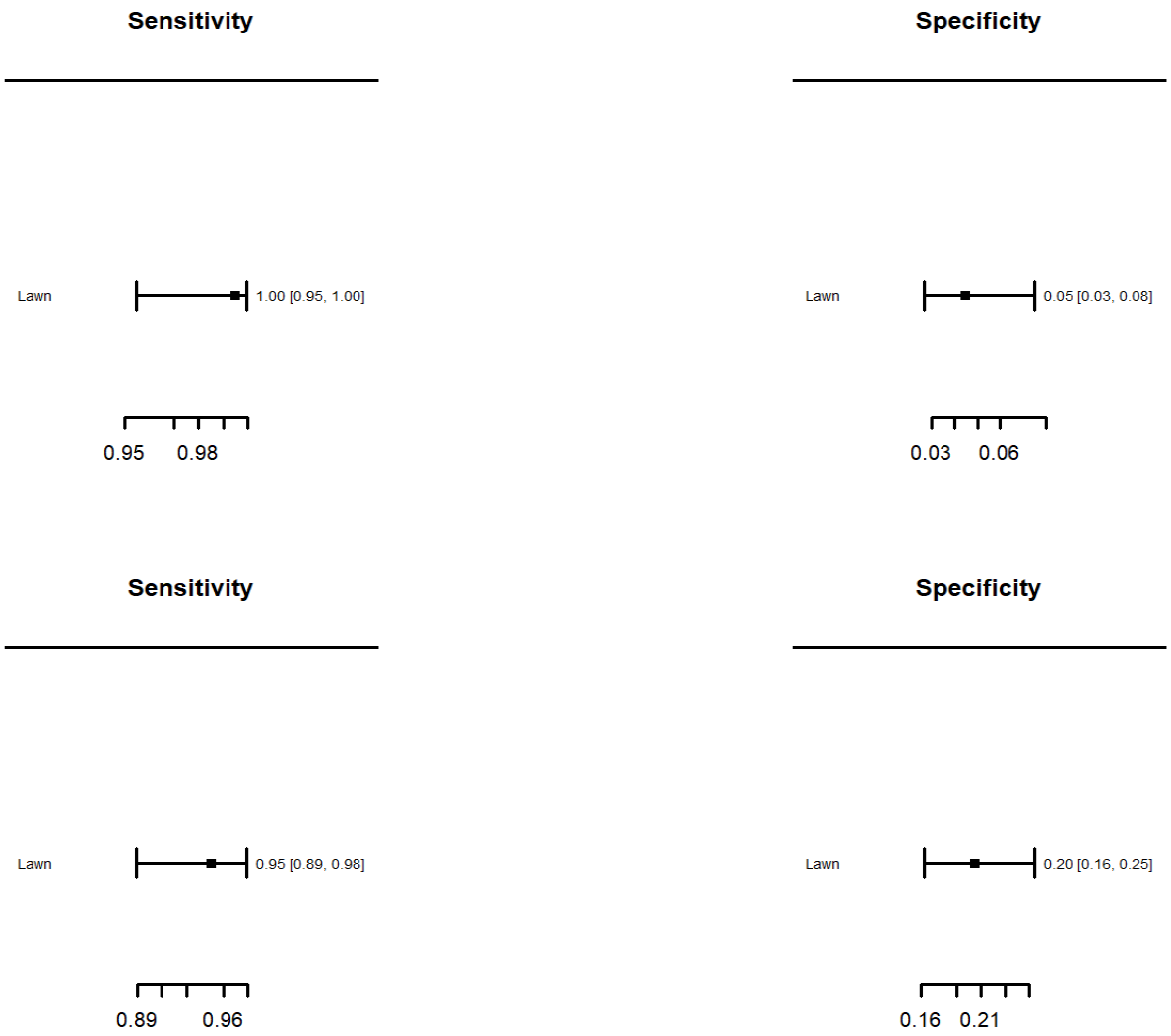
Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)



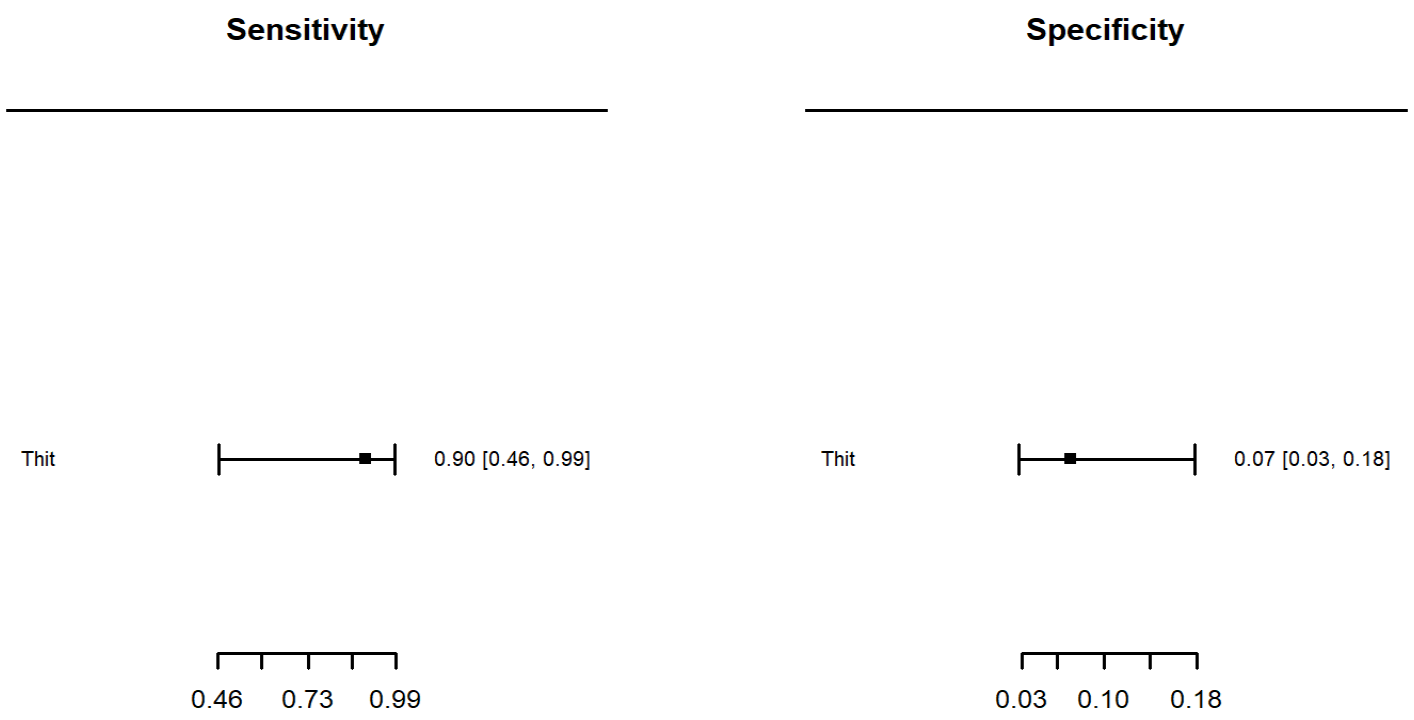

Forest plot for
BMI (<18.5 kg/m²)
Sensitivity
Specificity
Bjerrum
0.60 [0.31, 0.83]
Thit
0.50 [0.15, 0.85]
0.15
0.50
0.85
Bjerrum
0.57 [0.43, 0.70]
Thit
0.67 [0.53, 0.79]
0.43
0.61
0.79


Forest plot for
Lymphadenopathy
Sensitivity
Specificity
Bjerrum
0.08 [0.01, 0.33]
Thit
0.25 [0.05, 0.70]
0.01
0.36
0.70
Bjerrum
0.88 [0.76, 0.94]
Thit
0.86 [0.74, 0.93]
0.74
0.84
0.94

Forest plot for
Top: W4SS or CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)



Forest plot for
W4SS or CXR (abnormal)



Forest plot for
Top: WHO Xpert algorithm and Bottom: Xpert alone

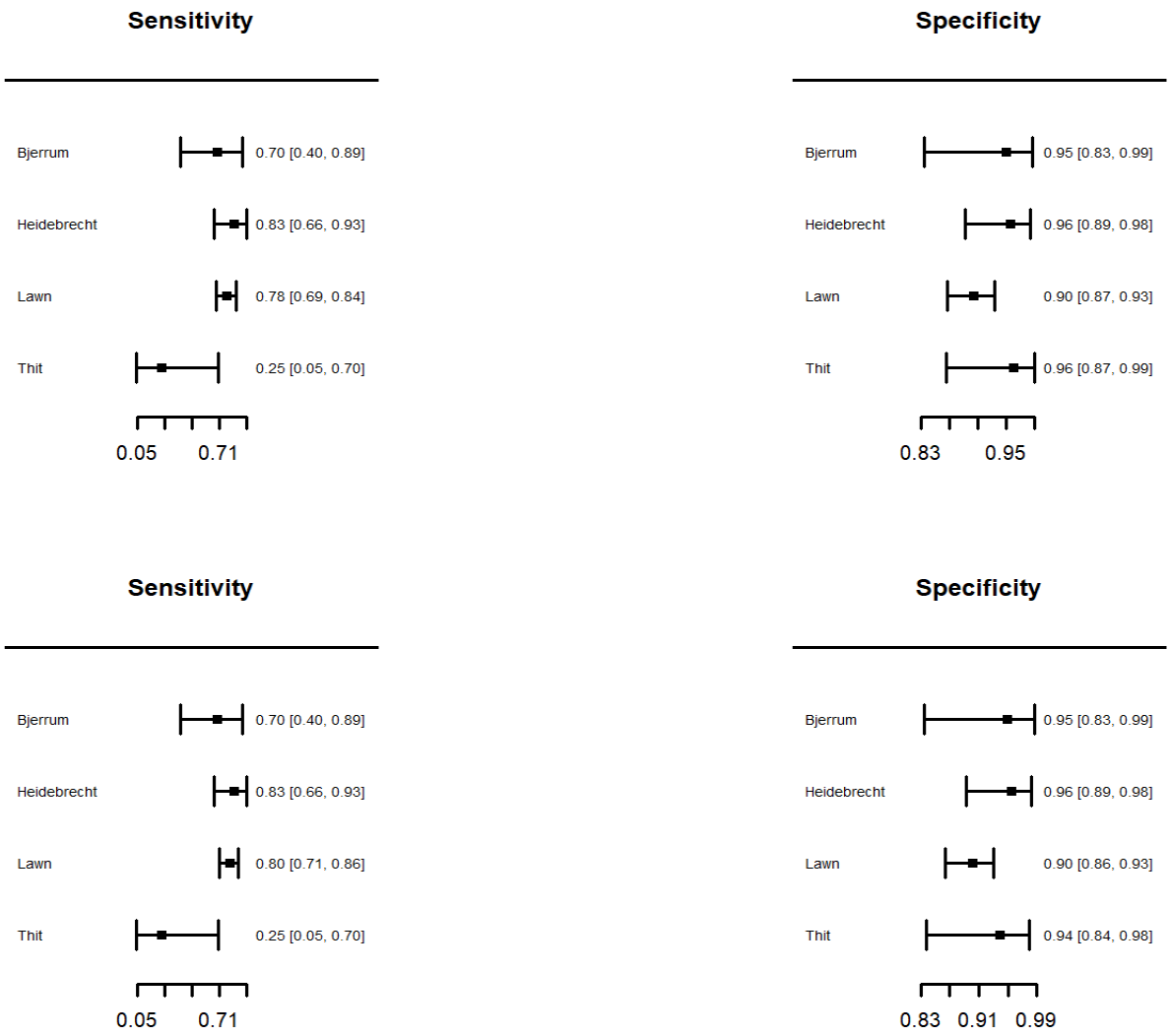
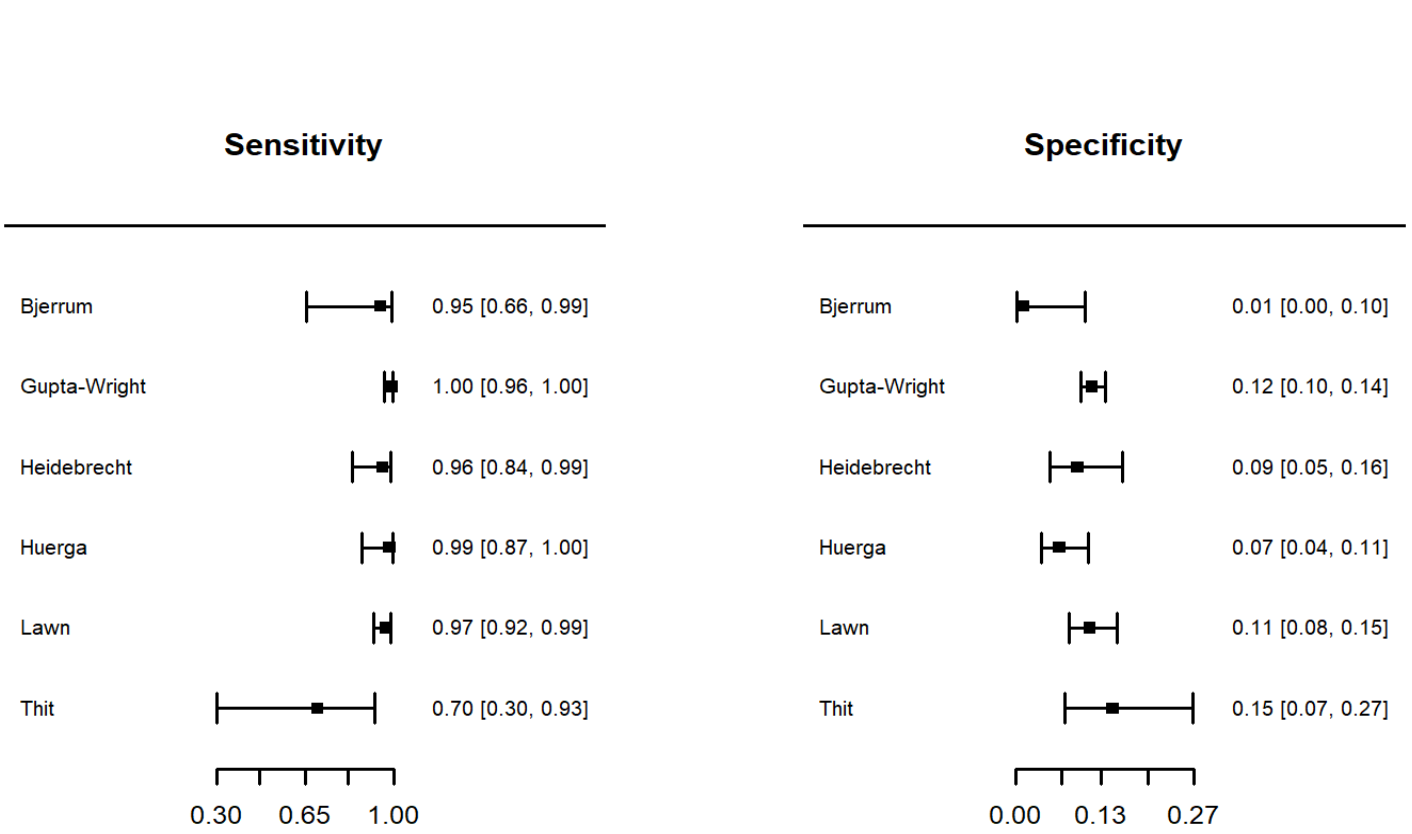
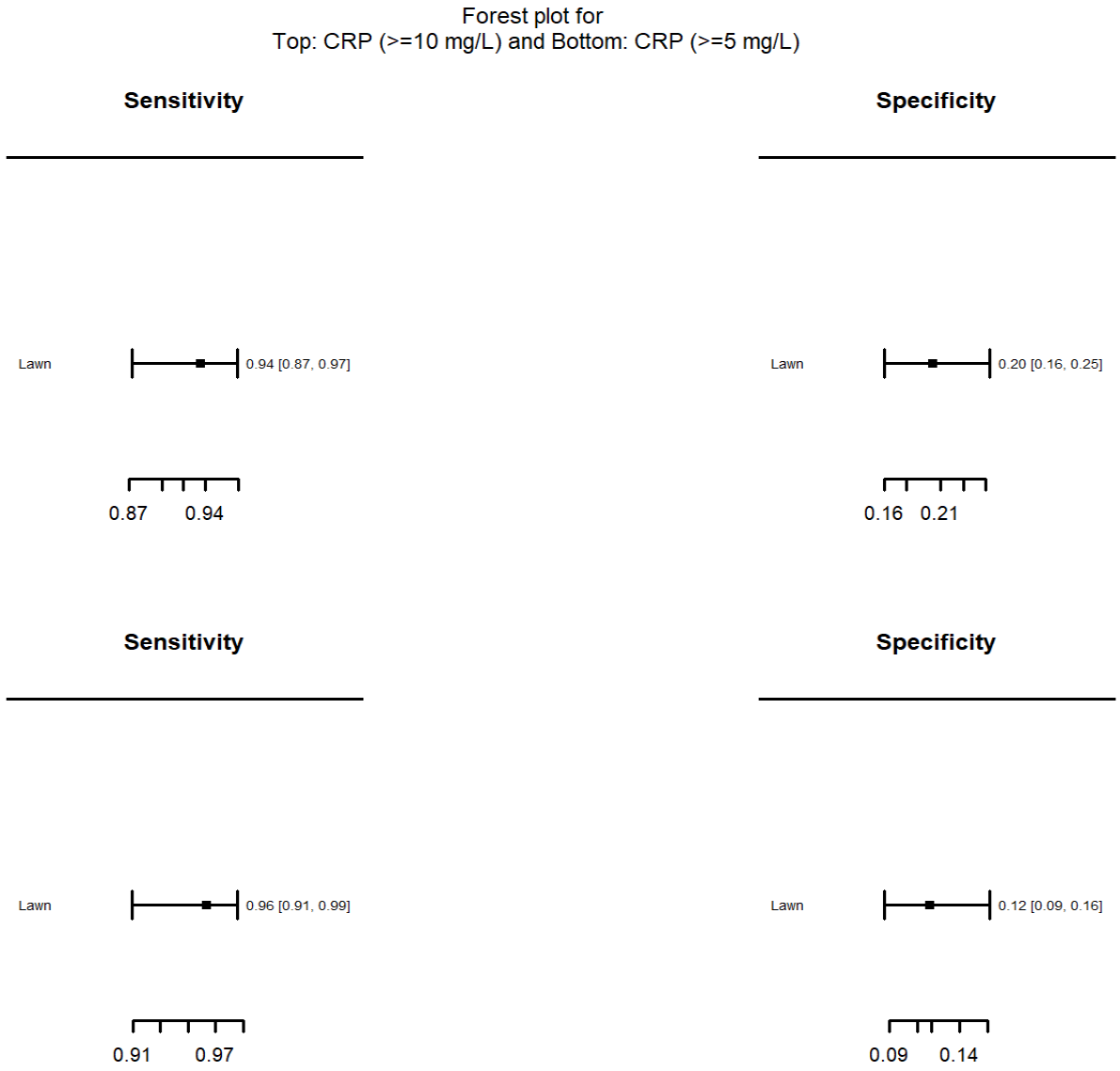
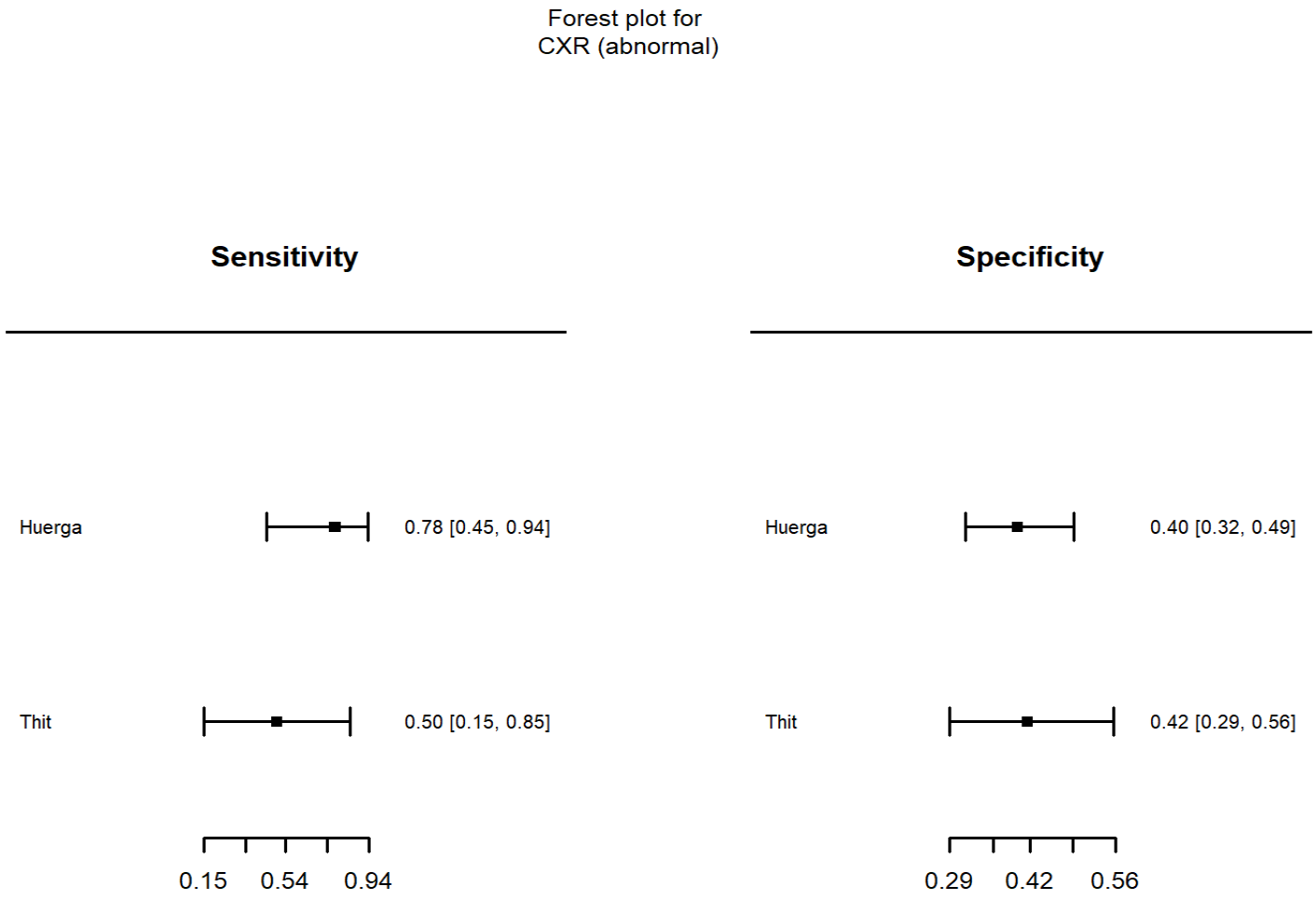


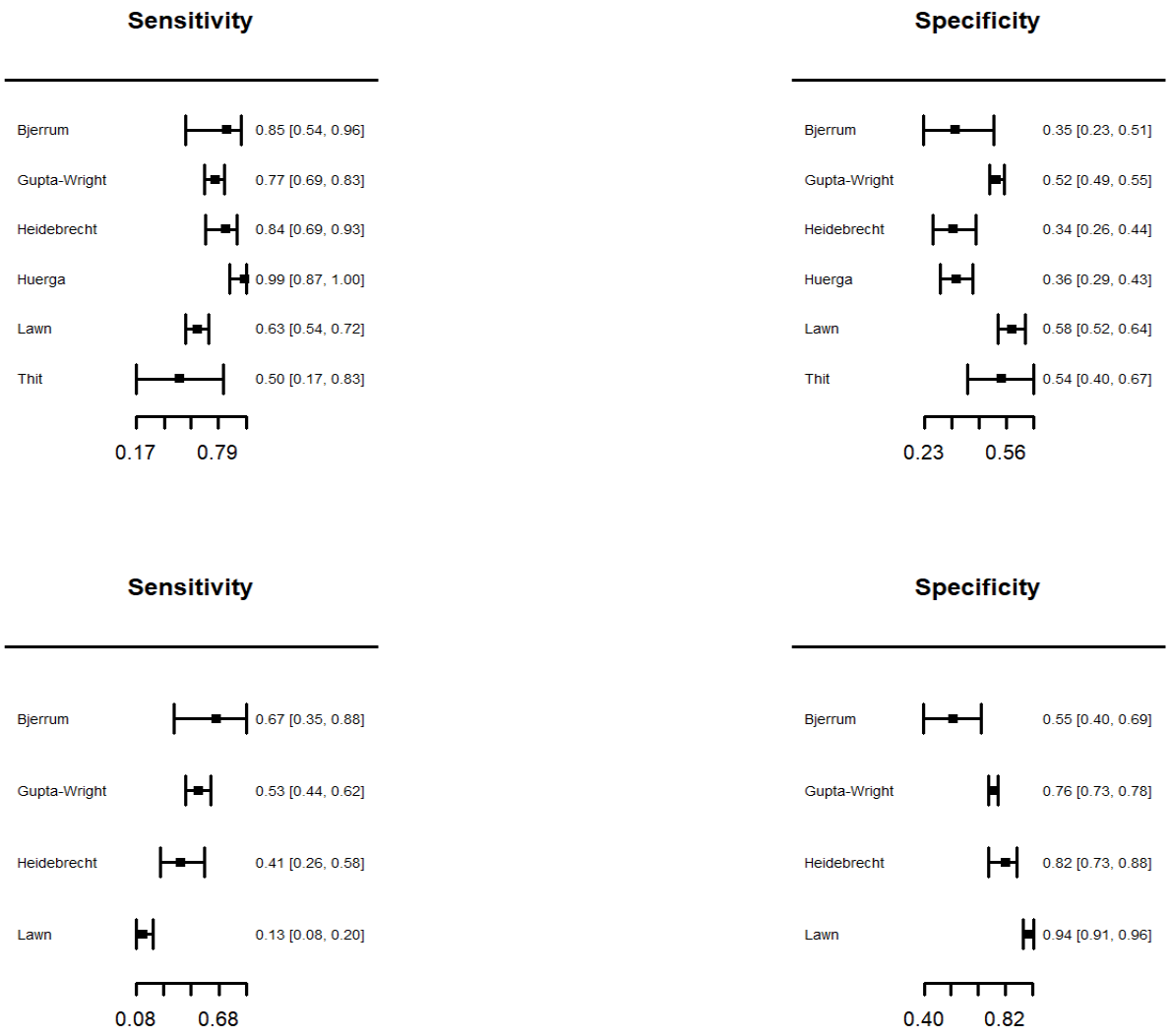
Figure S8-6B - Forest plots of sensitivity and specificity estimates for each screening test/strategy using Xpert as a reference standard



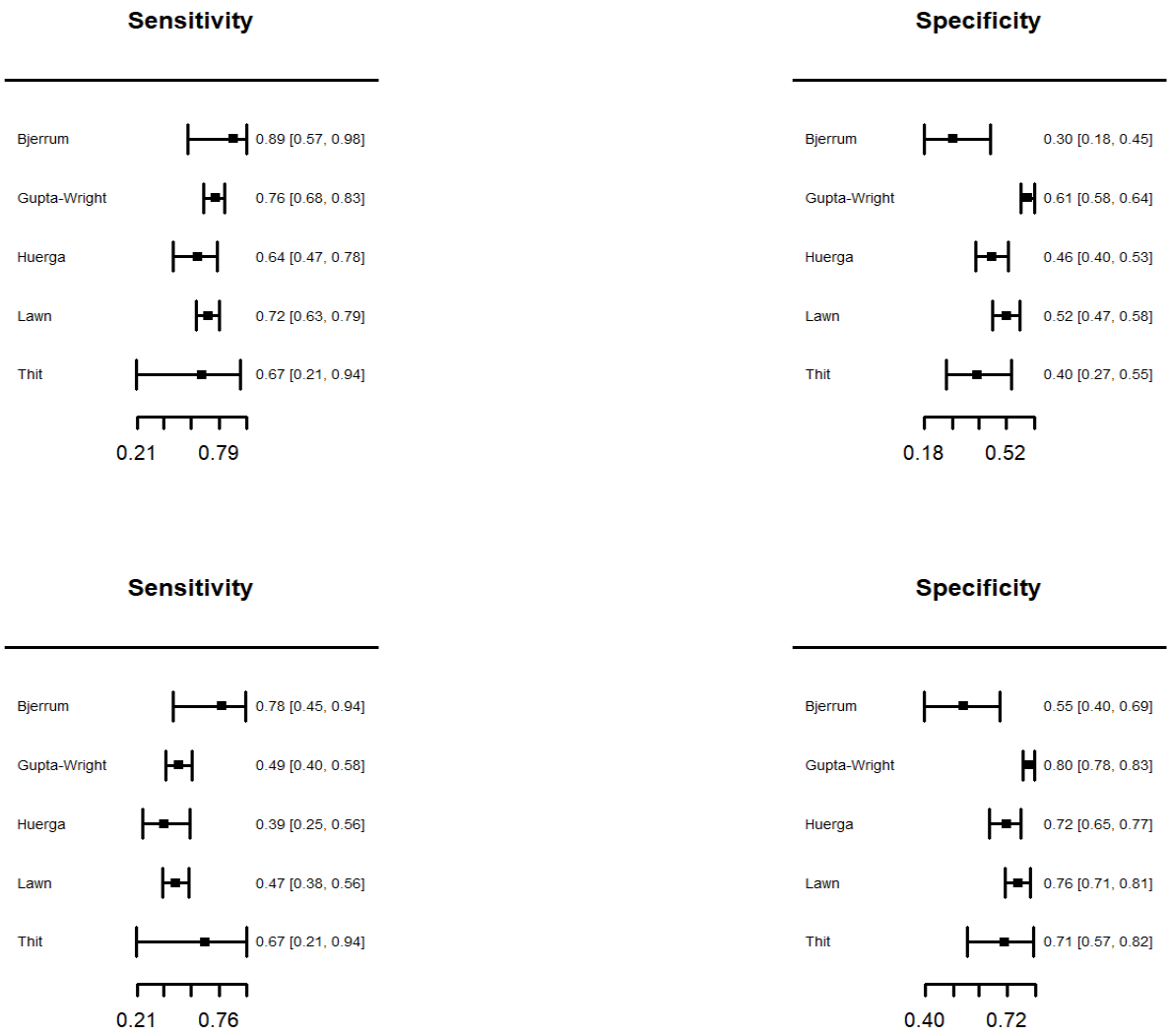

Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
0.94 [0.87, 0.97]
0.20 [0.16, 0.25]
0.87
0.94
0.16
0.21
Sensitivity
Specificity
Lawn
0.96 [0.91, 0.99]
0.12 [0.09, 0.16]
0.91
0.97
0.09
0.14


Forest plot for
CXR (abnormal)
Sensitivity
Specificity
Huerga
0.78 [0.45, 0.94]
Thit
0.50 [0.15, 0.85]
0.15
0.54
0.94
Huerga
0.40 [0.32, 0.49]
Thit
0.42 [0.29, 0.56]
0.29
0.42
0.56

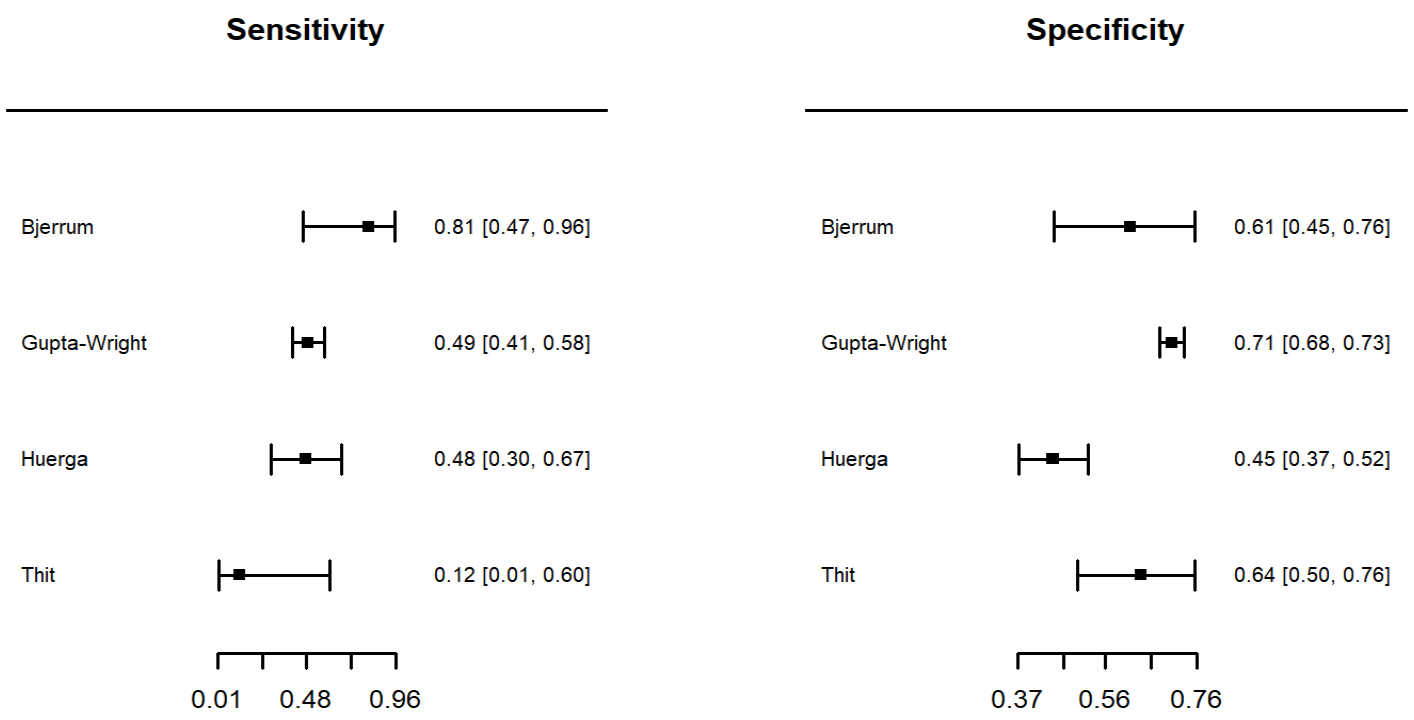
Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)

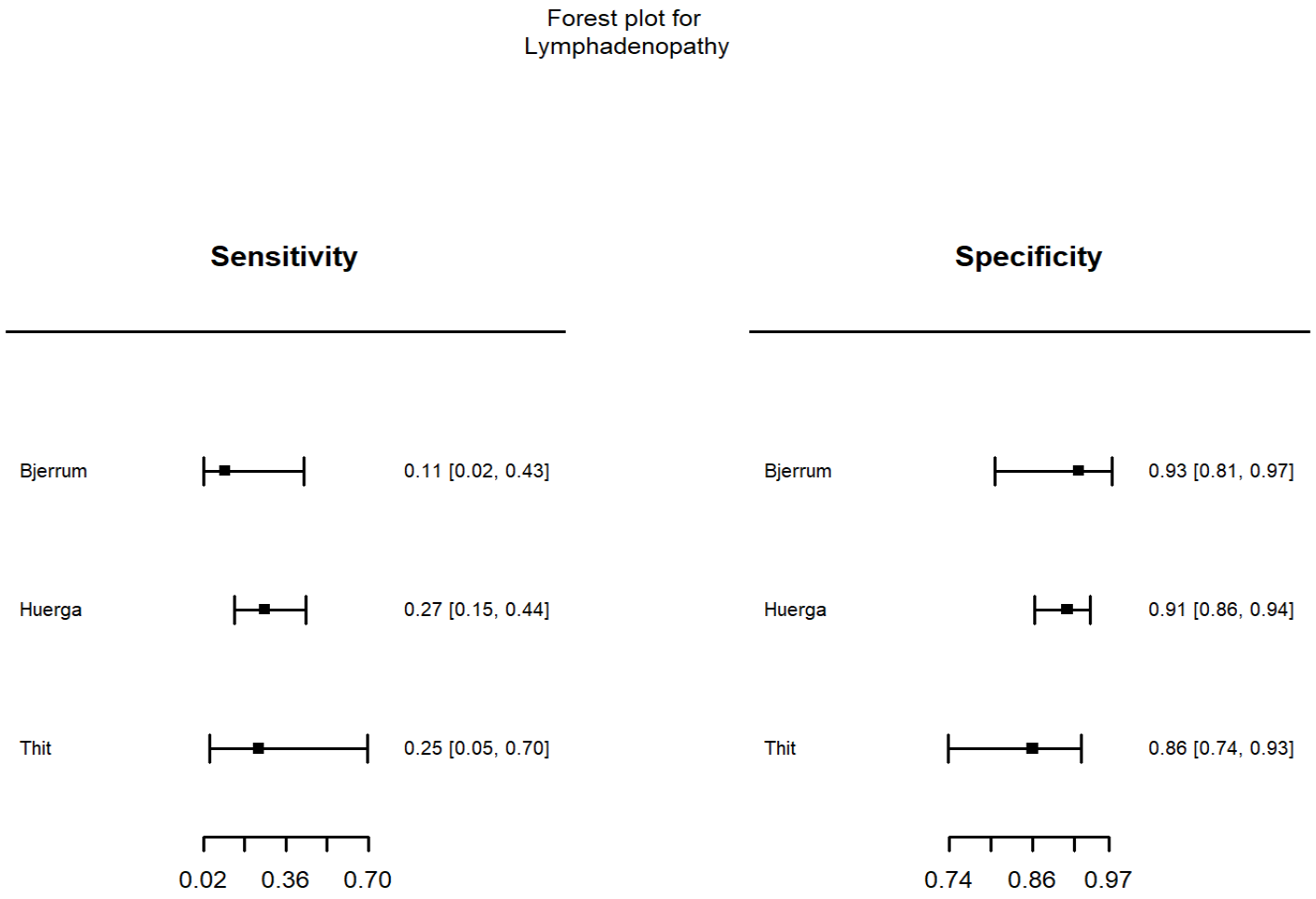


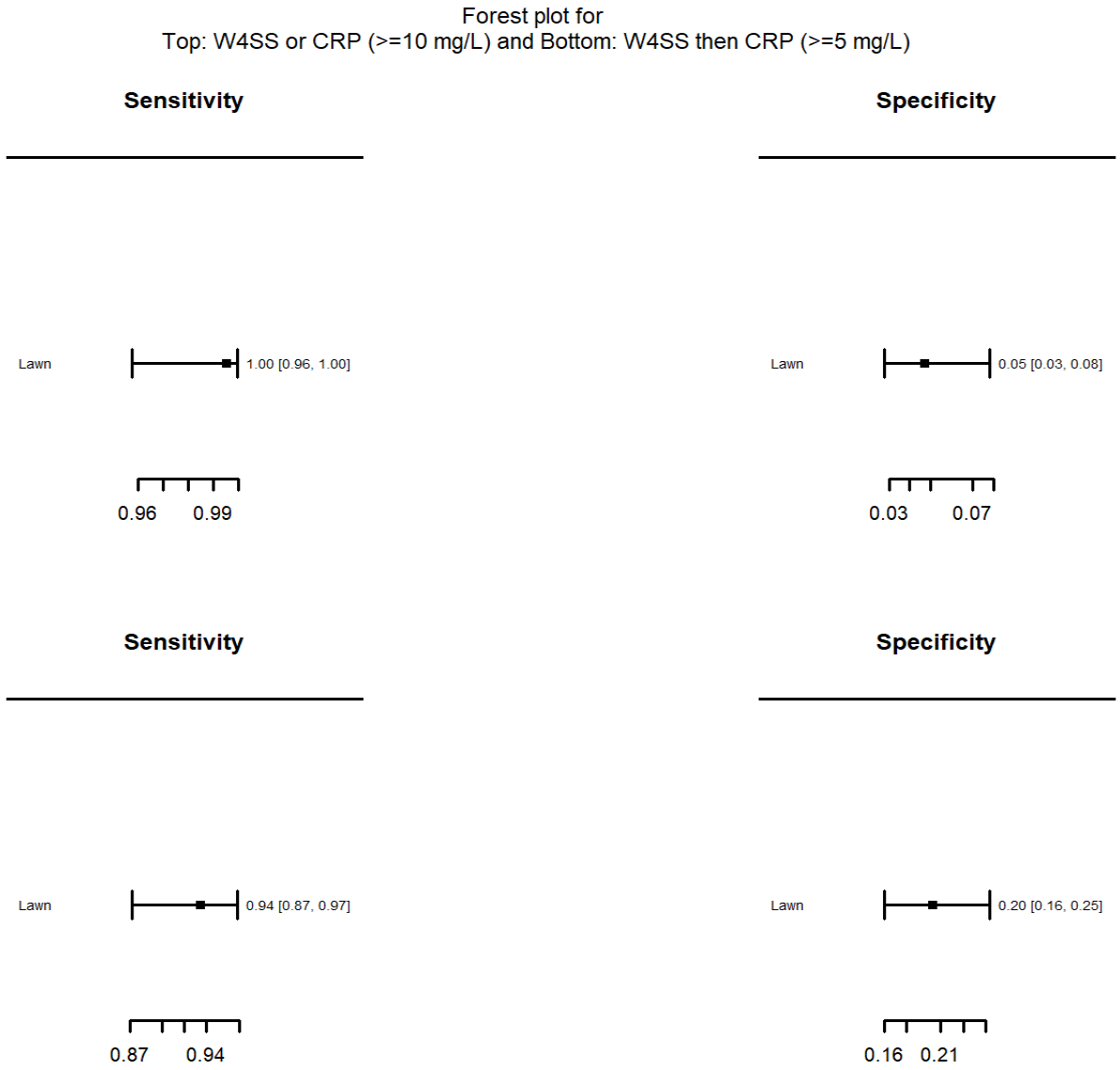
Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)



Forest plot for
BMI (<18.5 kg/m²)




Forest plot for
Lymphadenopathy
Sensitivity
Specificity
Bjerrum
0.11 [0.02, 0.43]
Huerga
0.27 [0.15, 0.44]
Thit
0.25 [0.05, 0.70]
0.02
0.36
0.70
Bjerrum
0.93 [0.81, 0.97]
Huerga
0.91 [0.86, 0.94]
Thit
0.86 [0.74, 0.93]
0.74
0.86
0.97


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]
0.96
0.99
Lawn
0.05 [0.03, 0.08]
0.03
0.07
Sensitivity
Specificity
Lawn
0.94 [0.87, 0.97]
0.87
0.94
Lawn
0.20 [0.16, 0.25]
0.16
0.21

Forest plot for
W4SS or CXR (abnormal)

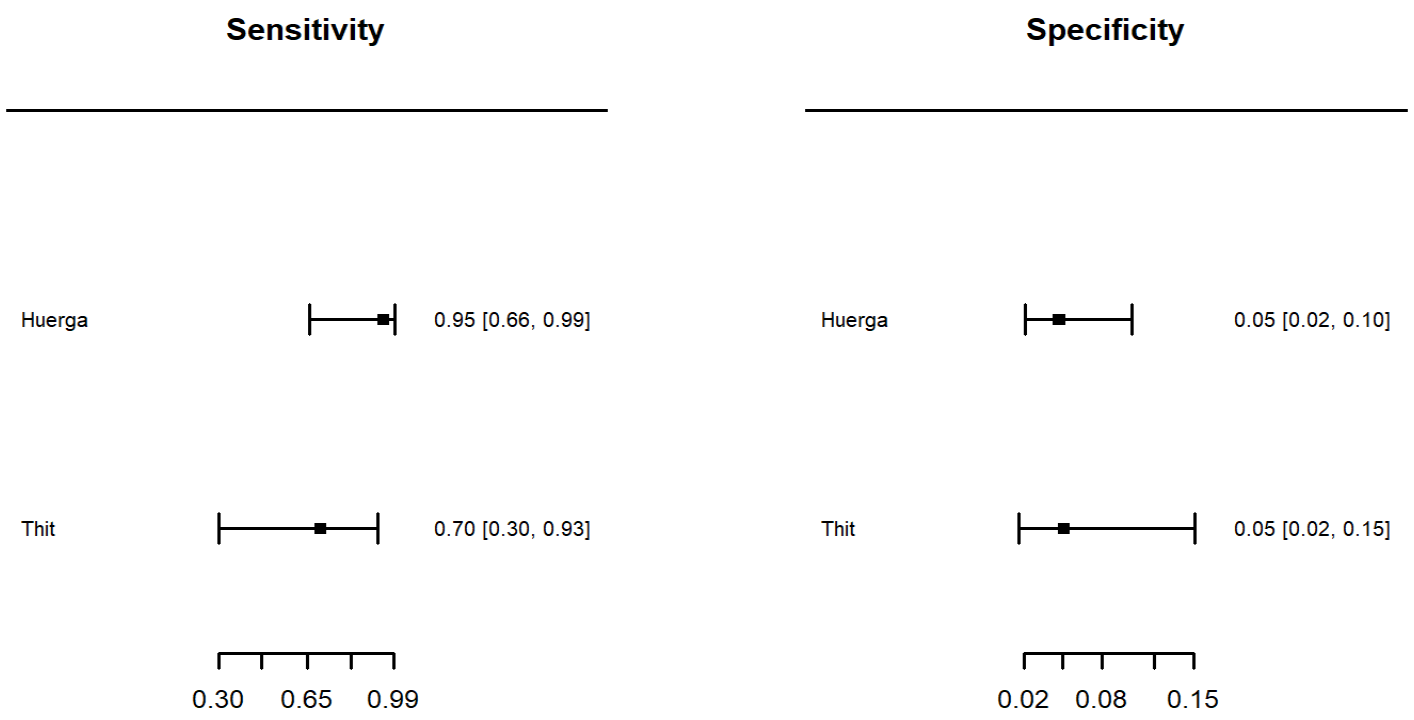
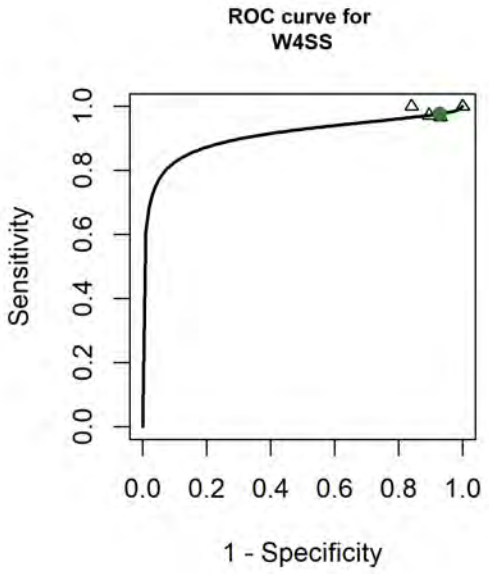
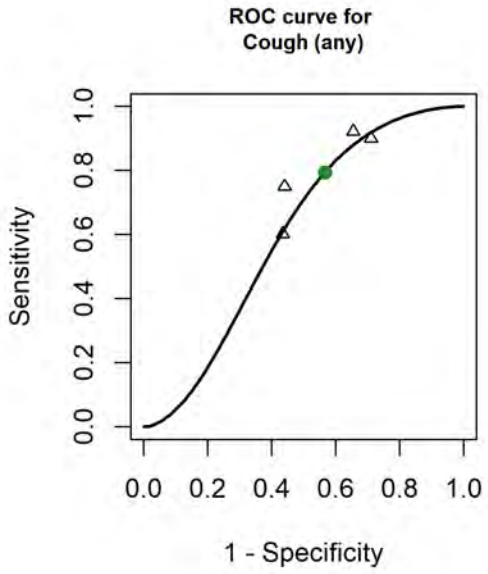
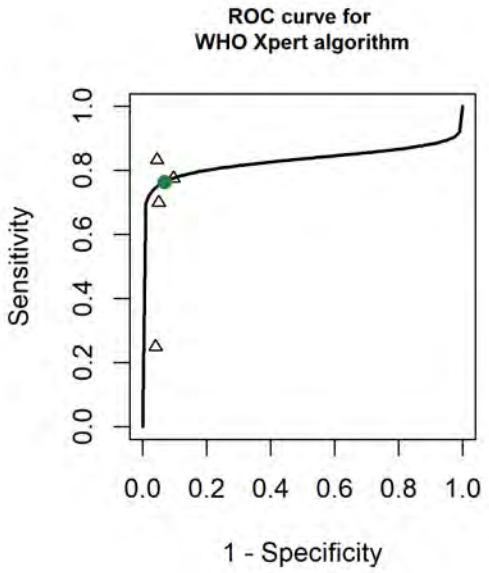


Figure 8-7: Summary receiver operating characteristics curves for each screening test/strategy (for tests/strategies with >=4 studies available)

Figure 8-7A - Summary receiver operating characteristics curves for each screening test/strategy using culture as a reference standard







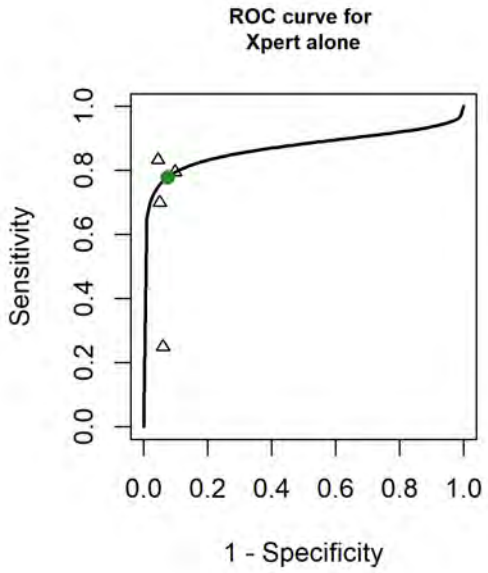
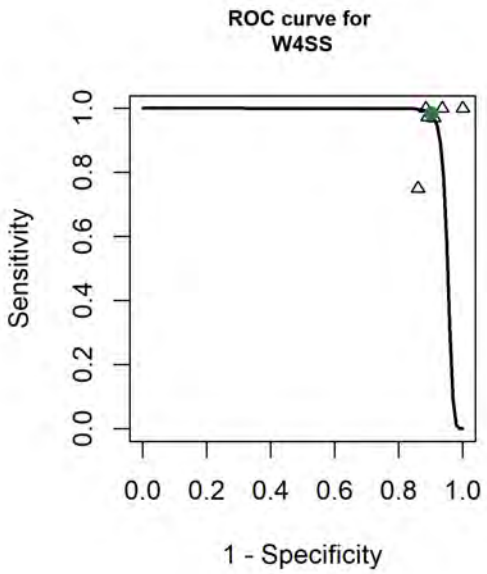
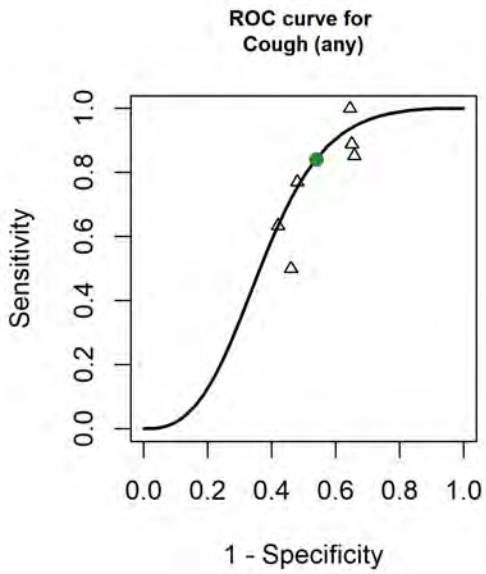
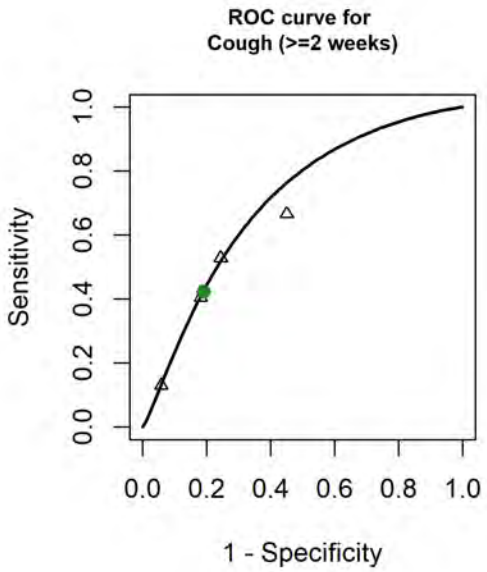
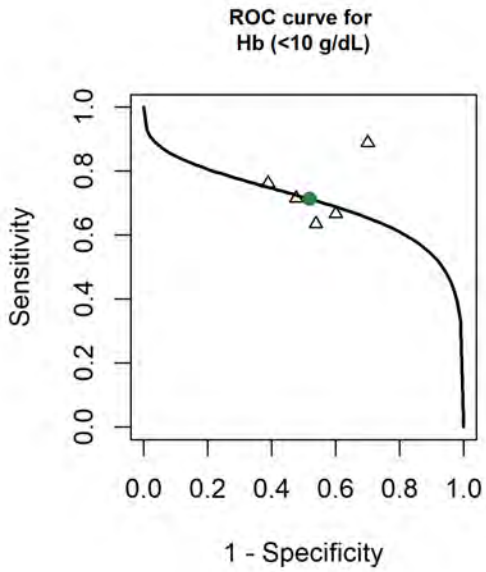


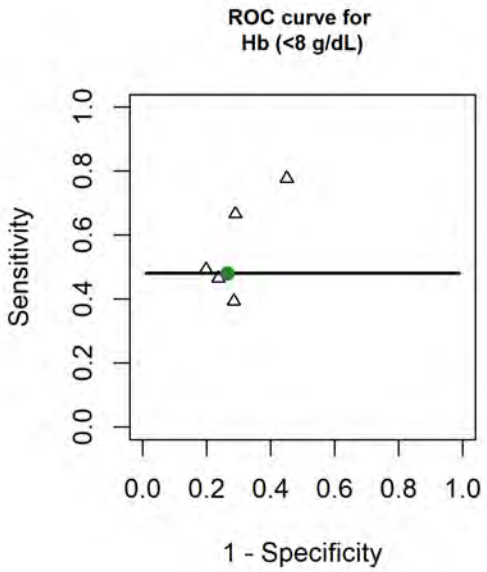
Figure 8-7B - Summary receiver operating characteristics curves for each screening test/strategy using Xpert as a reference standard






ROC curve for
Cough (>=2 weeks)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


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


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

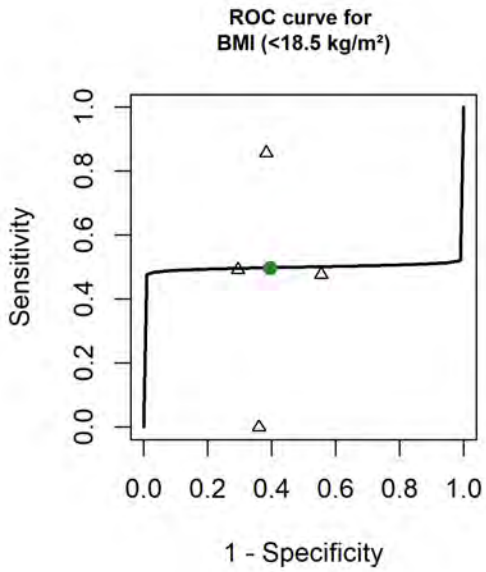
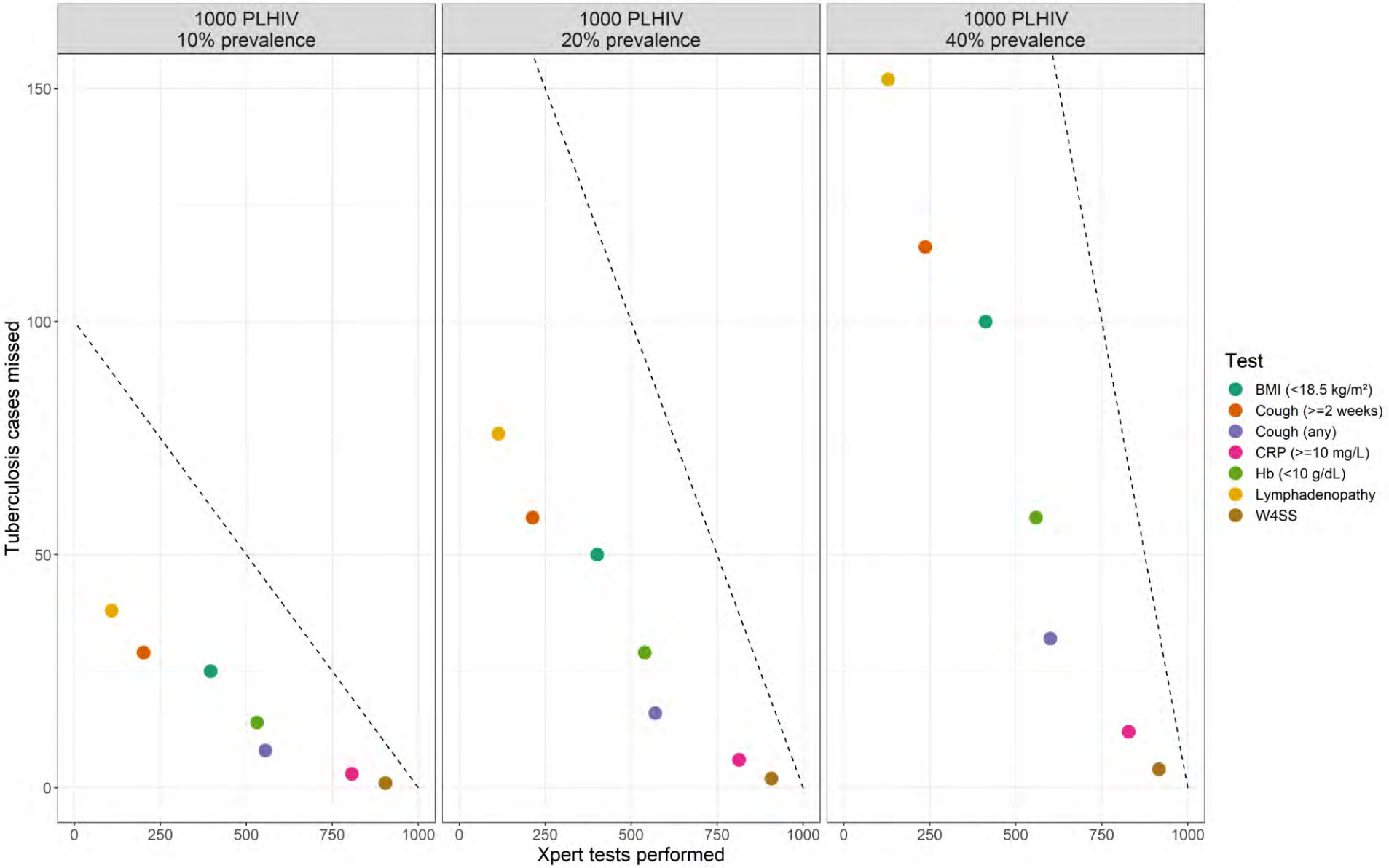

ROC curve for
BMI (<18.5 kg/m²)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

Figure 8-8: Plot comparing number of tuberculosis cases missed with number of Xpert tests required for different tuberculosis screening tests when screening a population of 1000 persons†



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

8.3 Appendix for Chapter 4

Table 8-24: 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 8-25: 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
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 8-26: 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 8-27: 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 8-28: 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 8-29: 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

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 8-30: 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)	-
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

Variable†	All	Bjerrum	Gupta-Wright Intervention††	Gupta-Wright Control††	Huerga	Lawn	Thit
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 8-31: 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 8-32: 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 8-33: 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 8-9: Study flow diagram

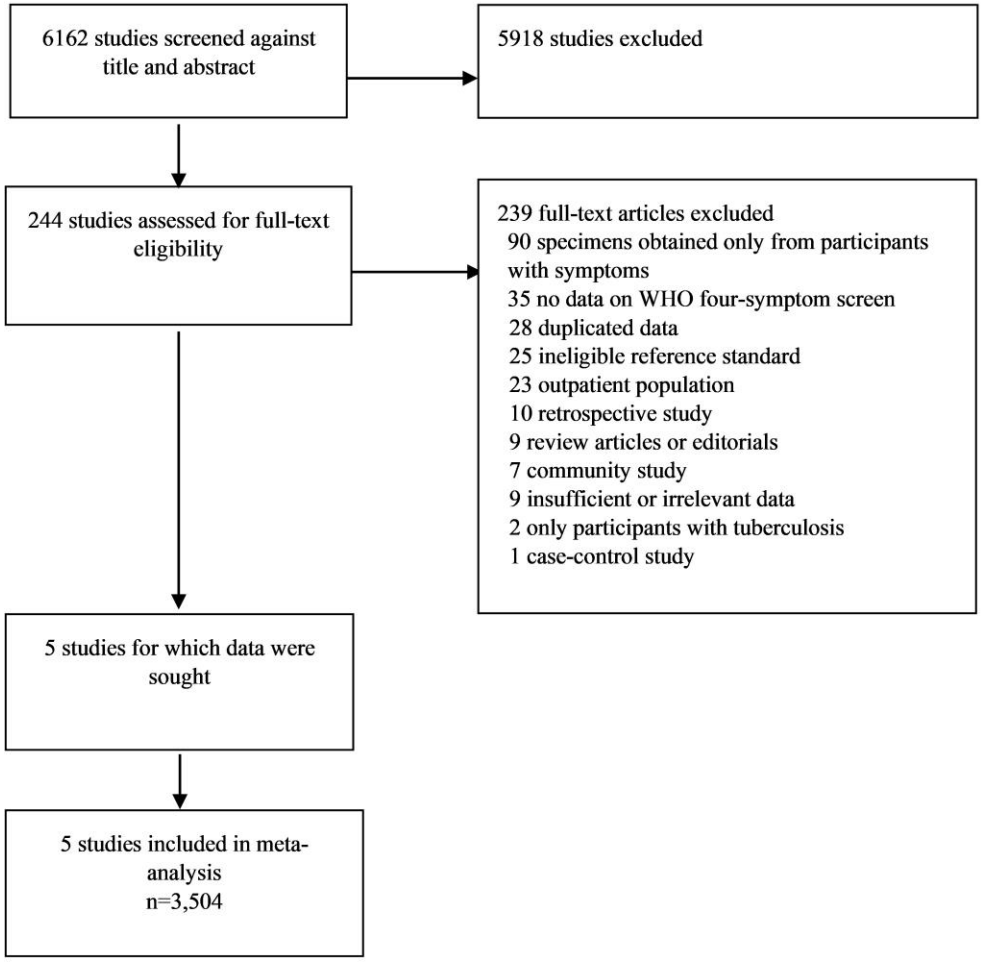
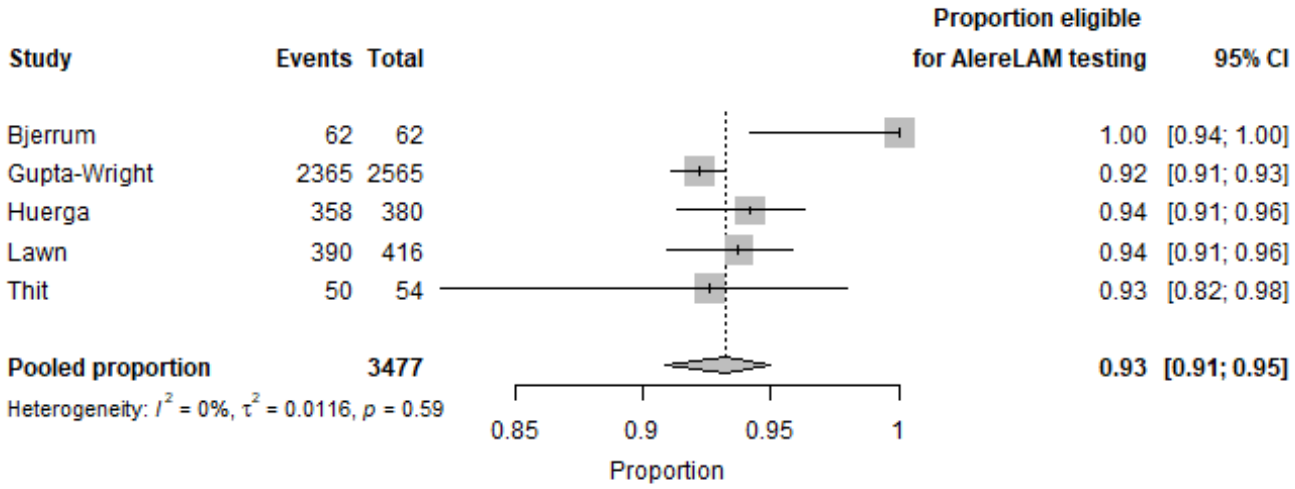
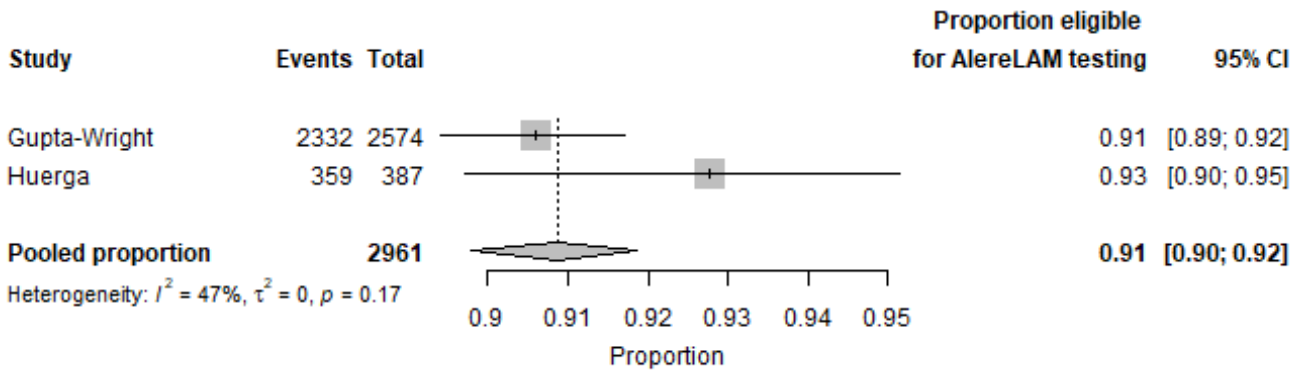


Figure 8-10: Forest plots of studies reporting proportion of inpatients eligible for AlereLAM testing according to WHO AlereLAM inpatient algorithm

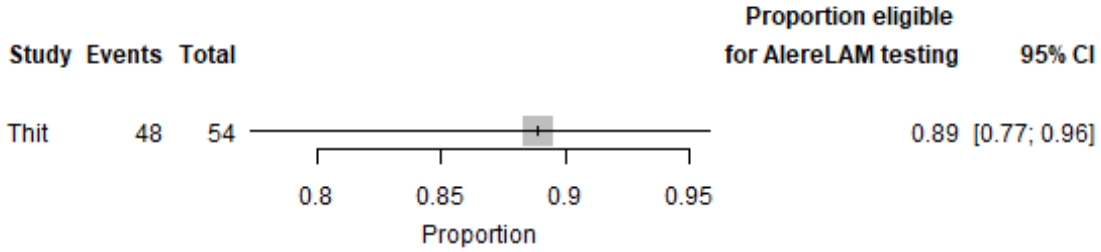
Random-effects meta-analysis of proportion studies reporting screen positivity by W4SS or CD4 <=200 cells/µL in HIV-infected inpatients



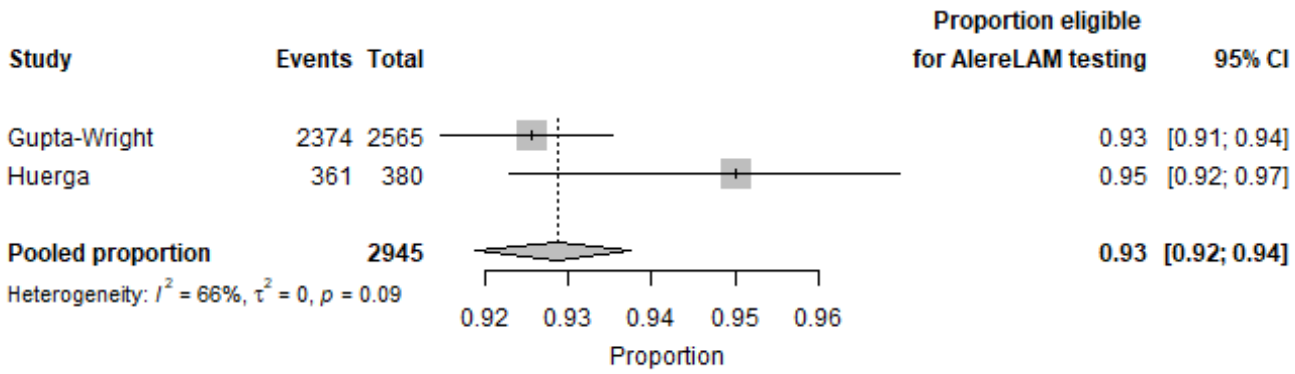
Random-effects meta-analysis of proportion studies reporting screen positivity by W4SS or WHO-defined danger signs in HIV-infected inpatients



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

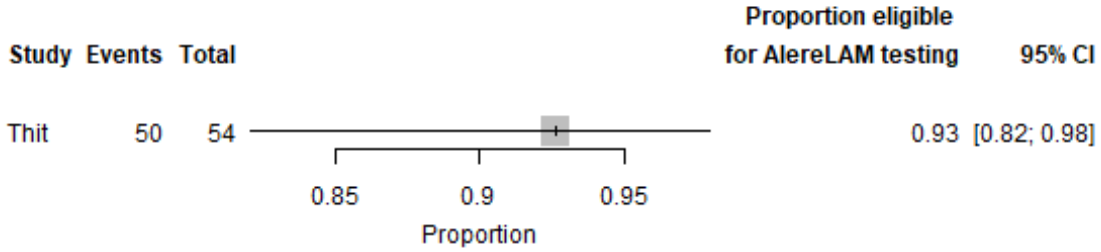
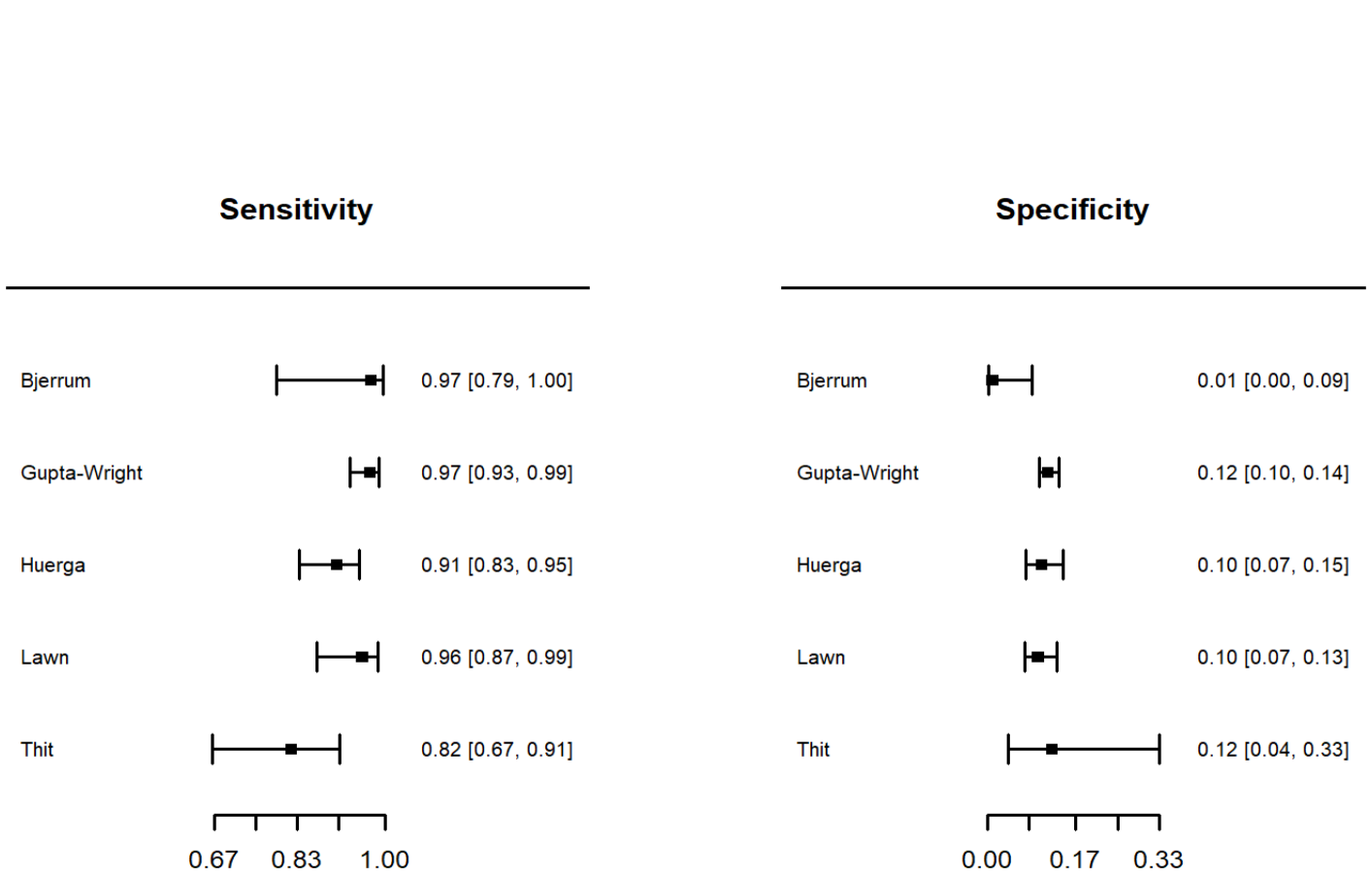
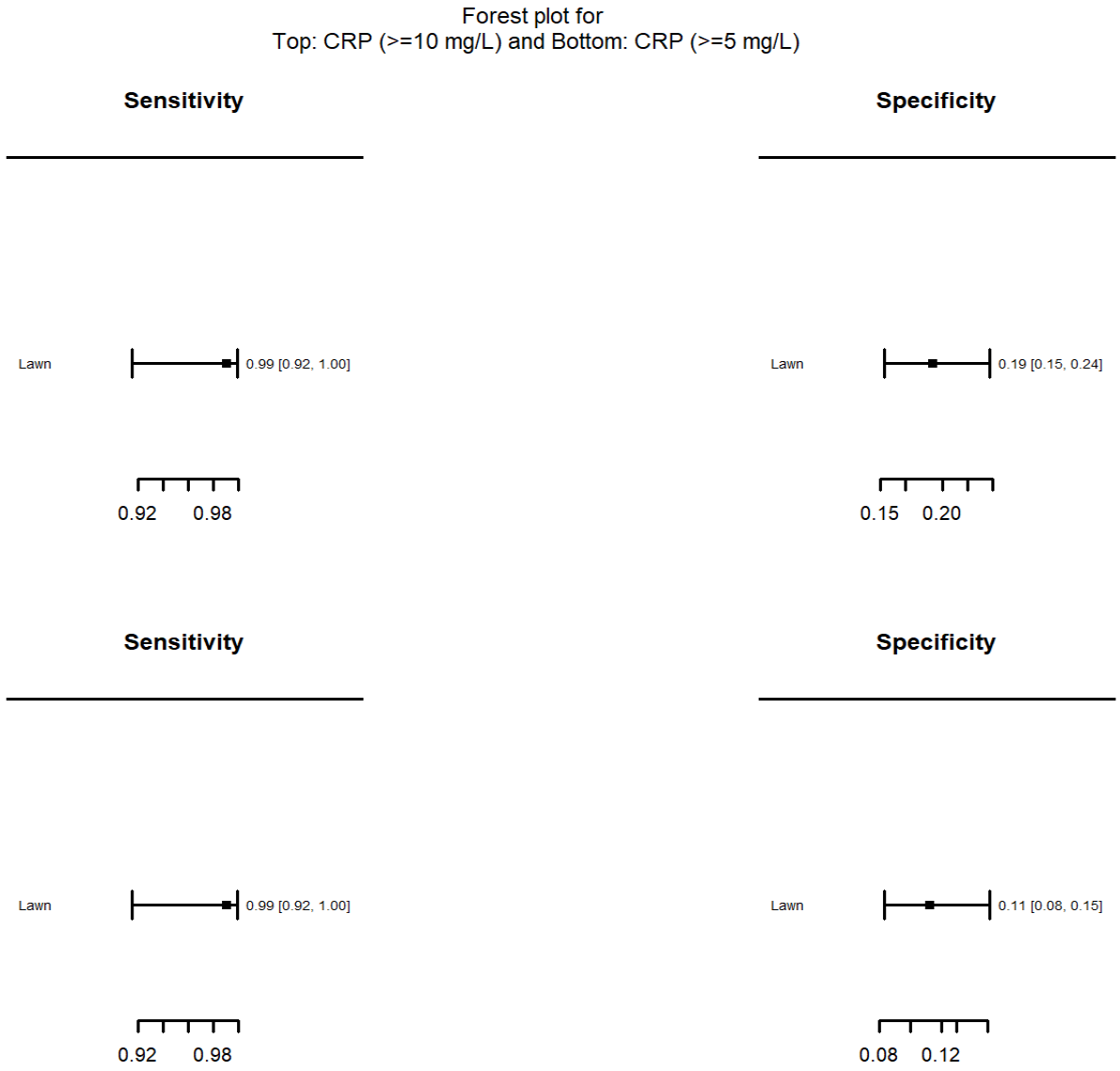


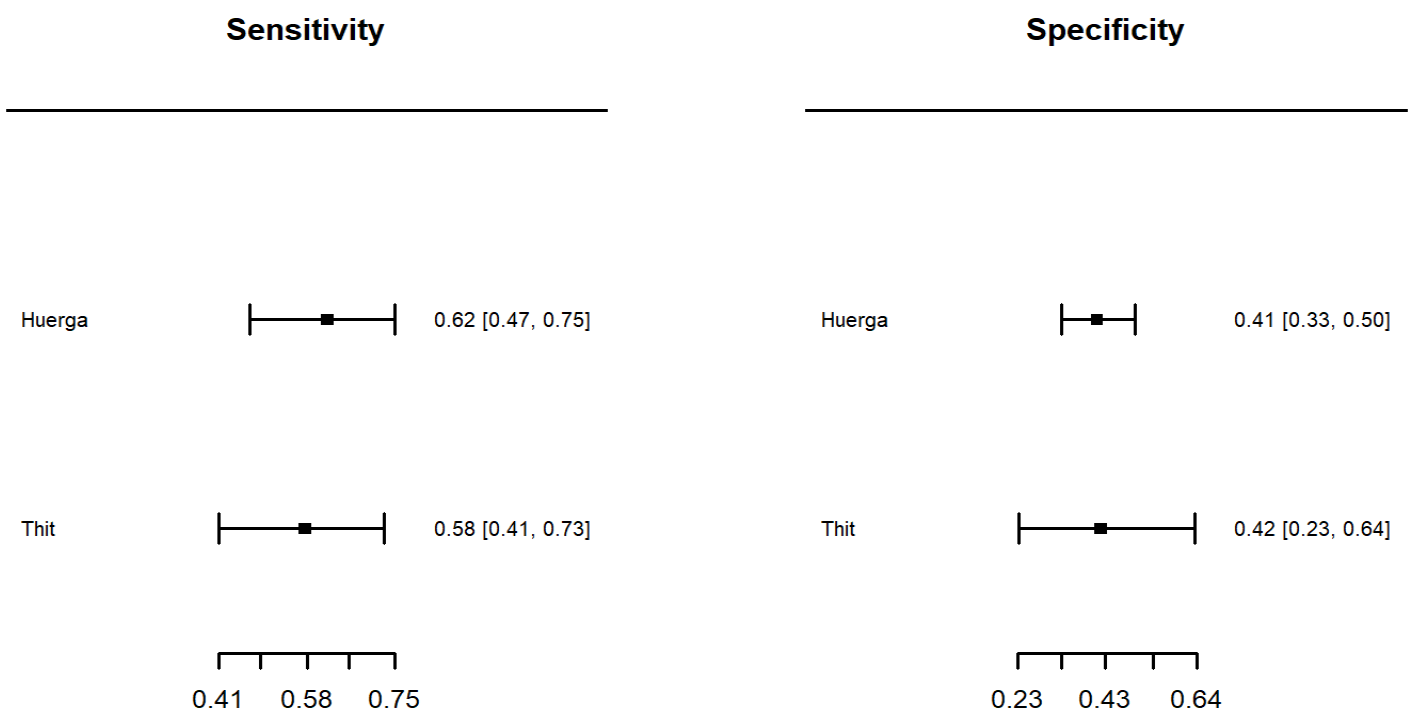
Figure 8-11: Forest plots of sensitivity and specificity estimates

Figure 8-11A: Forest plots of sensitivity and specificity estimates for each screening test/strategy using AlereLAM as a reference standard

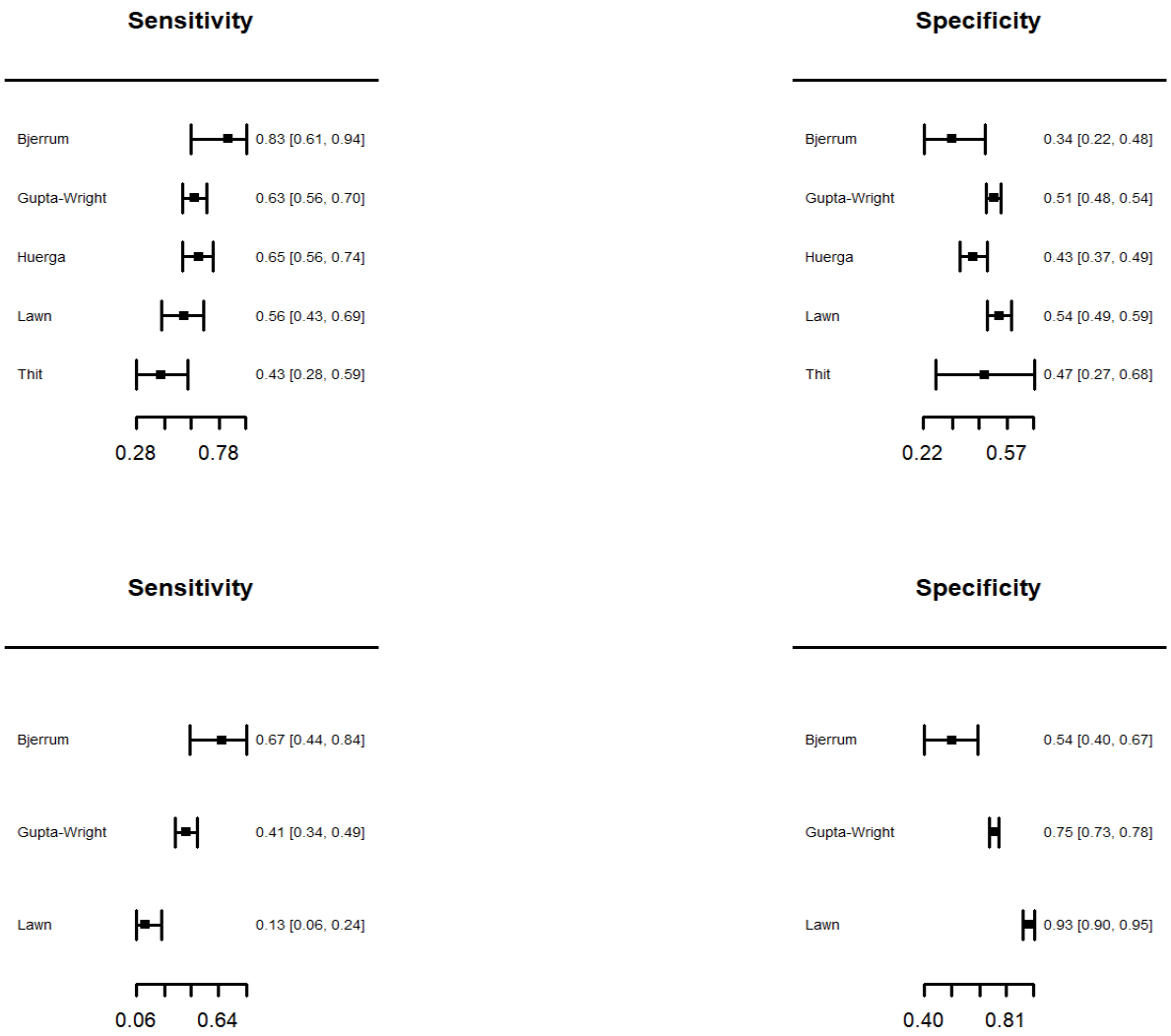



Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
0.99 [0.92, 1.00]
0.19 [0.15, 0.24]
0.92
0.98
0.15
0.20
Sensitivity
Specificity
Lawn
0.99 [0.92, 1.00]
0.11 [0.08, 0.15]
0.92
0.98
0.08
0.12

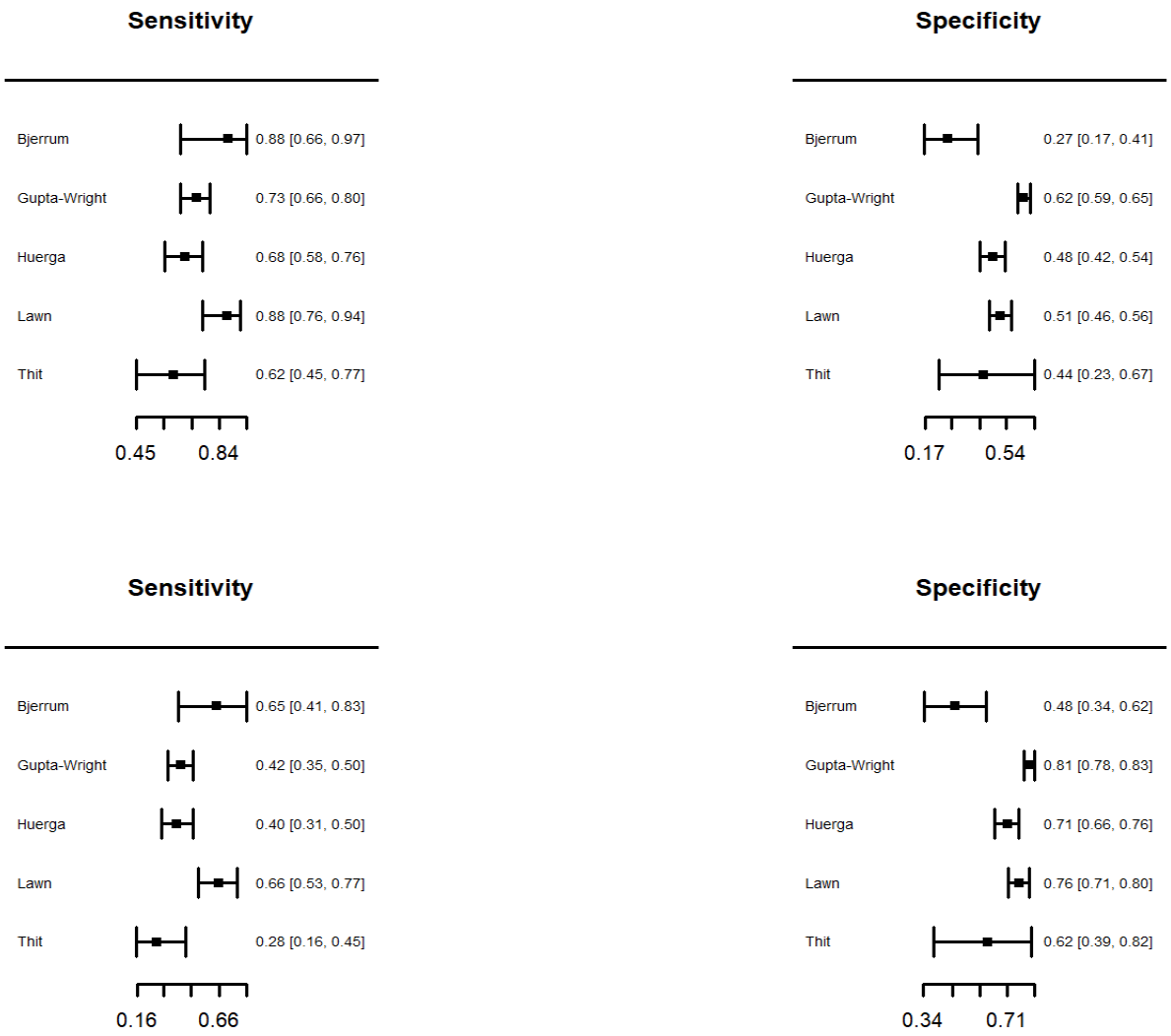
Forest plot for
CXR (abnormal)



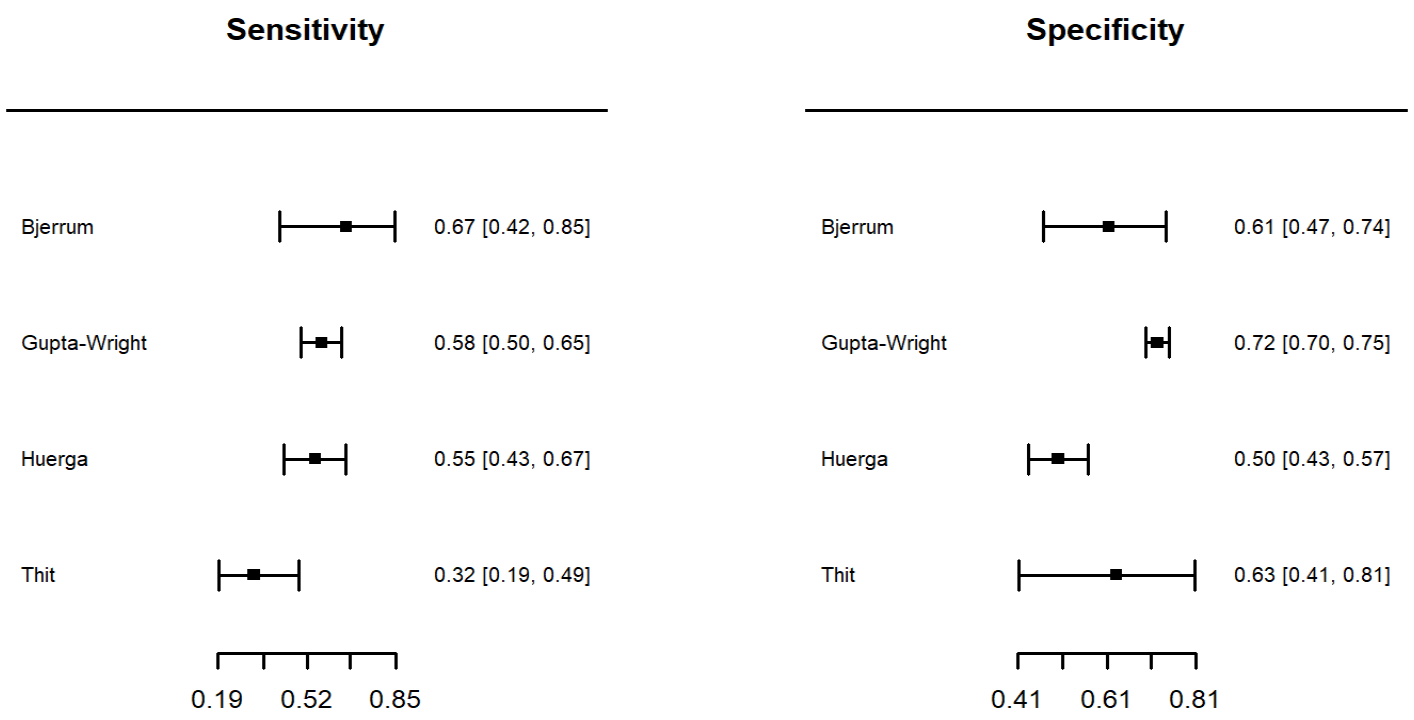
Forest plot for
Top: Cough (any) and Bottom: Cough (>=2 weeks)



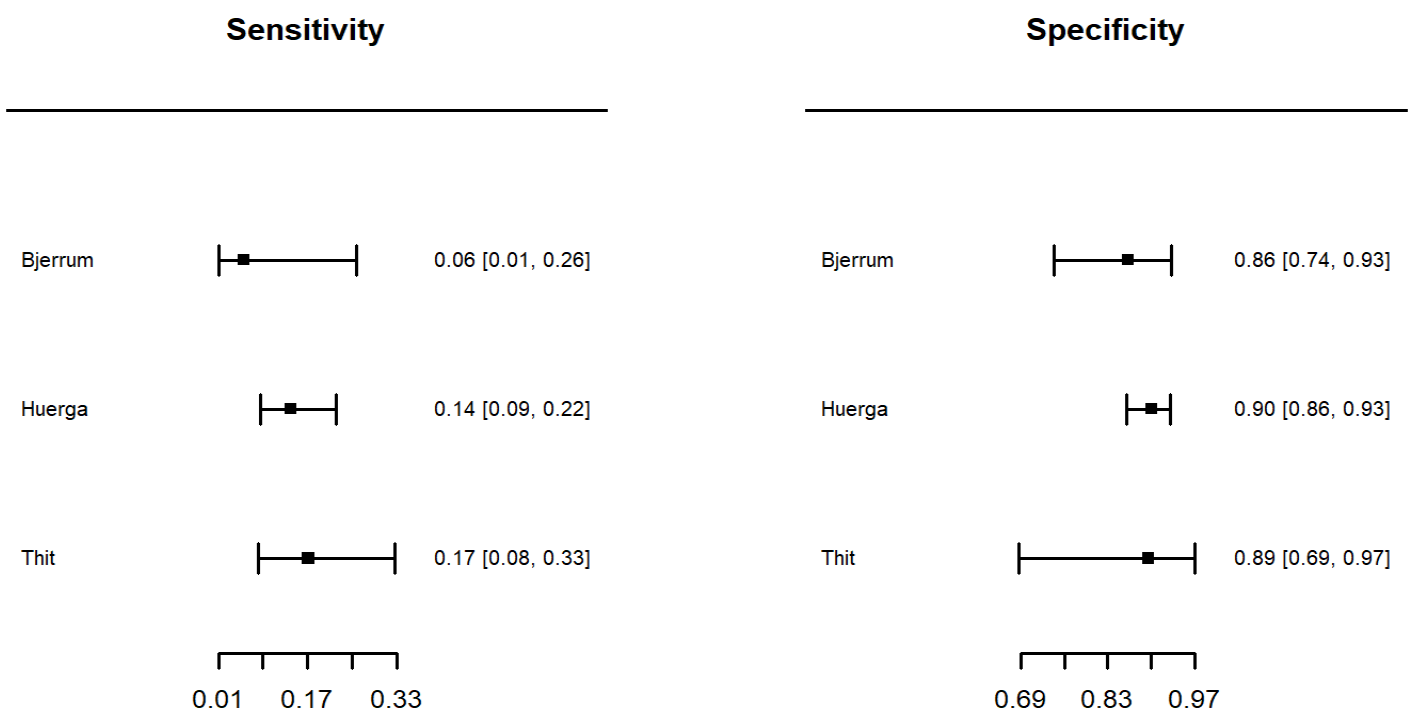
Forest plot for
Top: Hb (<10 g/dL) and Bottom: Hb (<8 g/dL)

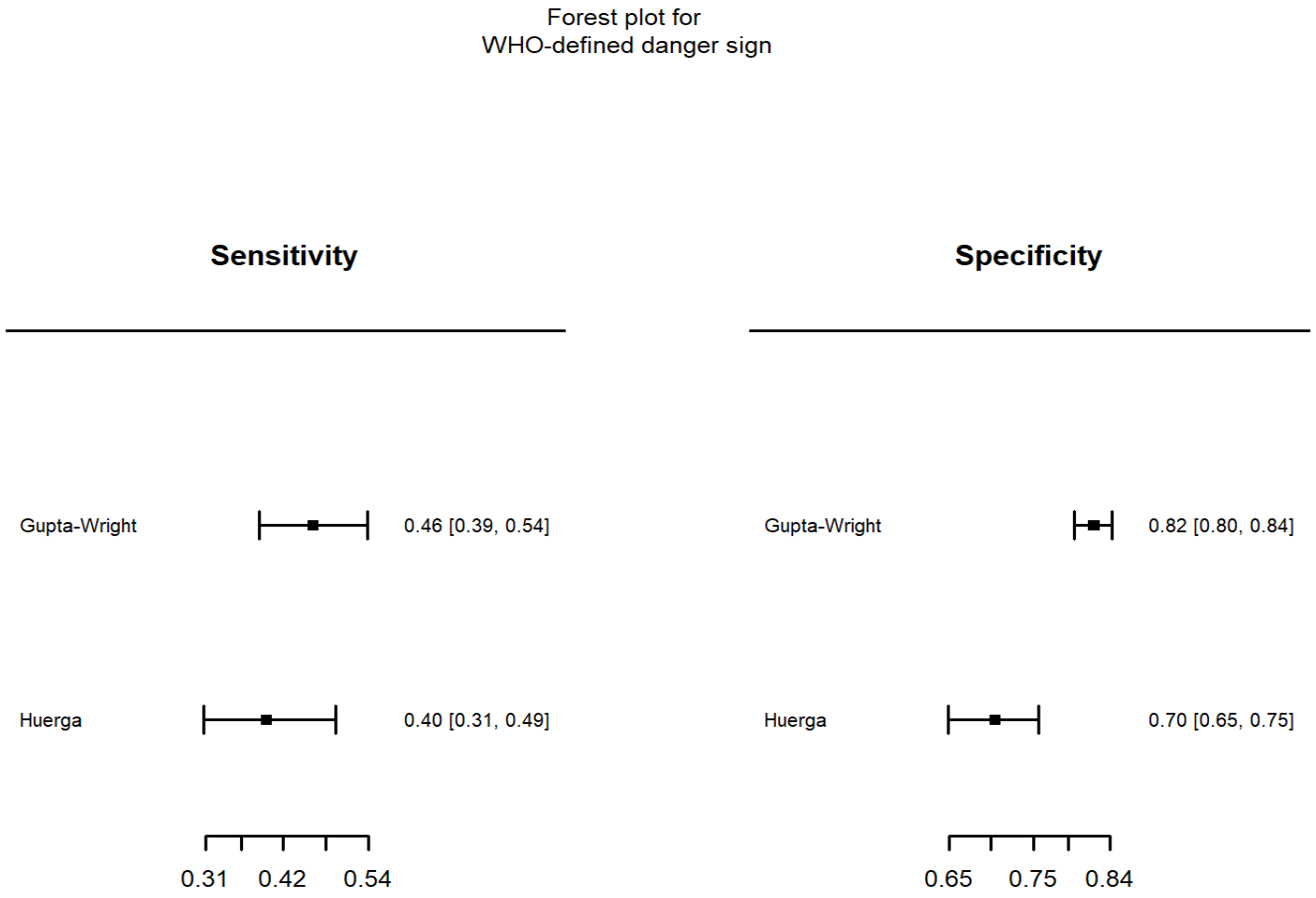


Forest plot for
BMI (<18.5 kg/m²)

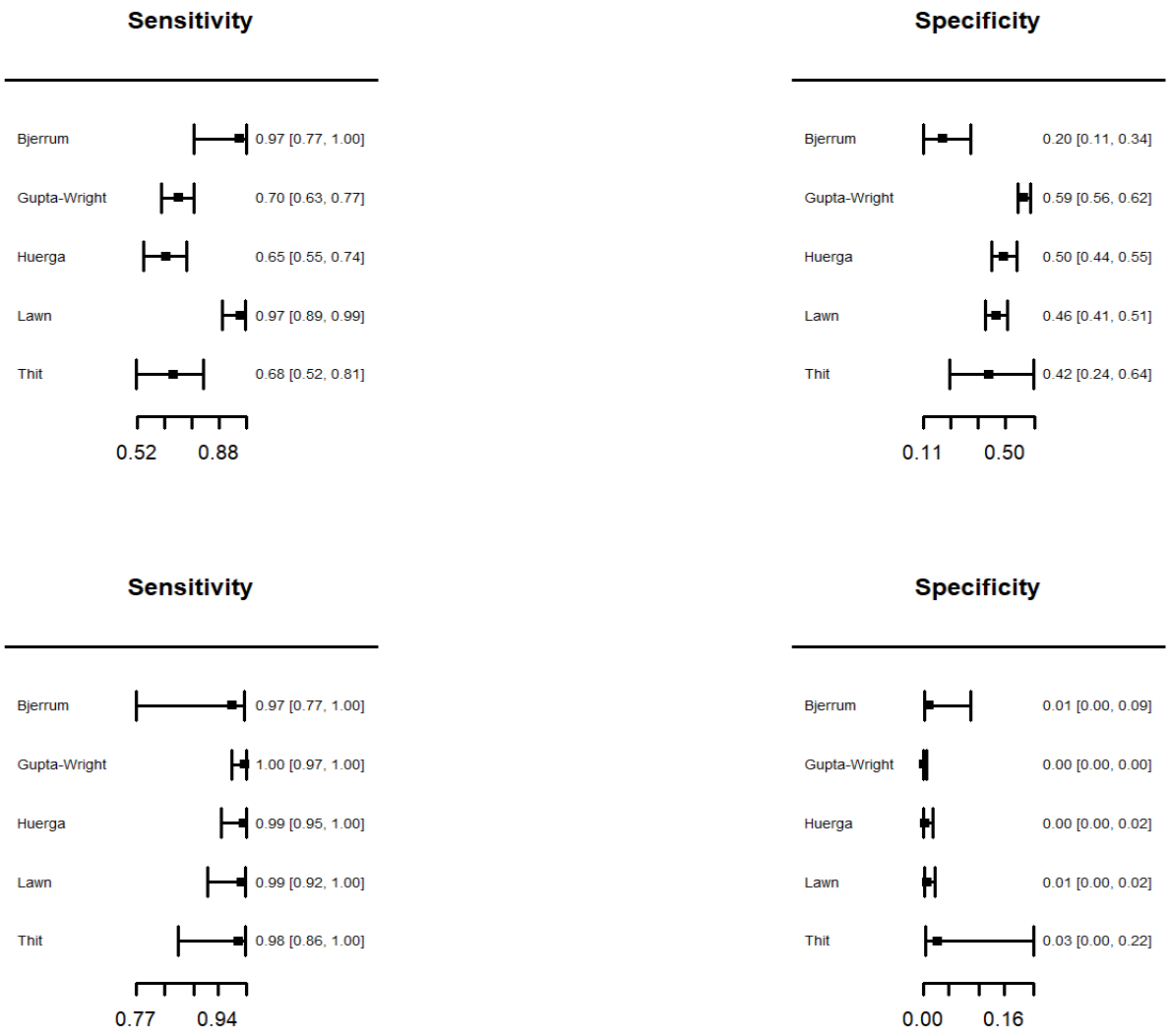


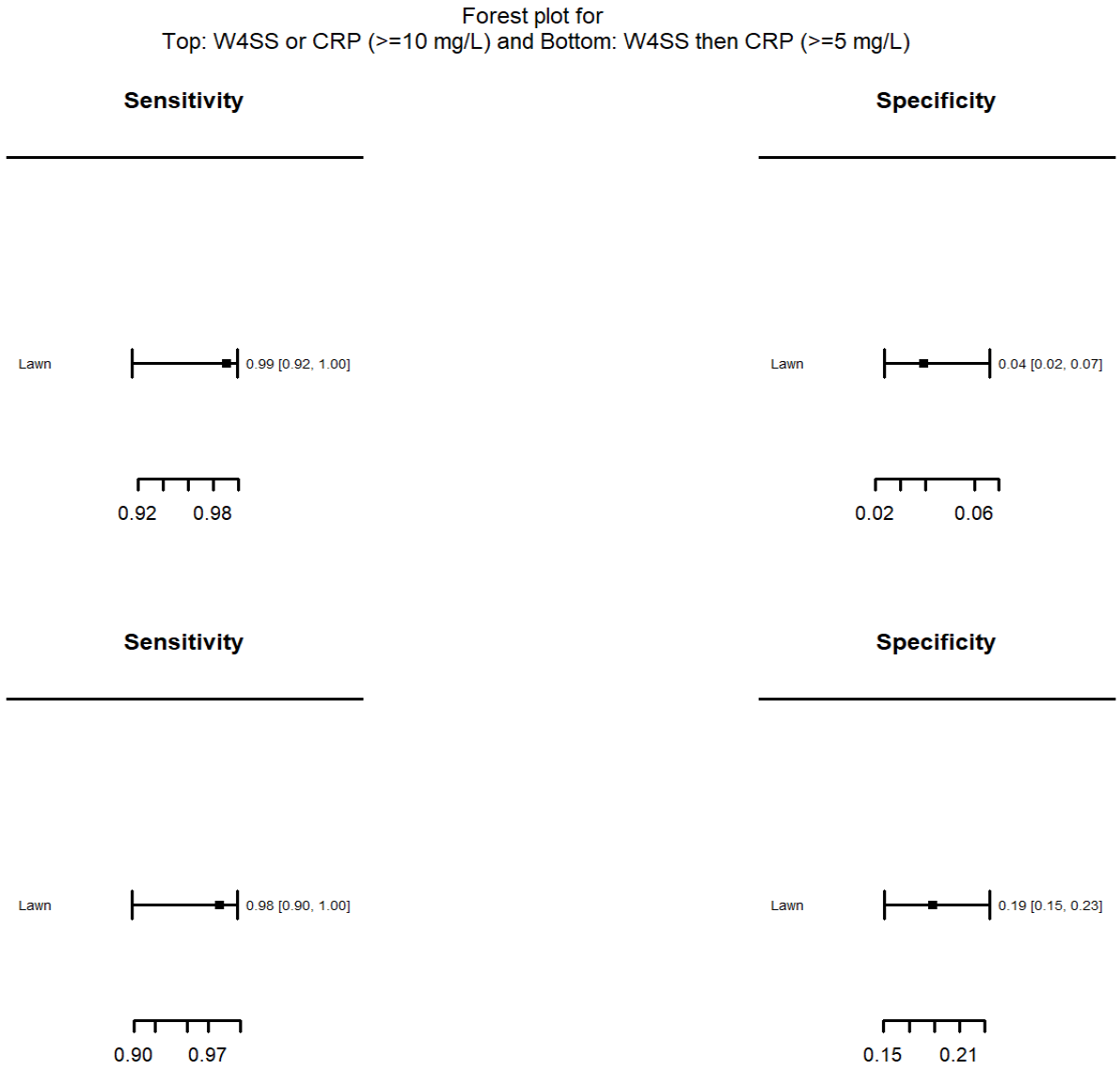
Forest plot for
Lymphadenopathy




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]
0.92
0.98
Lawn
0.04 [0.02, 0.07]
0.02
0.06
Sensitivity
Specificity
Lawn
0.98 [0.90, 1.00]
0.90
0.97
Lawn
0.19 [0.15, 0.23]
0.15
0.21

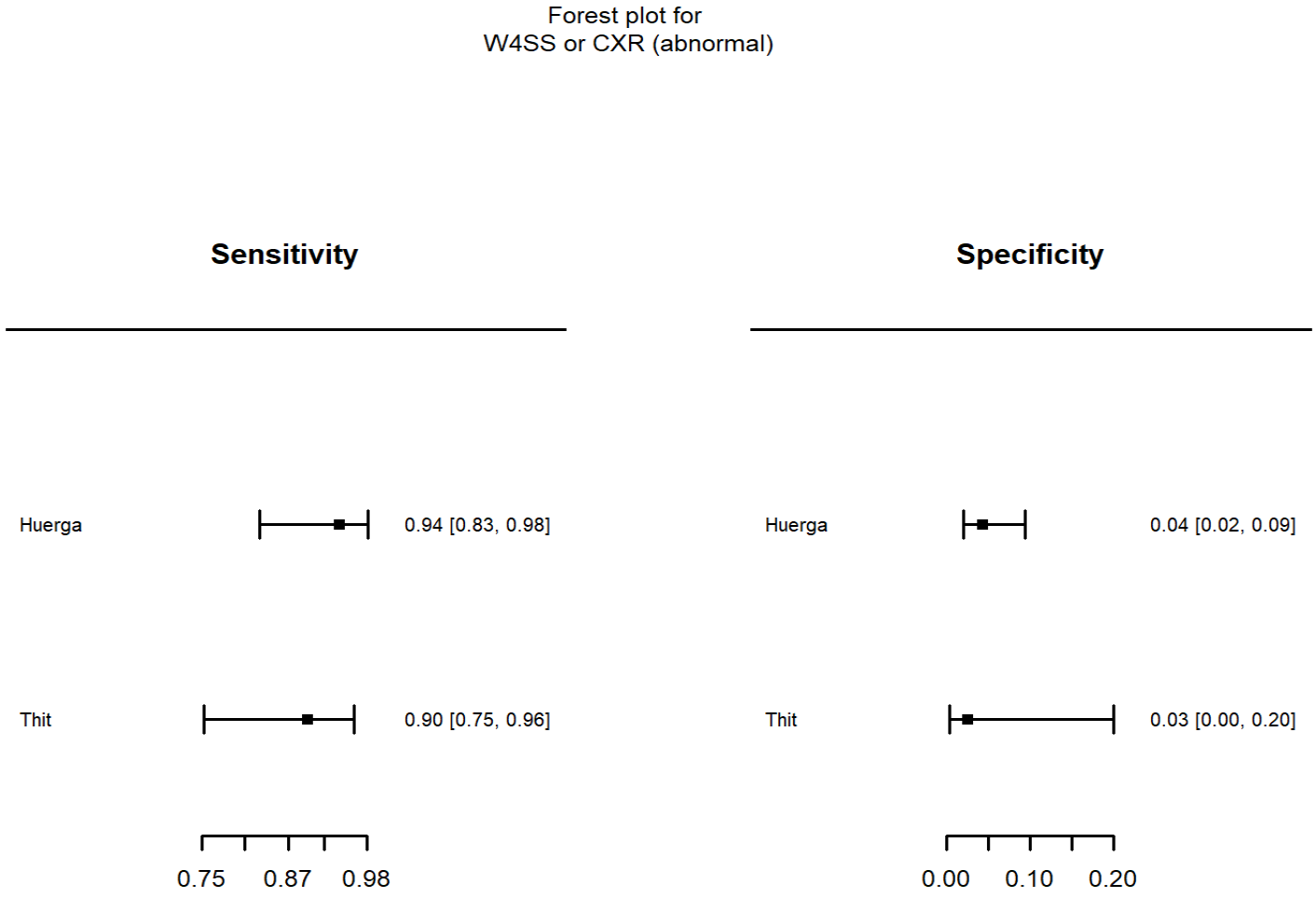
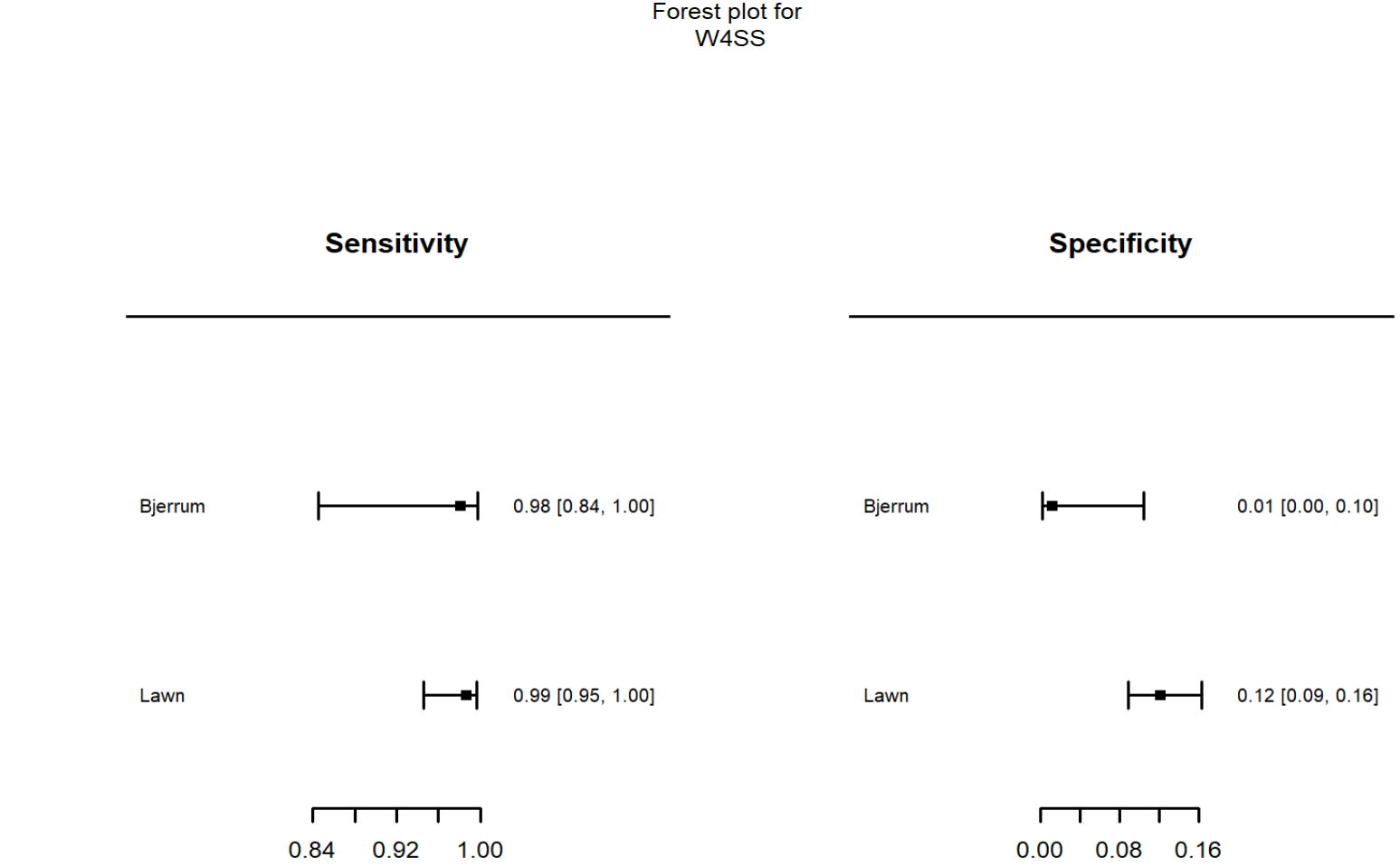
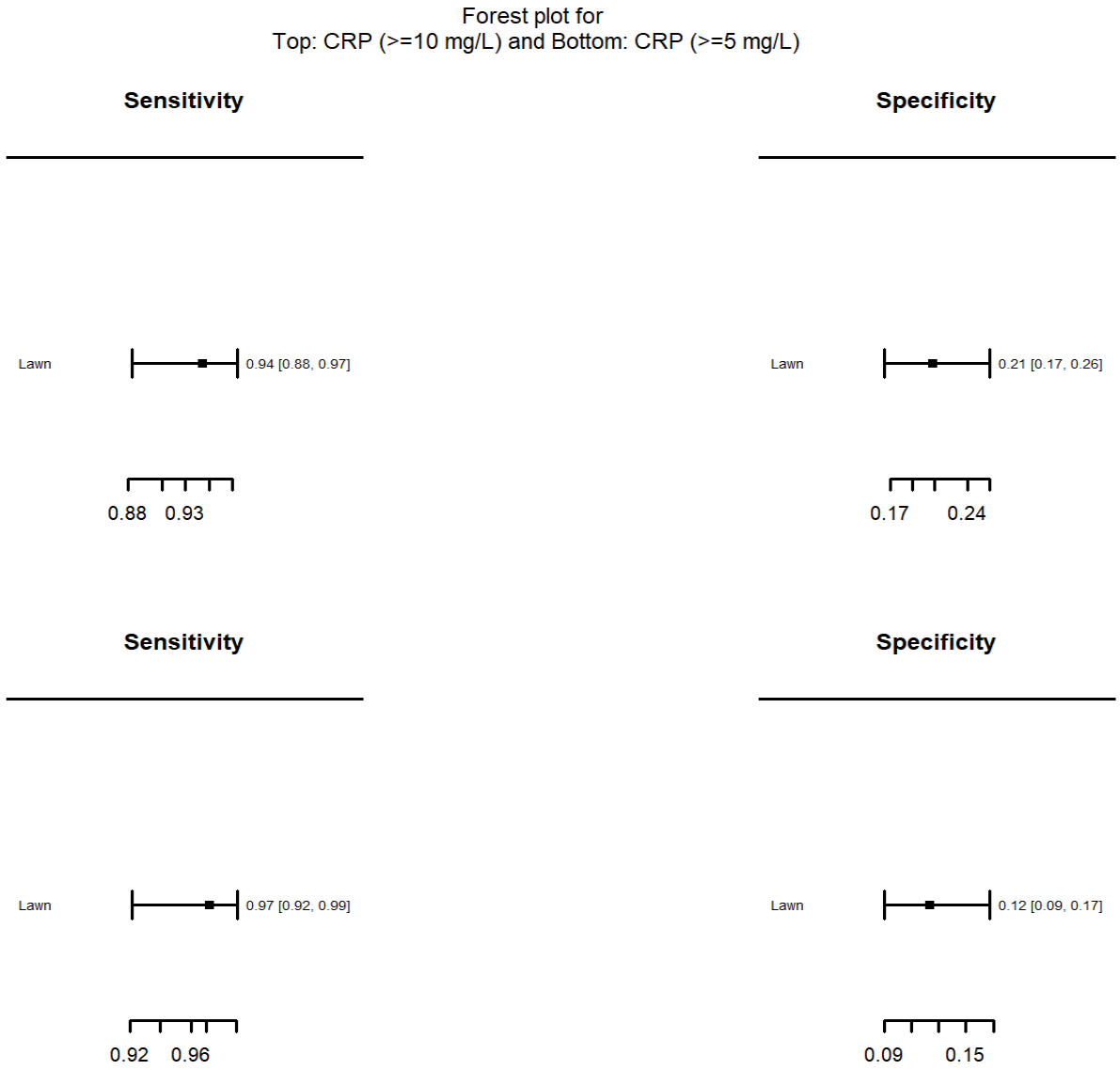
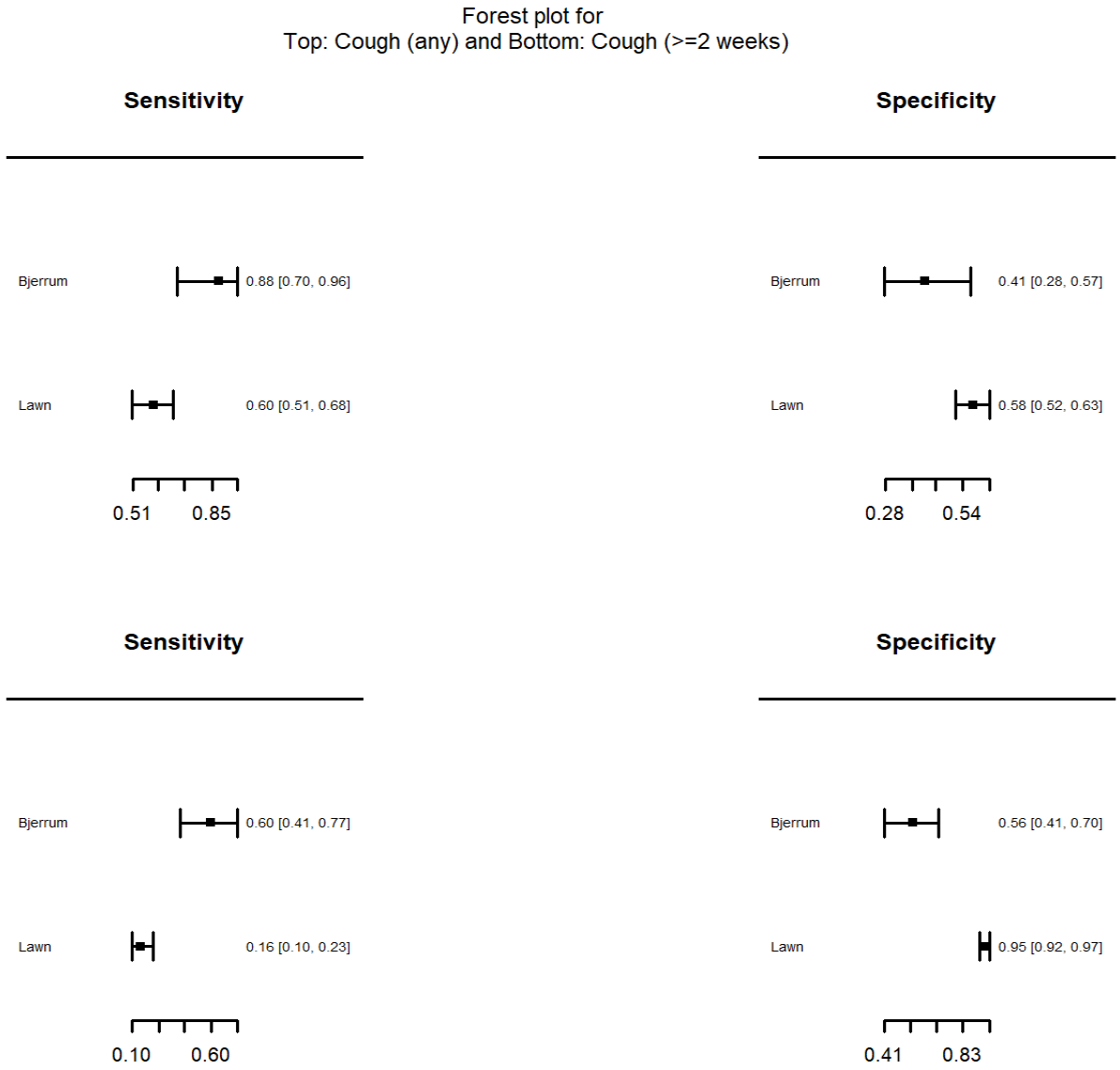
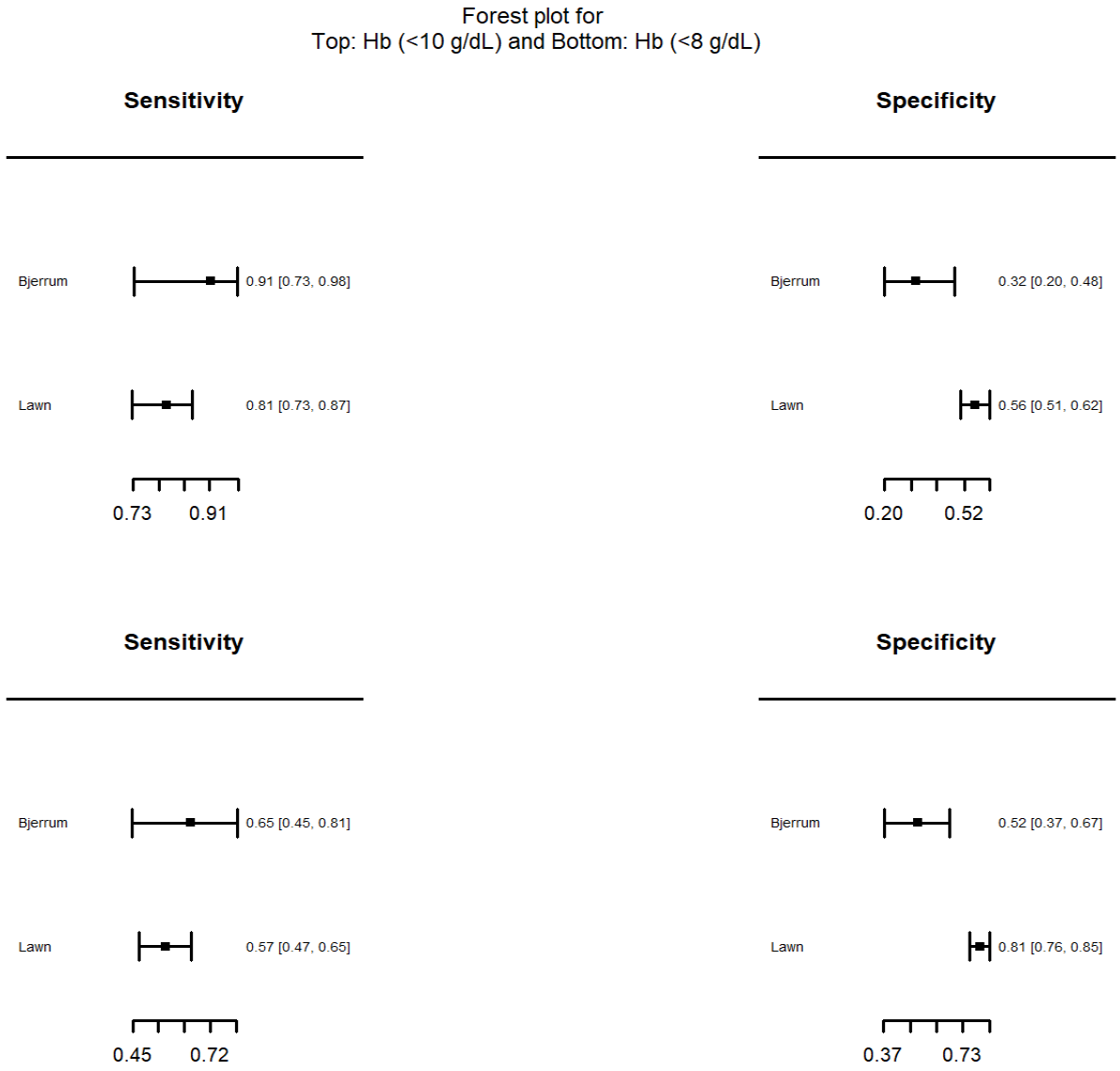

Forest plot for
W4SS or CXR (abnormal)
Sensitivity
Specificity
Huerga
0.94 [0.83, 0.98]
Thit
0.90 [0.75, 0.96]
0.75
0.87
0.98
Huerga
0.04 [0.02, 0.09]
Thit
0.03 [0.00, 0.20]
0.00
0.10
0.20

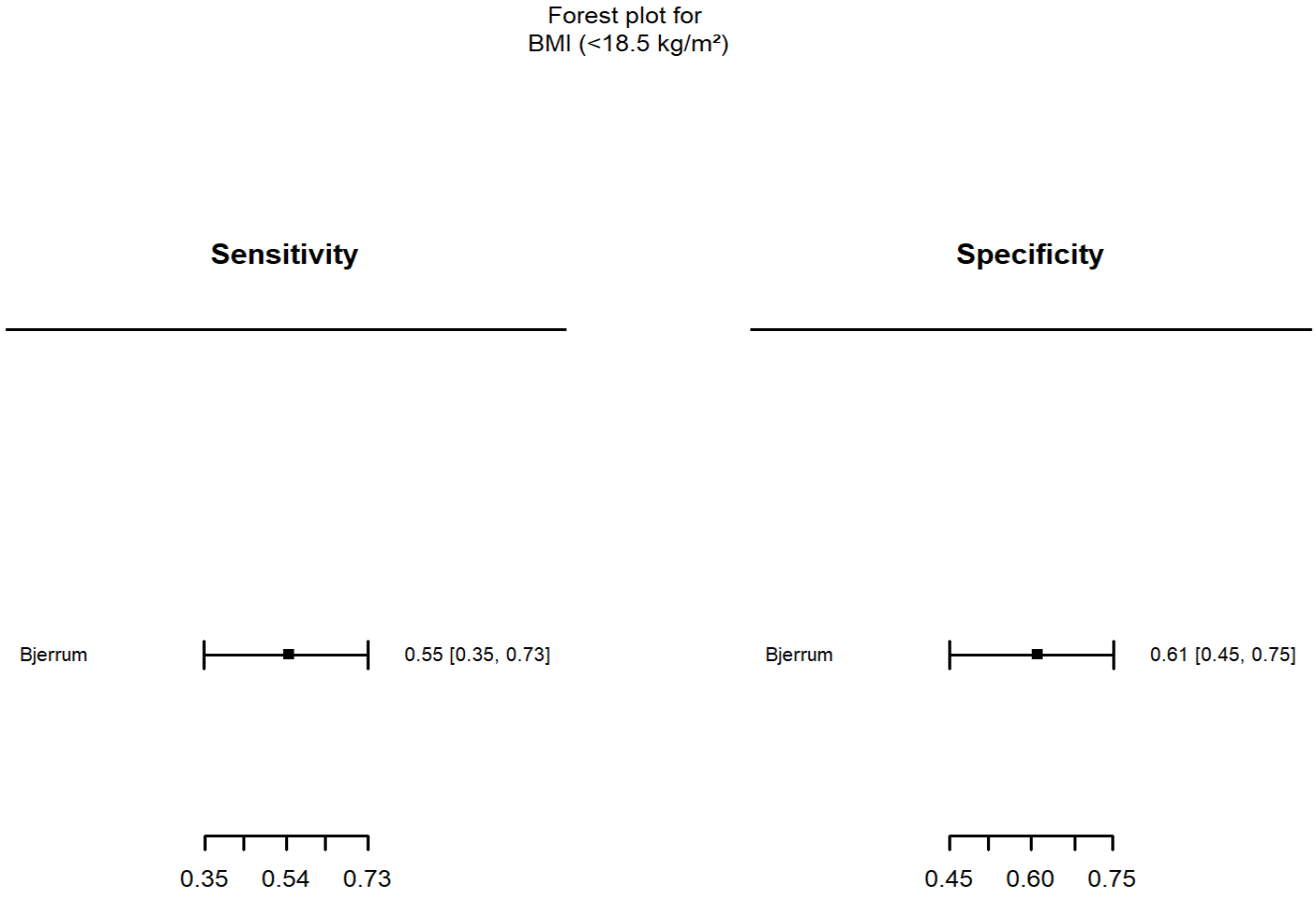
Figure 8-11B - Forest plots of sensitivity and specificity estimates for each screening test/strategy using FujiLAM as a reference standard

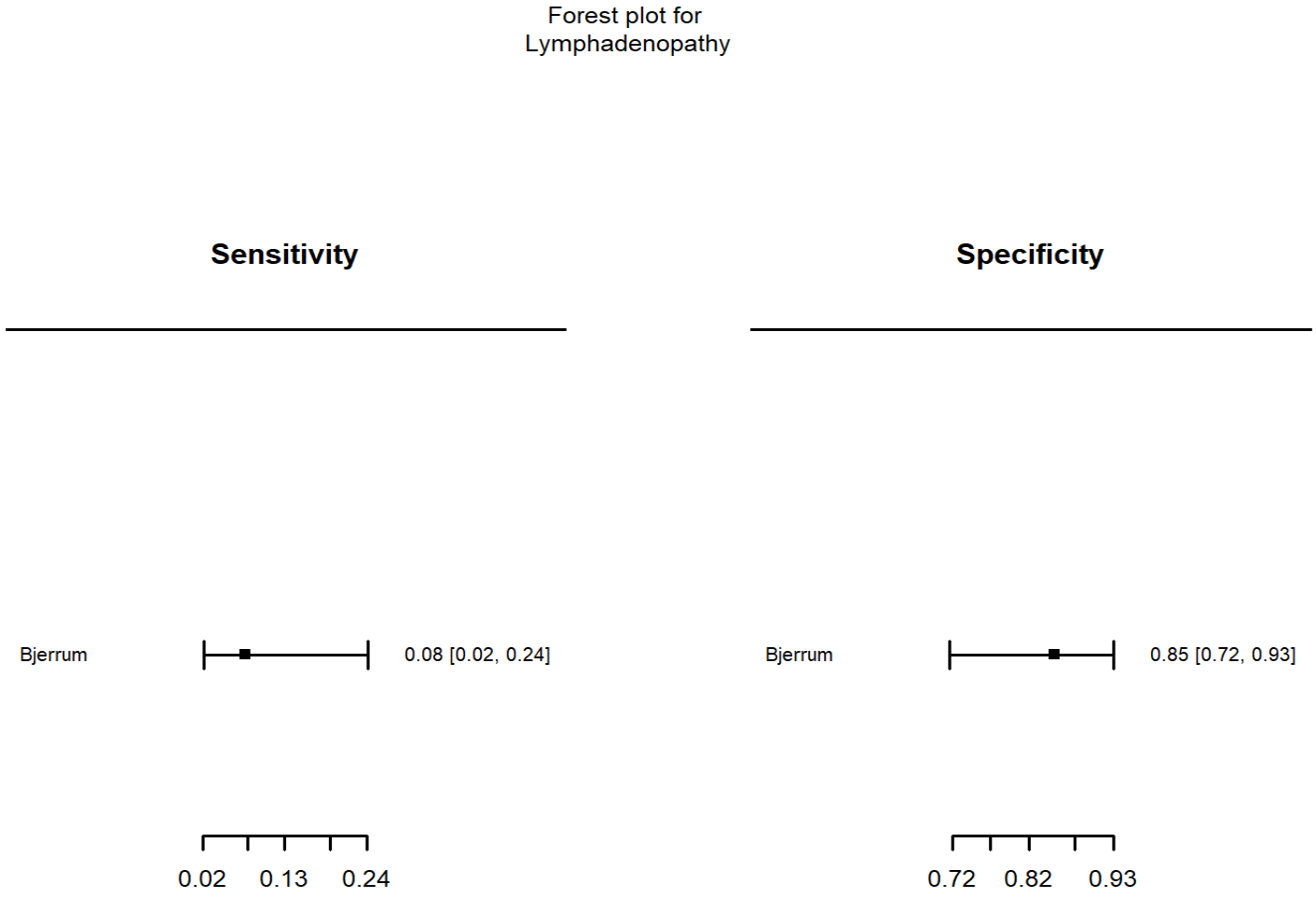


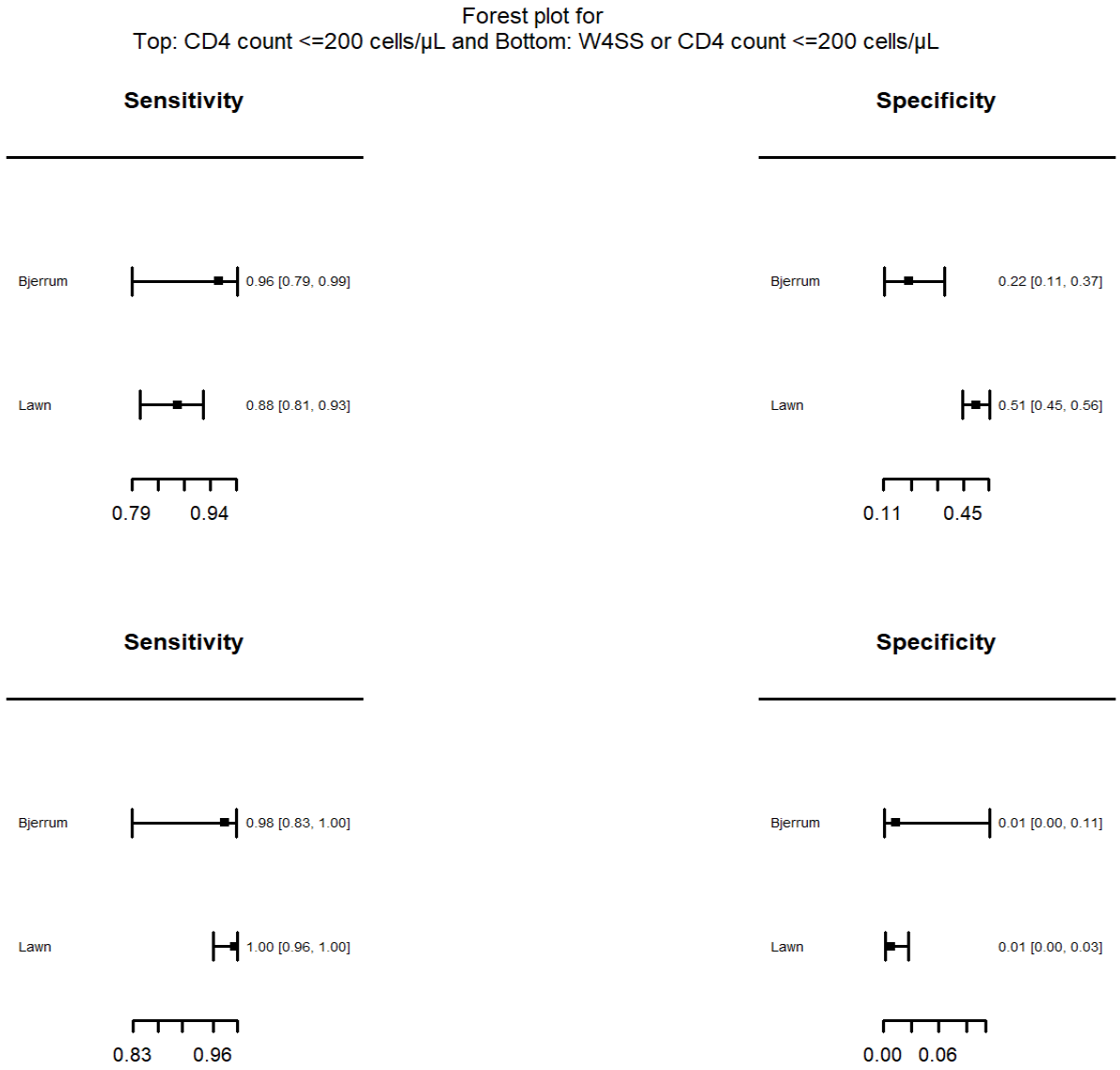

Forest plot for
Top: CRP (>=10 mg/L) and Bottom: CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
0.94 [0.88, 0.97]
0.21 [0.17, 0.26]
0.88
0.93
0.17
0.24
Sensitivity
Specificity
Lawn
0.97 [0.92, 0.99]
0.12 [0.09, 0.17]
0.92
0.96
0.09
0.15


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)
Sensitivity
Specificity
Bjerrum
0.55 [0.35, 0.73]
0.35
0.54
0.73
Bjerrum
0.61 [0.45, 0.75]
0.45
0.60
0.75


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

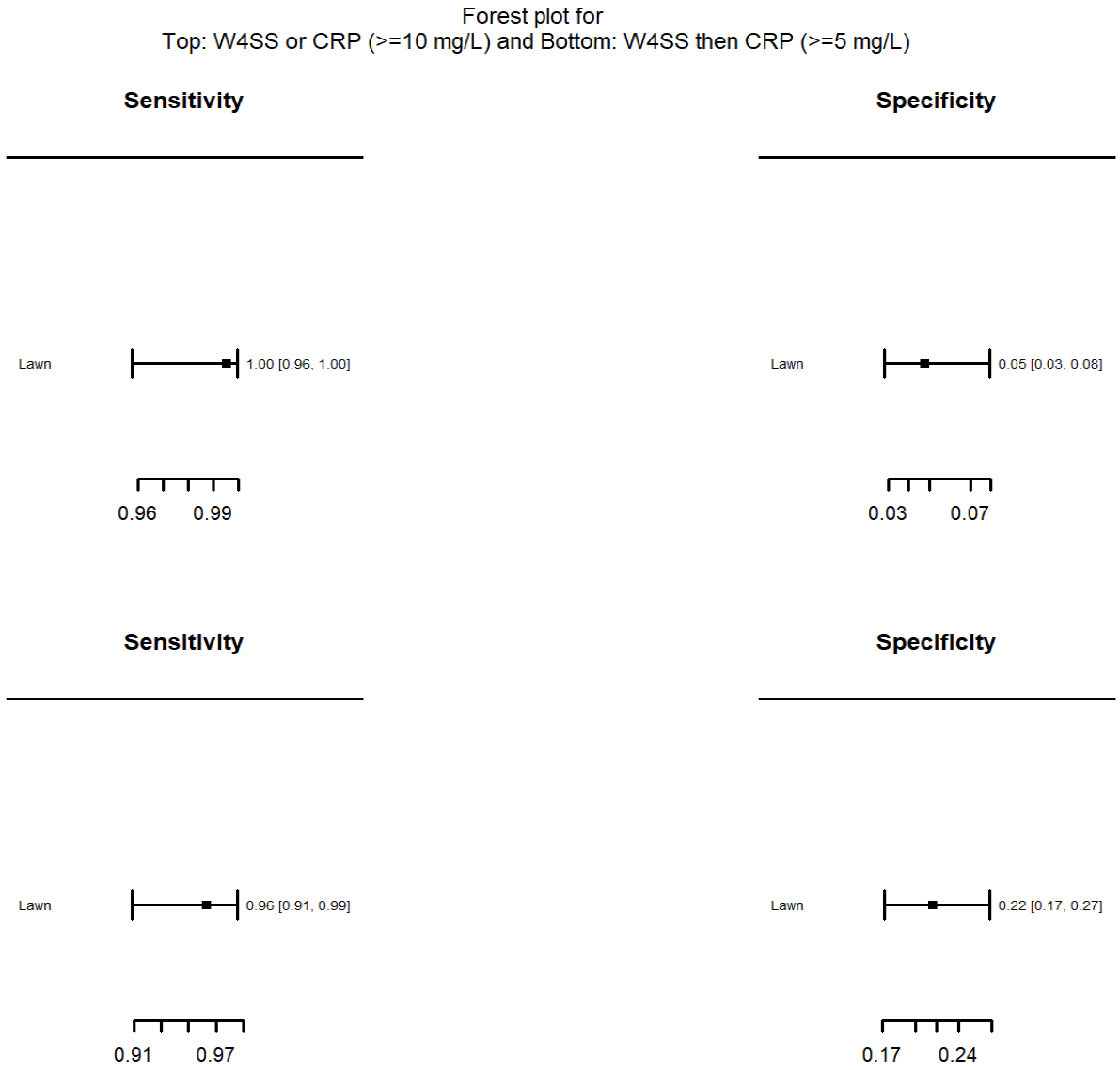
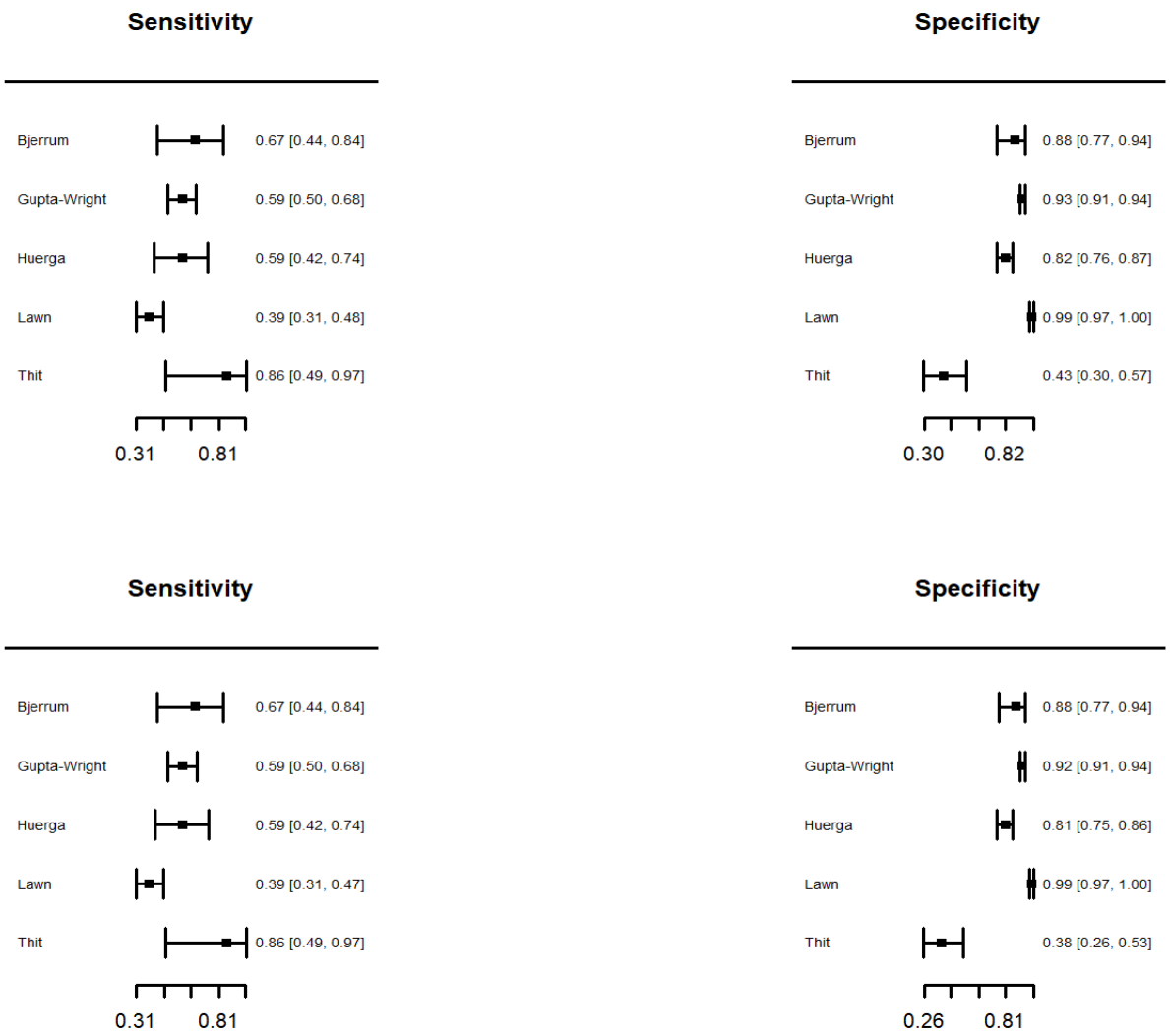

Forest plot for
Top: W4SS or CRP (>=10 mg/L) and Bottom: W4SS then CRP (>=5 mg/L)
Sensitivity
Specificity
Lawn
1.00 [0.96, 1.00]
0.96
0.99
Lawn
0.05 [0.03, 0.08]
0.03
0.07
Sensitivity
Specificity
Lawn
0.96 [0.91, 0.99]
0.91
0.97
Lawn
0.22 [0.17, 0.27]
0.17
0.24

Figure 8-11C - 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

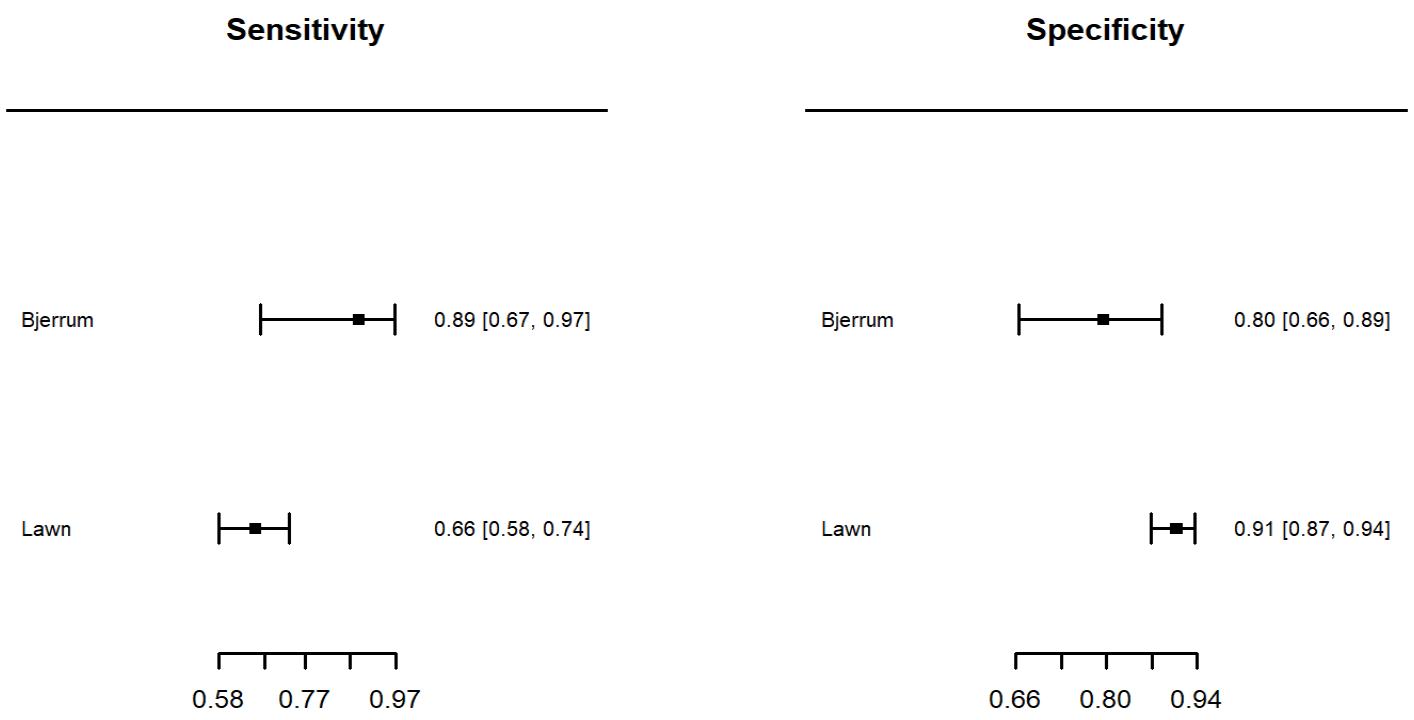
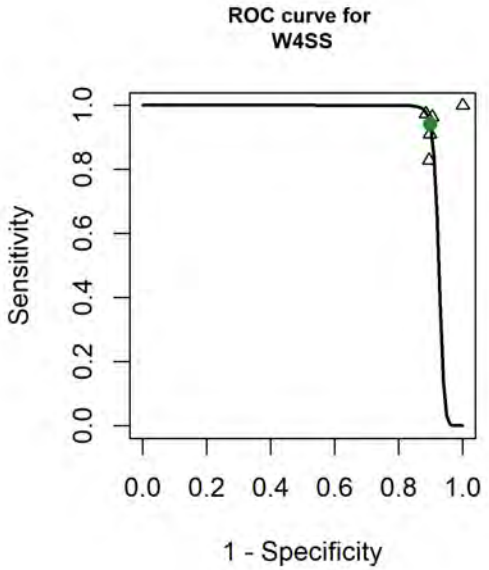
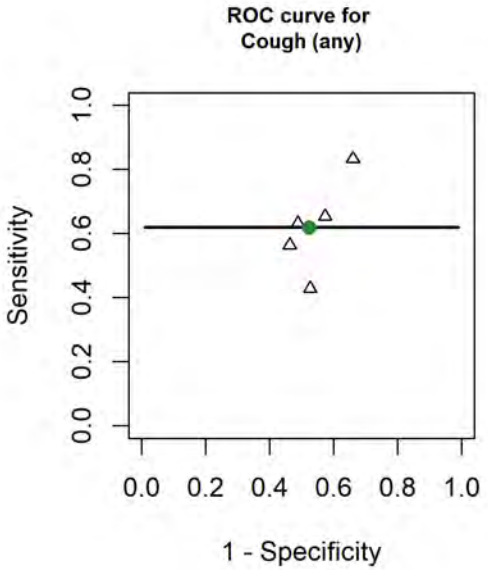
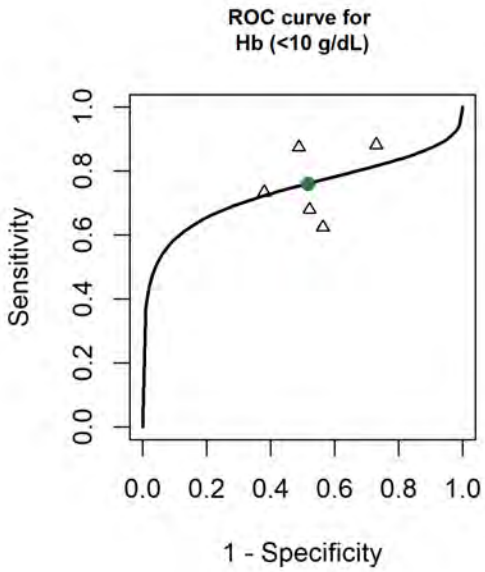


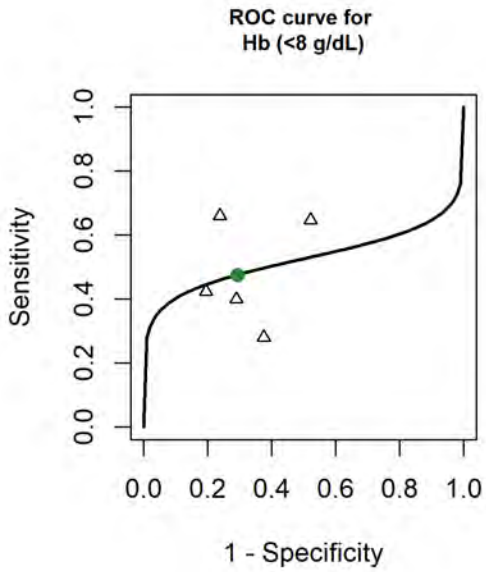
Figure 8-12: Summary receiver operating characteristics curves (for tests/strategies with >=4 studies available)

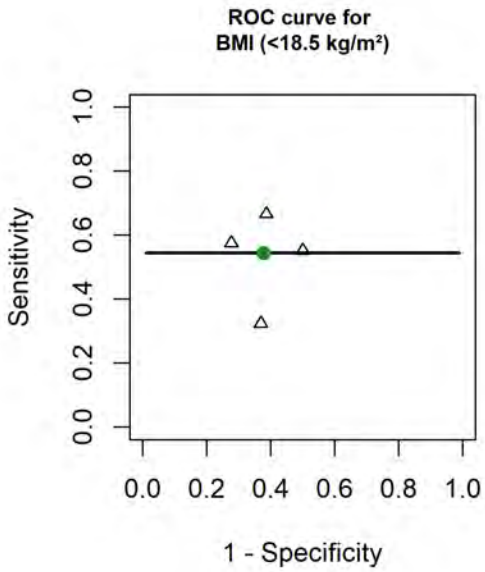
Figure 8-12A - Summary receiver operating characteristics curves for each screening test/strategy using AlereLAM as a reference standard

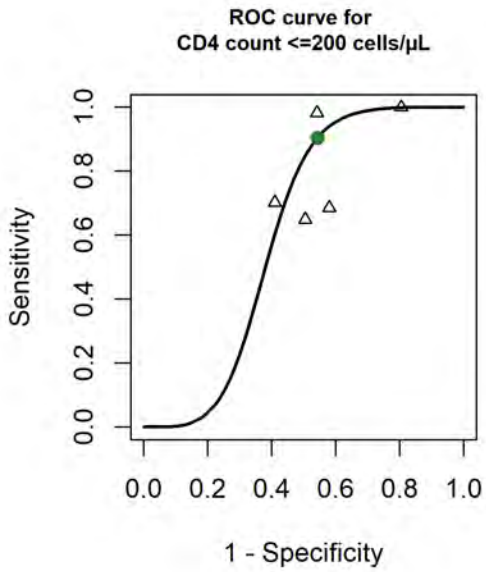





ROC curve for
Hb (<10 g/dL)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0


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


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

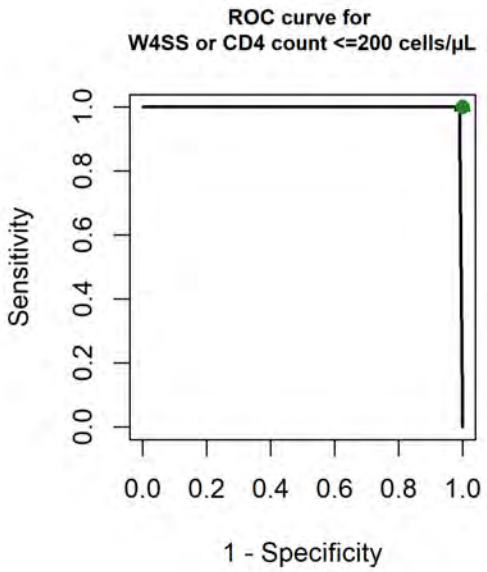
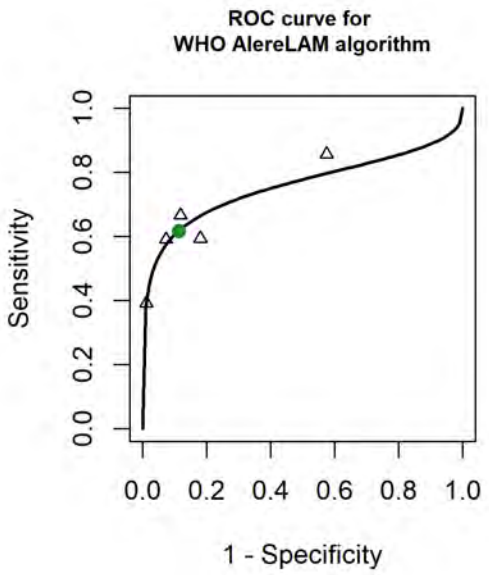

ROC curve for
W4SS or CD4 count <=200 cells/µL
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity

Figure 8-13 - Summary receiver operating characteristics curves for each LF-LAM strategy using culture or Xpert as a reference standard



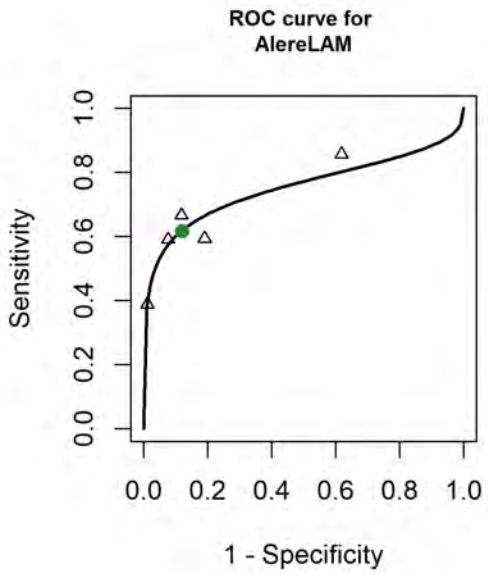
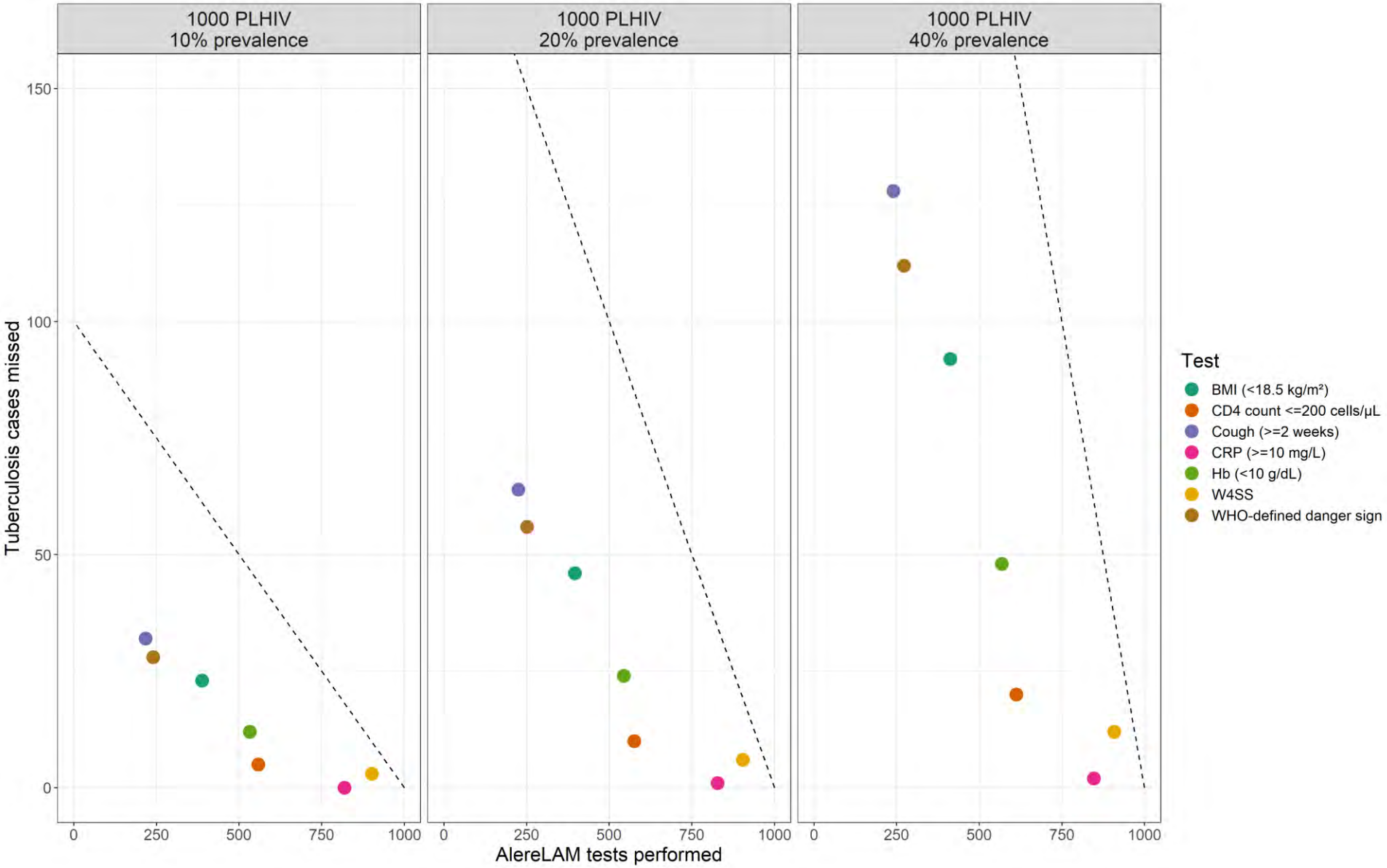


Figure 8-13: Plot comparing number of AlereLAM positive tuberculosis cases missed with number of AlereLAM tests required for different tuberculosis screening tests when screening a population of 1000 persons†



8.4 Appendix for Chapter 5

Supplementary methods

Sample size calculations

To ensure the development of a robust prediction model, sample size calculations are provided for active tuberculosis as a binary outcome to target (B1) a precise overall outcome proportion estimate, (B2) a small mean absolute prediction error, (B3) a shrinkage factor of 0.9, and (B4) small optimism of 0.05 in the apparent R^2. We used the pmsampsize package in R for criteria B1, B3, and B4. We assumed an anticipated value of 0.114 for R^2Cox-Snell. We derived R^2Cox-Snell from a C-statistic of 0.80 for CRP, which we considered the minimum anticipated performance. We chose the overall proportion of tuberculosis as 0.14.Using 58 predictor parameters, at least 4,283 participants are required, corresponding to 600 events (where the prevalence of tuberculosis is 0.14) and events per candidate predictor parameter of 10.3. Based on these criteria, the study sample size (4,315) was considered sufficient.

Missing data

We used the mice package in R to perform single-level multiple imputation within each cohort to deal with missing data.[1] For the imputation model, we included all variables (including transformations and the outcome) that were included in the CPM, as well as auxiliary variables. We first log-transformed CRP (log(C-reactive protein)) and CD4 count (log(CD4 count + 1)) and then scaled all continuous variables by subtracting the mean of the original variable from the raw value and dividing by the standard deviation of the original variable. For imputation methods, we used predictive mean matching for continuous variables, logistic regression for binary variables, polytomous regression for categorical variables and proportional odds model for ordered categorical variables.[2] We compared the distribution of imputations with the distribution of observed values and visually examined diagnostic plots to check for algorithm convergence. We created 10 imputed datasets. All further analyses were performed in each of the 10 imputed datasets. We pooled subsequent model and validation parameters using Rubin's rules.[3]

Table 8-34: Study-level characteristics

Author, year	Country	Study period	Study population	Study setting	Case-finding	Exclusion criteria	Sputum culture	Liquid or solid culture
Boyles, 2020[4]	South Africa	2018-2019	PLHIV with ≥1 W4SS symptom	1 community health clinic	Passive	On ATT or received antibiotics ≤14 days	2 samples, induced if necessary	Liquid
Lawn, 2013[5]	South Africa	2010-2011	ART-naive PLHIV aged ≥18 years	1 community-based ART clinic	Active	No current TB diagnosis	2 spot samples with ≥ 1 induced	Liquid
Reeve, 2019[6]	South Africa	2017-2020	ART-naive PLHIV aged ≥18 years	1 outpatient clinic	Active	On ATT <60 days before or has unknown treatment status	2 spot samples, majority induced	Liquid
Shapiro, 2018[7]	South Africa	2014-2015	ART-naive PLHIV aged ≥18 years	1 urban HIV clinic	Active	-	2 samples, induced if necessary	Liquid
Theron, unpublished	South Africa	2016-2020	PLHIV aged ≥18 years with ≥1 W4SS symptom	2 primary health clinics (Scottsdene and Wallacedene)	Passive	On ATT ≤2 months before or has unknown treatment status	1 sample, induced if necessary	Liquid
Yoon, 2018[8]	Uganda	2013-2016	ART-naive PLHIV aged ≥18 years with CD4 cell count ≤350 per µL	2 HIV clinics in Kampala (MJAP and TASO)	Active	Diagnosis of active tuberculosis, taking ATT (anti-TB or TB preventive therapy, fluoroquinolones) ≤3 days before	2 spot samples, 2nd induced if necessary	Both

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

Table 8-35: Summary of main characteristics for all participants and by each study

Variable	Overall†	Boyles	Lawn	Reeve	Shapiro	Theron_Scottsdene	Theron_Wallacedene	Yoon_MJAP	Yoon_TASO
Participants	4315 (100)	207 (5)	602 (14)	831 (19)	439 (10)	112 (3)	329 (8)	1612 (37)	183 (4)
Demographics									
Active case-finding	3667 (85)	0 (0)	602 (100)	831 (100)	439 (100)	0 (0)	0 (0)	1612 (100)	183 (100)
N	4315	207	602	831	439	112	329	1612	183
Age (years)	33.2 (27-40)	36 (31-41)	33.2 (27.7-40.6)	32 (26-39)	31.7 (26.5-39.2)	37 (32-44)	37 (31-44)	33 (27-40)	33 (26-40)
N	4315	207	602	831	439	112	329	1612	183
Female	2381 (55)	69 (33)	402 (67)	493 (59)	257 (59)	65 (58)	155 (47)	841 (52)	99 (54)
N	4315	207	602	831	439	112	329	1612	183
HIV history									
On ART	380 (9)	80 (39)	0 (0)	0 (0)	0 (0)	55 (49)	245 (74)	0 (0)	0 (0)
N	4315	207	602	831	439	112	329	1612	183
CD4 count (cells/μL)	204 (93-319)	185 (50-363)	168 (96-232)	303 (170-487)	301 (171-466)	330 (150-518)	353 (186-549)	158 (66-260)	146 (68-259)
N	4188	207	599	747	430	108	302	1612	183
Clinical characteristics									
History of tuberculosis	602 (14)	34 (17)	161 (27)	115 (14)	19 (4)	54 (48)	159 (48)	57 (4)	3 (2)
N	4309	203	602	830	439	112	329	1611	183
Cough	2256 (52)	196 (95)	292 (49)	257 (31)	196 (45)	108 (96)	296 (90)	824 (51)	87 (48)
N	4314	207	602	831	439	112	328	1612	183
Fever	1541 (36)	124 (60)	169 (28)	57 (7)	134 (31)	33 (40)	87 (27)	844 (52)	93 (51)
N	4283	207	602	831	439	82	327	1612	183
Weight loss	2638 (62)	186 (90)	406 (68)	343 (41)	149 (34)	60 (72)	162 (49)	1182 (73)	150 (82)
N	4283	207	601	830	439	83	328	1612	183
Night sweats	1674 (39)	155 (75)	242 (40)	196 (24)	142 (32)	90 (80)	217 (66)	560 (35)	72 (39)
N	4313	206	602	831	439	112	328	1612	183
Cough >= 2 weeks	1455 (34)	94 (47)	117 (19)	172 (21)	133 (30)	88 (79)	244 (74)	548 (34)	59 (32)
N	4306	201	602	831	437	112	328	1612	183
Tuberculosis diagnostic tests									
Sputum culture +	652 (15)	65 (32)	93 (17)	93 (12)	42 (10)	42 (38)	81 (25)	208 (13)	28 (15)
N	4209	206	541	796	439	111	322	1611	183
Laboratory tests									
BMI (kg/m2)	22 (19.4-25.8)	20.1 (18.3-22.3)	23.5 (20.8-27.2)	24.2 (21-29.2)	24.2 (21.2-28.9)	19.1 (16.7-22.7)	20.7 (18.4-24.8)	21 (18.9-23.9)	21.2 (19-23.2)
N	4306	207	601	827	439	112	327	1610	183
CRP (mg/L)	6.4 (2.5-38.6)	57 (16.8-115)	9.9 (2.5-32.5)	6.2 (2.5-29.2)	4.1 (2.5-17.3)	63.1 (11.3-124.9)	41.4 (6.3-127.7)	4.1 (2.5-22.2)	3.5 (2.5-31)
N	4093	204	502	763	425	107	299	1610	183
Hb (g/dL)	12.5 (11-13.9)	-	11.9 (10.4-13.2)	12.7 (11.3-13.9)	-	12.6 (10.8-13.7)	13.1 (11.6-14.3)	12.7 (11-14.3)	11.7 (10.5-13.4)
N	2453	-	565	665	-	110	324	736	53

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

‡Participants with data available for variable

Variable	Overall†	Boyles	Lawn	Reeve	Shapiro	Theron_Scottsdene	Theron_Wallacedene	Yoon_MJAP	Yoon_TASO

Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein, Hb = haemoglobin

Table 8-36: Percentage of missing data by each study

Variable†	Boyles	Lawn	Reeve	Shapiro	Theron_Scottsdene	Theron_Wallacedene	Yoon_MJAP	Yoon_TASO
Case finding	0	0	0	0	0	0	0	0
Age	0	0	0	0	0	0	0	0
Sex	0	0	0	0	0	0	0	0
ART status	0	0	0	0	0	0	0	0
History of tuberculosis	2	0	0	0	0	0	0	0
Cough	0	0	0	0	0	0	0	0
Fever	0	0	0	0	27	1	0	0
Weight loss	0	0	0	0	26	0	0	0
Night sweats	0	0	0	0	0	0	0	0
Cough >=2 weeks	3	0	0	0	0	0	0	0
BMI	0	0	0	0	0	1	0	0
CD4 count	0	0	10	2	4	8	0	0
CRP	1	17	8	3	4	9	0	0
Haemoglobin	100	6	20	100	2	2	54	71
Sputum culture	0	10	4	0	1	2	0	0

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

Definition of abbreviations: ART = antiretroviral therapy, BMI = body mass index, CRP = C-reactive protein

Table 8-37: Backward stepwise selection using Akaike information criterion in bootstrap samples for each imputed dataset

Multiply imputed dataset	Age	Body mass index	CD4 count	C-reactive protein	Sex	Cough	Fever	Weight loss	Night sweats	Case finding
1	65	100	98	100	30	69	28	56	51	41
2	76	100	99	100	45	64	19	57	44	51
3	56	100	100	100	62	65	30	32	57	47
4	79	100	99	100	60	75	22	46	53	39
5	65	100	100	100	56	72	23	61	50	55
6	69	100	99	100	49	66	17	42	67	47
7	80	100	99	100	67	56	17	50	67	43
8	74	100	99	100	48	76	18	68	54	48
9	71	100	99	100	54	74	22	58	57	40
10	80	100	96	100	54	79	31	37	46	37

Variable selected in at least 70% (green) or less than 70% (red) of bootstrap samples for each of 10 imputed datasets

Table 8-38: Extended CPM with variables retained in >= 70% of bootstrap samples and in >= 50% of imputed datasets

Term	Estimate
Intercept	-1.7209
Age (years)	0.6511
Age (spline 1)	-2.0033
Age (spline 2)	5.0965
Body mass index (kg/m2)	-0.8335
Body mass index (spline 1)	2.2347
Body mass index (spline 2)	-4.8940
CD4 count (cells/μL)	0.1726
CD4 count (spline 1)	-0.5500
CD4 count (spline 2)	2.4367
C-reactive protein (mg/L)	1.3609
C-reactive protein (spline 1)	5.4569
C-reactive protein (spline 2)	-6.7330
Cough (yes)	0.2983

To recreate the above model, raw data values need to undergo the following preperation steps.
First, log transformed C-reactive protein (log(C-reactive protein)) and CD4 count (log(CD4 count + 1)) need to be calculated.
Second, all continous variables need to be scaled by subtracting the mean of the original variable from raw value and then dividing it by standard deviation of the original variable. The means and standard deviations are: Age (34.46; 9.20), Body mass index (23.26; 5.63), CD4 count (5.07; 1.08), C-reactive protein (2.38; 1.52).
Third, restricted cubic spline transformations need to be generated using 4 knots. The knot positions are: Age (-1.42; -0.48; 0.28; 1.80), Body mass index (-1.19; -0.50; 0.14; 1.93), CD4 count (-2.02; -0.12; 0.51; 1.34), C-reactive protein (-0.96; -0.85; 0.38; 2.01).

Table 8-39: C-reactive protein only CPM

Term	Estimate
Intercept	-2.5665
C-reactive protein (mg/L)	1.1262
C-reactive protein (spline 1)	18.8435
C-reactive protein (spline 2)	-21.6480

To recreate the above model, raw data values need to undergo the following preperation steps.
First, log transformed C-reactive protein (log(C-reactive protein)) needs to be calculated.
Second, C-reactive protein needs to be scaled by subtracting the mean of the original variable from raw value and then dividing it by standard deviation of the original variable. The means and standard deviations for C-reactive protein are 2.38 and 1.52.
Third, restricted cubic spline transformations need to be generated using 4 knots. The knot positions for C-reactive protein are -0.96; -0.85; 0.38; 2.01.

Table 8-40: Joint predicted probability of achieving a C-statistic of > 0.70, 0.75, or 0.80 and a CS between 0.8 and 1.2 in a new population

Model	0.70	0.75	0.80
C-reactive protein only CPM	0.814	0.695	0.382
Extended CPM	0.767	0.761	0.537
Hanifa model	0.273	0.003	0.000

Table 8-41: Model performance statistics using AregImpute

Model or tool	Concordance statistic	Calibration-in-the-large	Calibration slope
Extended CPM			
Summary	0.81 (0.77, 0.84)	0.04 (-0.18, 0.25)	0.97 (0.81, 1.13)

Figure 8-14: Plots of predictor effects (plot.Predict)

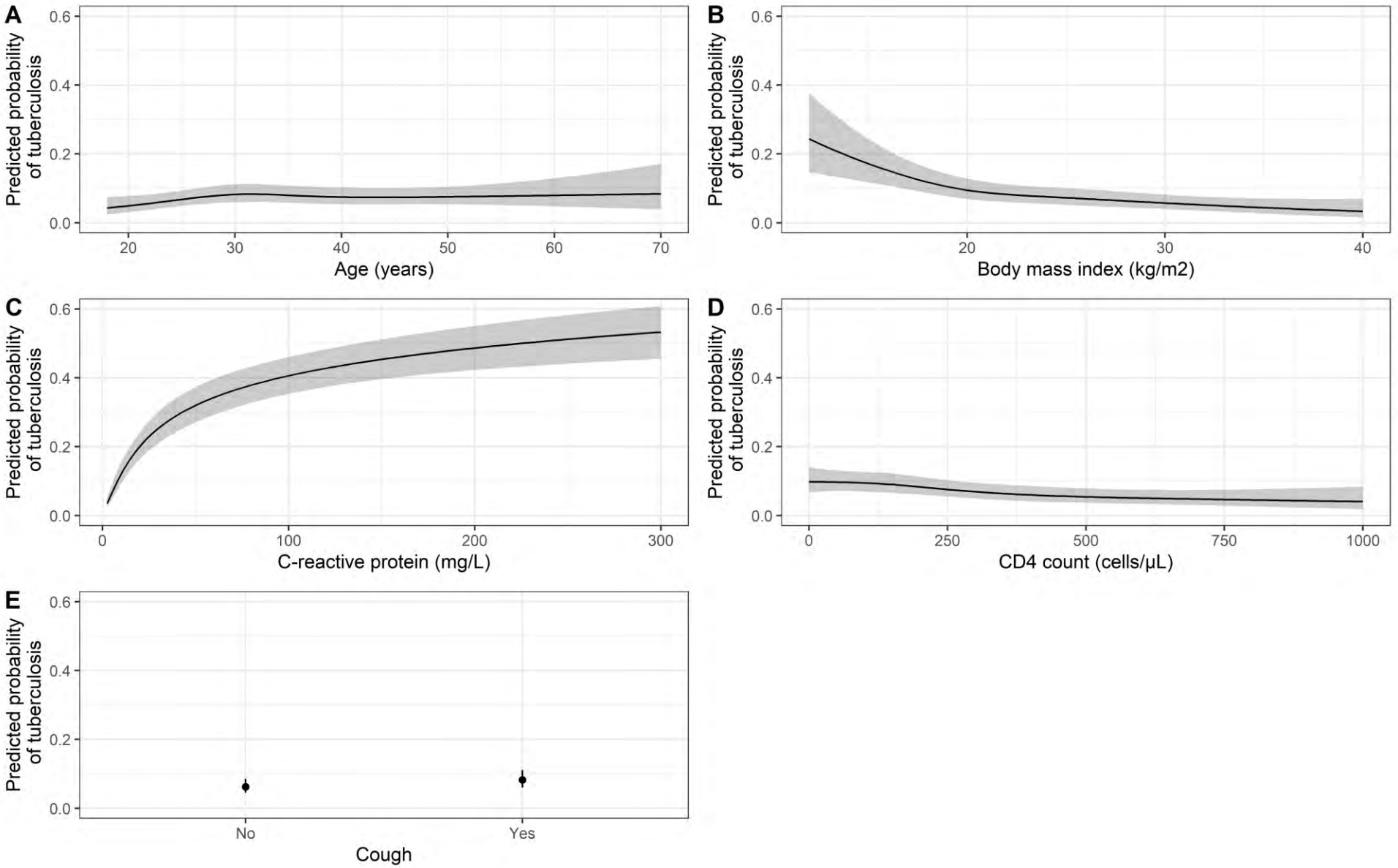
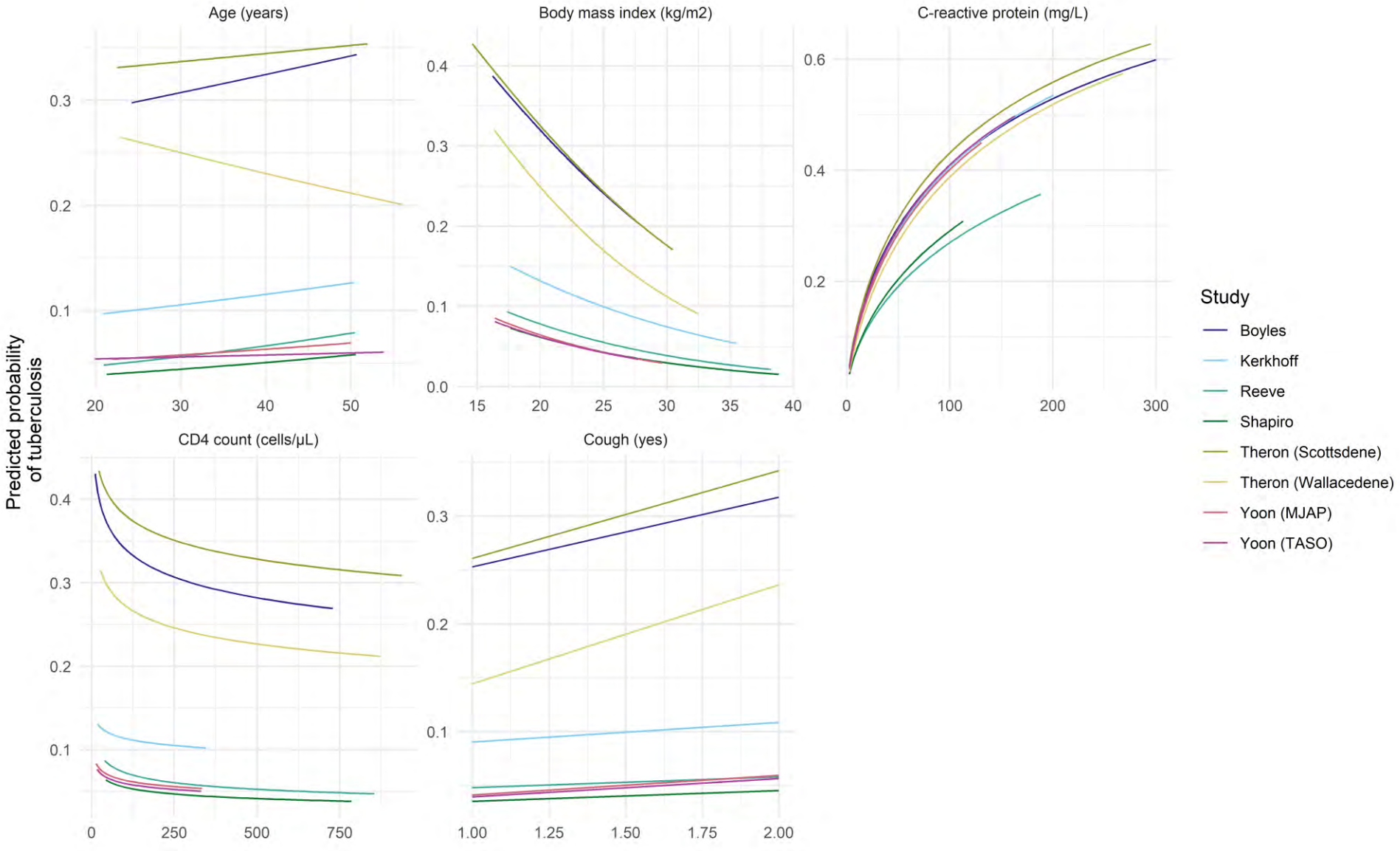
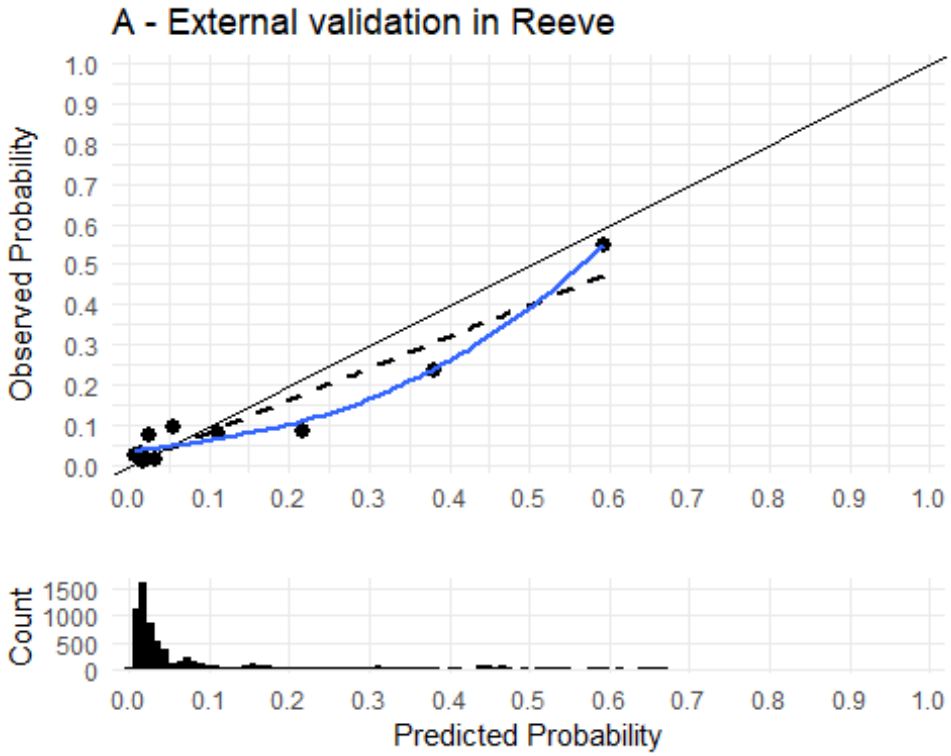


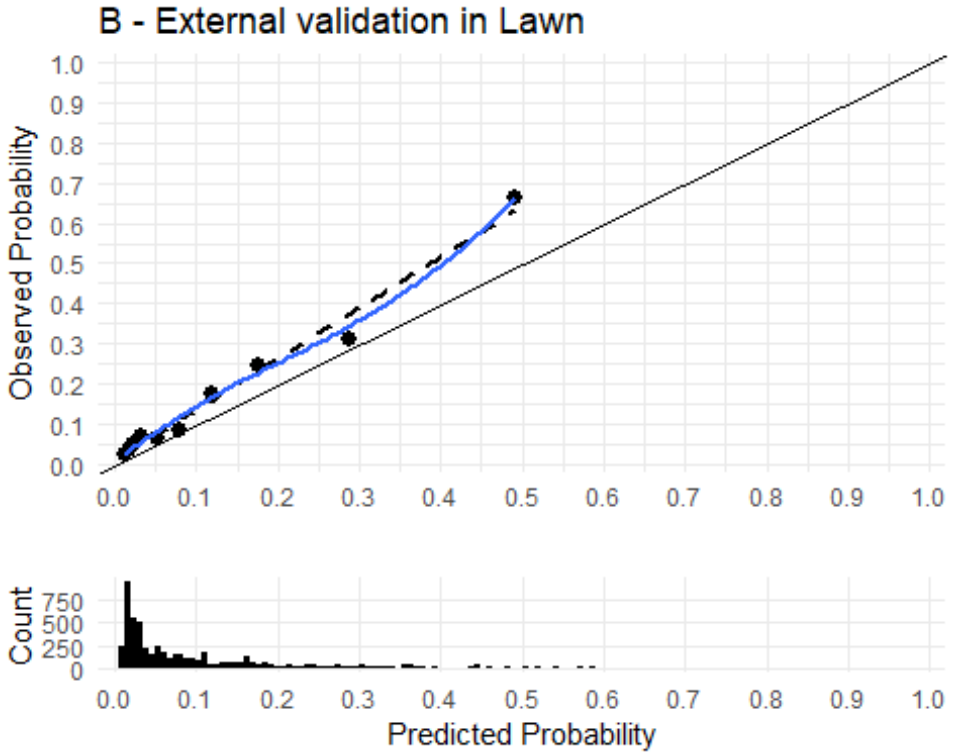
Figure 8-15: Heterogeneity of predictor effects using linear mixed-effects models fit across all datasets*

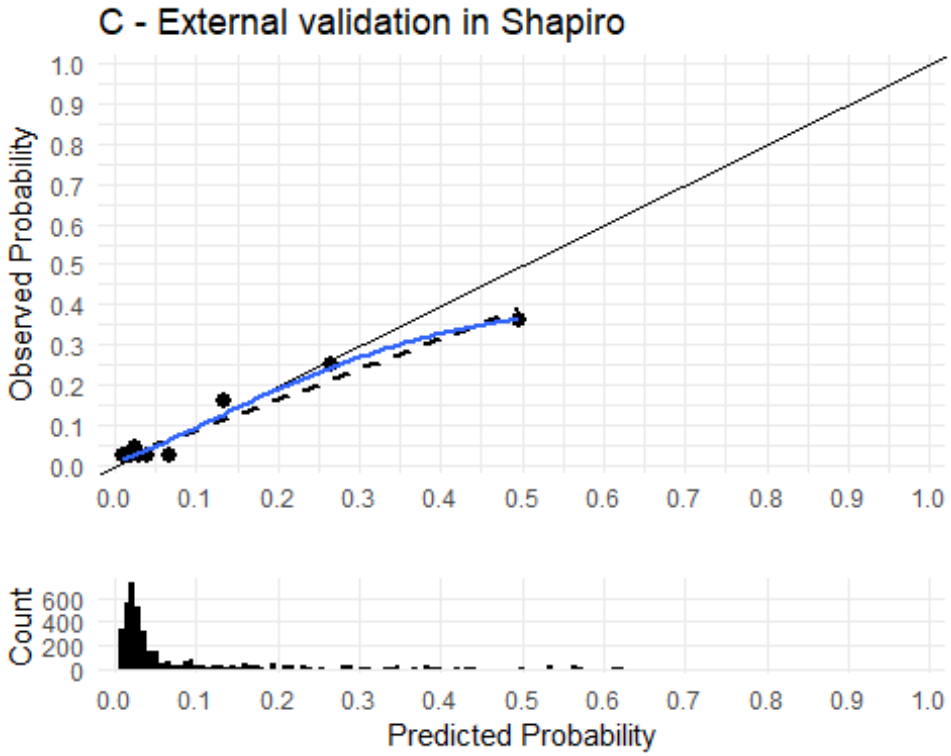


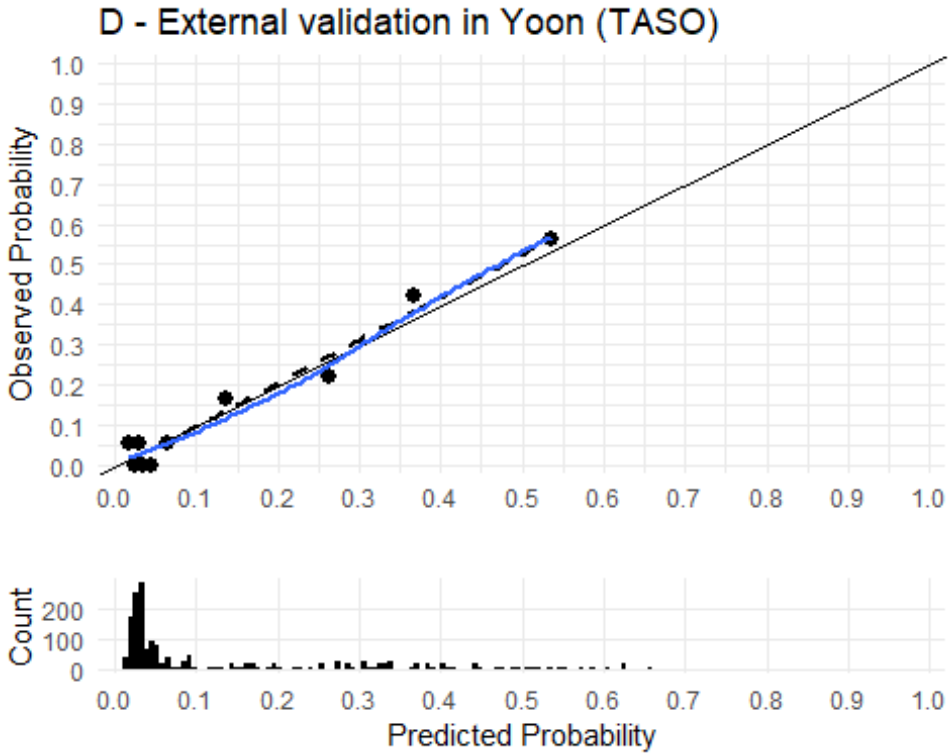
*Across 90% quantile range for each covariate in dataset with other covariates held at median (or mode) value

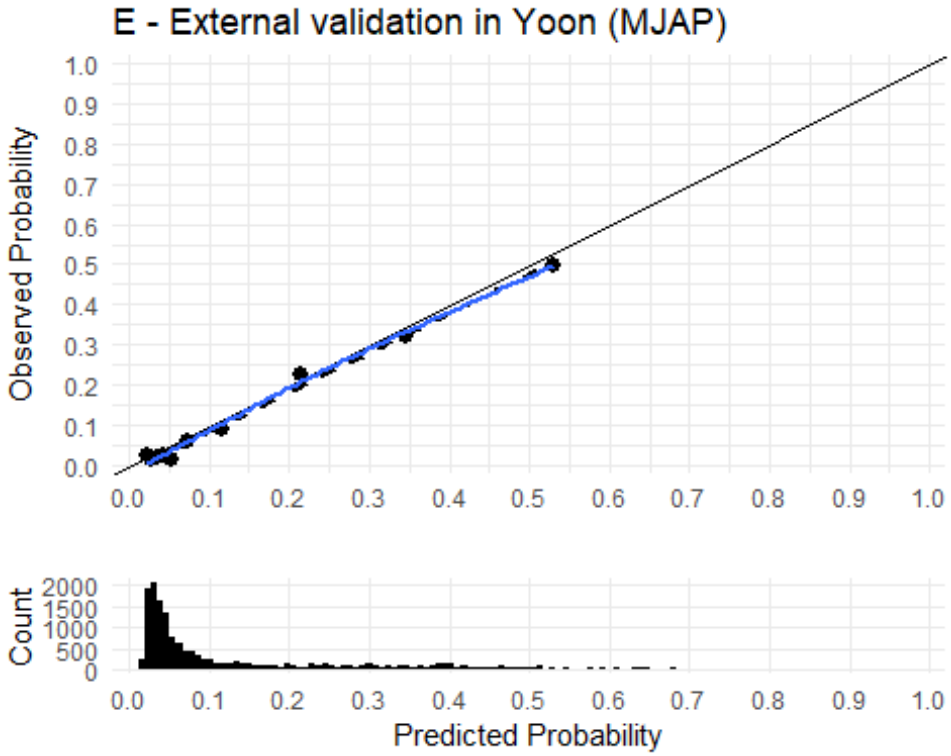
Figure 8-16: Extended CPM calibration plots

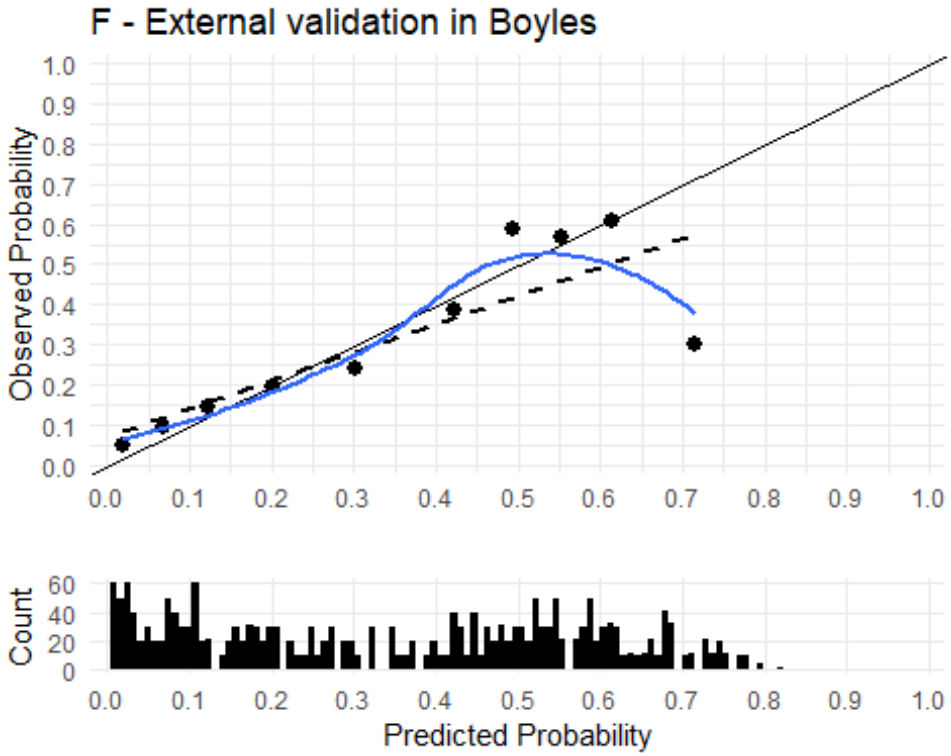

A - External validation in Reeve
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
1500
1000
500
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability

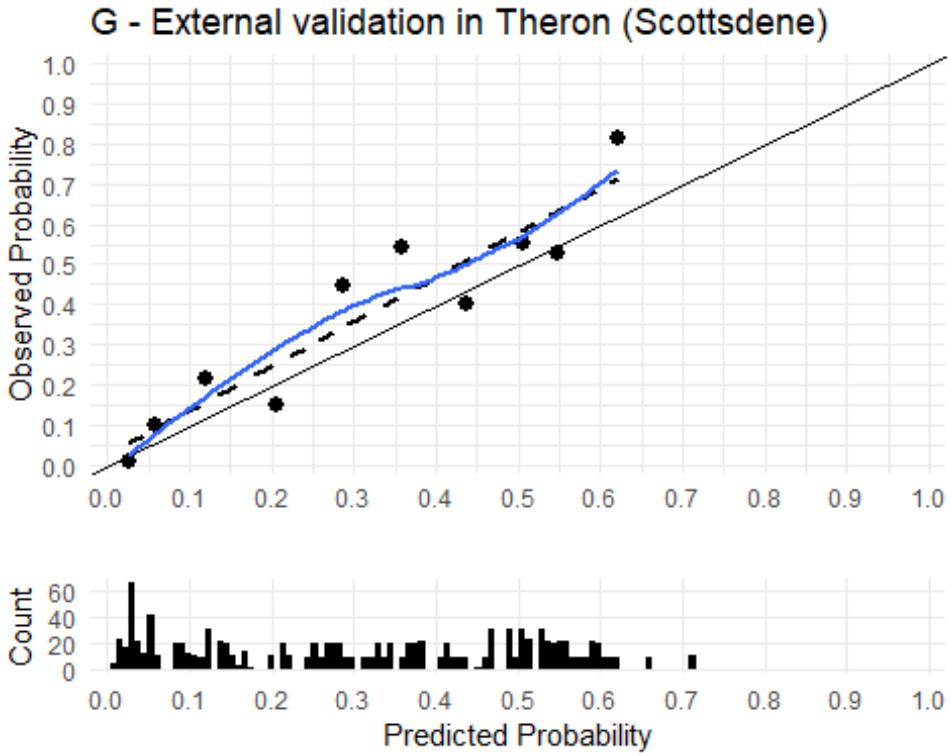

B - External validation in Lawn
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
750
500
250
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


C - External validation in Shapiro
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
600
400
200
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


D - External validation in Yoon (TASO)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
200
100
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


E - External validation in Yoon (MJAP)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
2000
1500
1000
500
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


F - External validation in Boyles
Observed Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Count
0
20
40
60
Predicted Probability


G - External validation in Theron (Scottsdene)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
60
40
20
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability

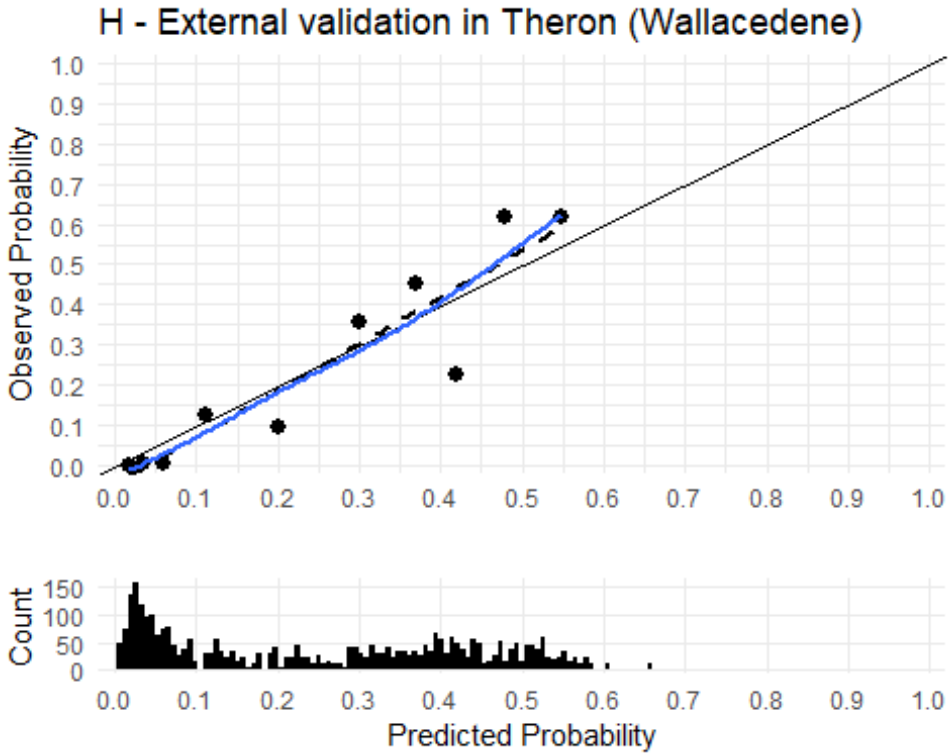
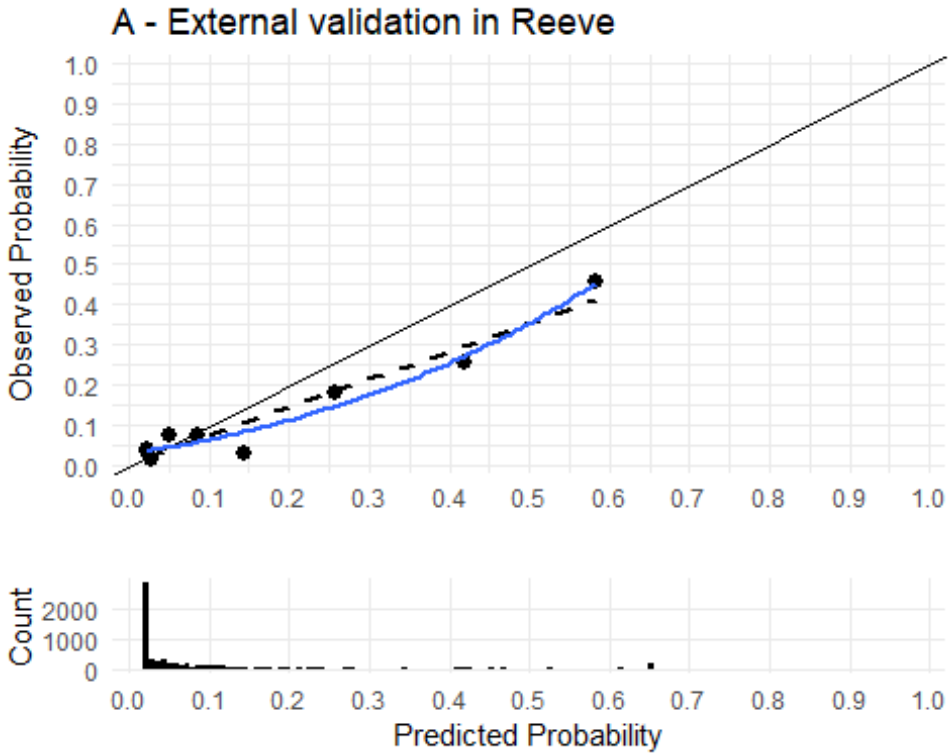
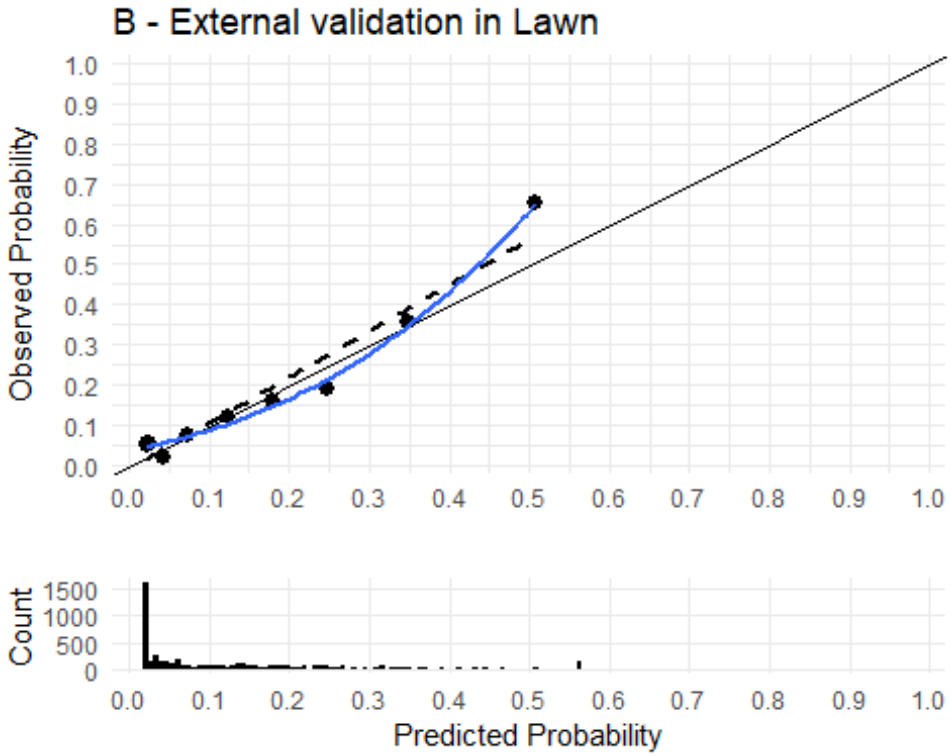
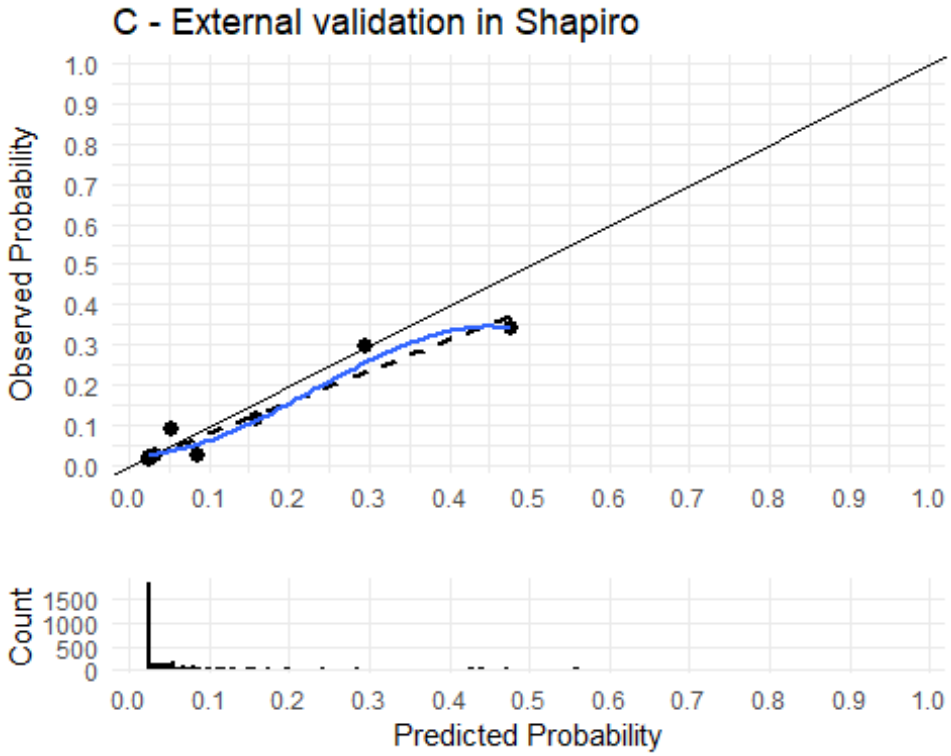
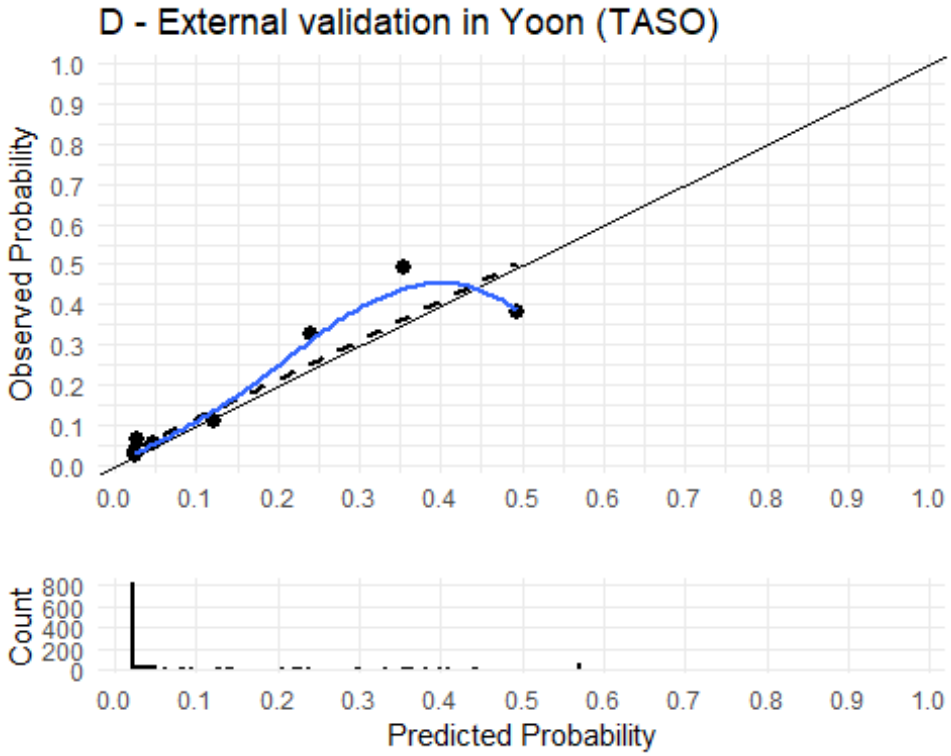

H - External validation in Theron (Wallacedene)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
150
100
50
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability

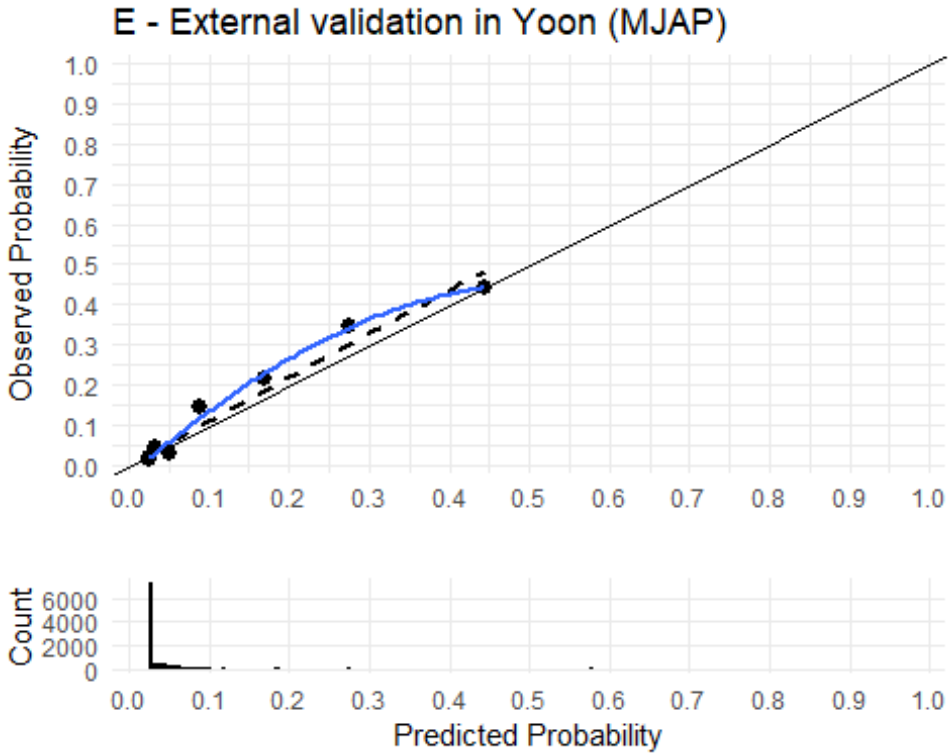
Figure 8-17: C-reactive protein only CPM calibration plots

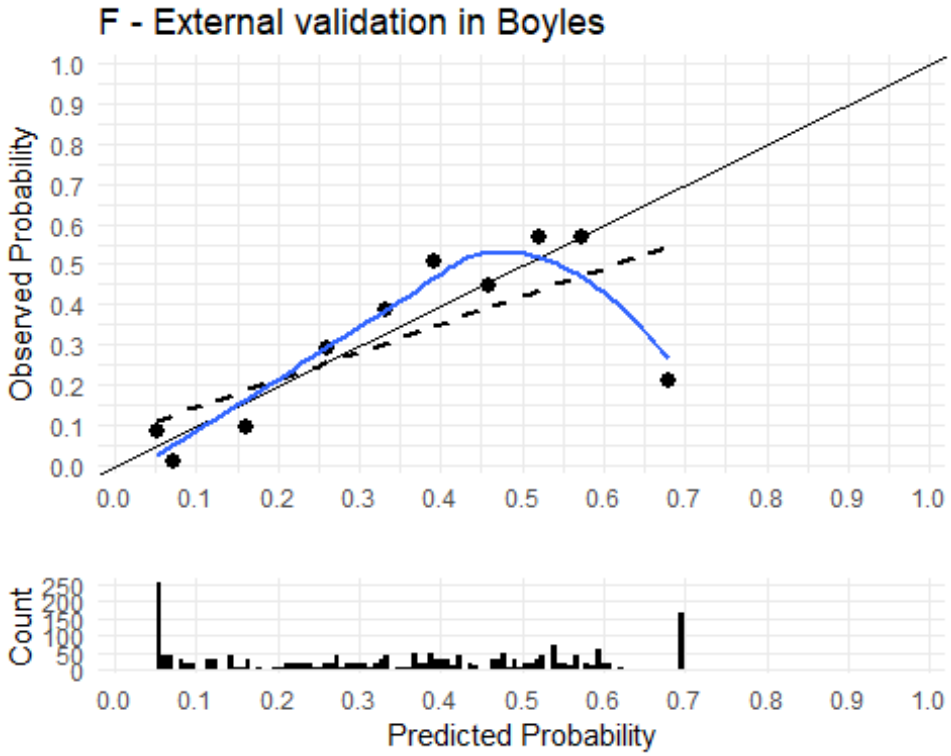

A - External validation in Reeve
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
2000
1000
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability

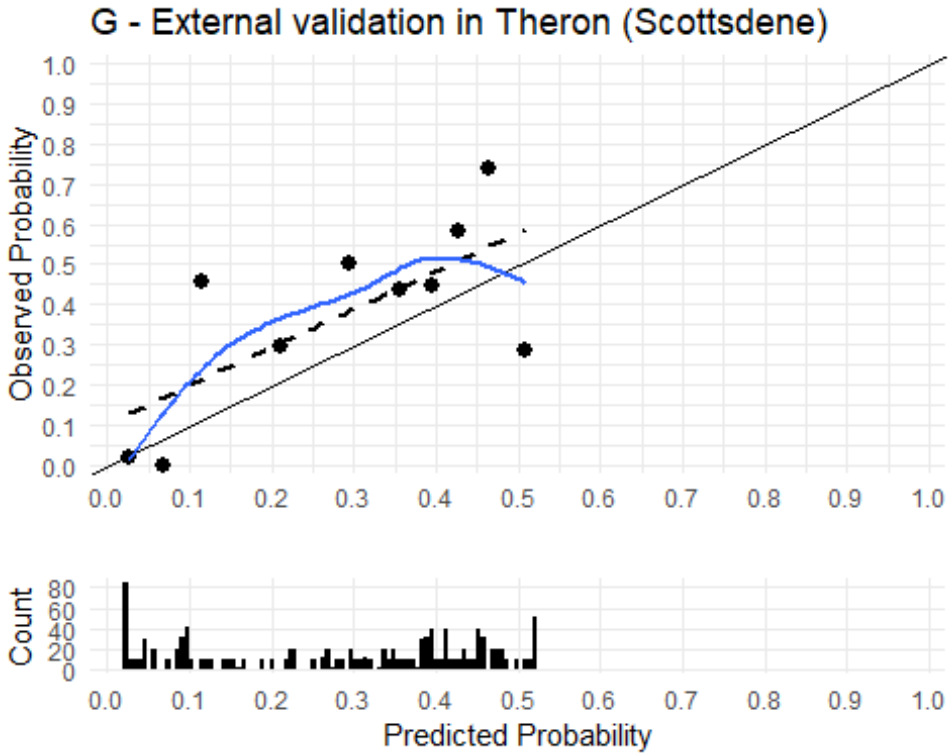

B - External validation in Lawn
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
1500
1000
500
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


C - External validation in Shapiro
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
1500
1000
500
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


D - External validation in Yoon (TASO)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
800
600
400
200
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


E - External validation in Yoon (MJAP)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
6000
4000
2000
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability


F - External validation in Boyles
Observed Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Count
0
50
100
150
200
250
Predicted Probability


G - External validation in Theron (Scottsdene)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
80
60
40
20
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability

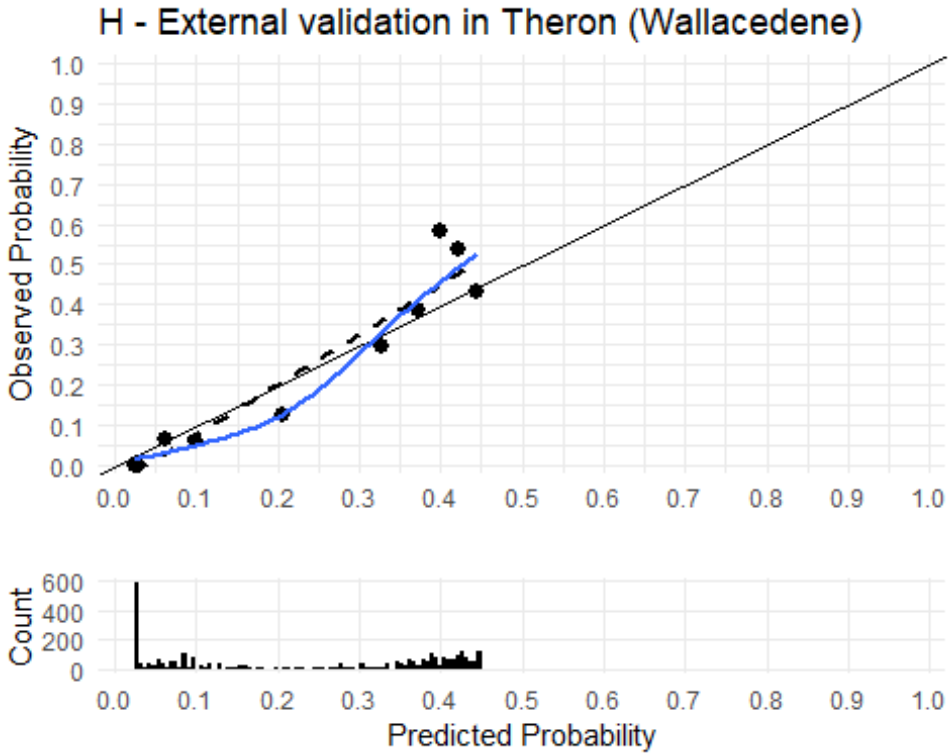
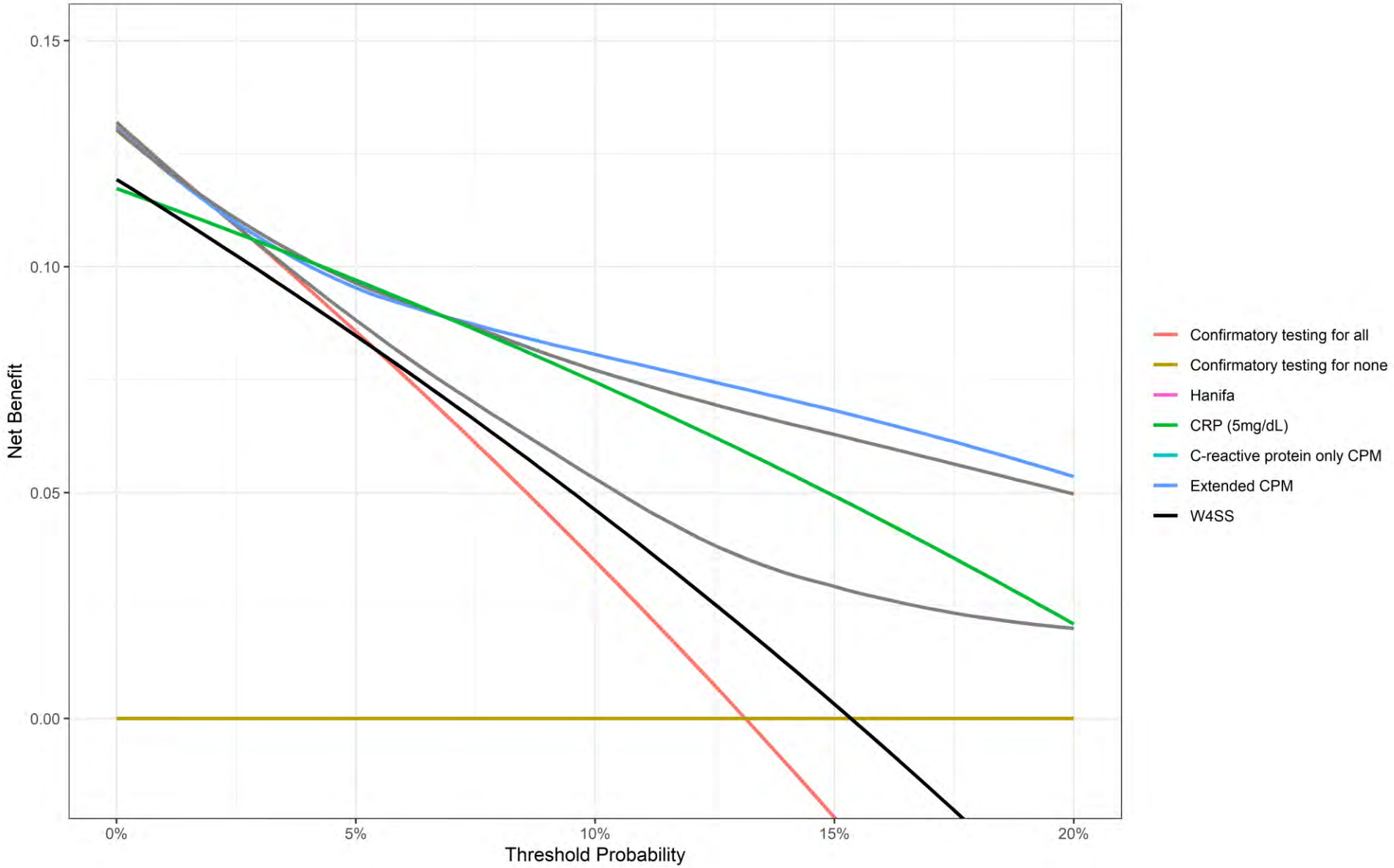

H - External validation in Theron (Wallacedene)
Observed Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Count
600
400
200
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability

Figure 8-18: Decision curve analysis comparing the final and C-reactive protein only CPMs to other tools or strategies among active case-finding cohorts*



*Excludes 3 passive case-finding cohorts, as all participants in that study had a positive W4SS
Definition of abbreviations: CRP = C-reactive protein

Figure 8-19: Plots of predictor effects (plot.Predict) using AregImpute