Point-of-care C-reactive protein and Xpert MTB/RIF Ultra for tuberculosis screening and diagnosis in unselected antiretroviral therapy initiators: a prospective diagnostic accuracy study

Byron WP Reeve (PhD)*[1], Gcobisa Ndlangalavu*[1], Hridesh Mishra[1,5] (PhD), Zaida Palmer[1], Happy Tshivhula[1] (PhD), Loren Rockman[1], Selisha Naidoo[1], Desiree L Mbu[1], Charissa C Naidoo[1] (PhD), Brigitta Derendinger[1] (PhD), Gerhard Walzl[1] (MD, PhD), Stephanus T. Malherbe[1] (MD, PhD), Paul D van Helden[1] (PhD), Fred C Semitala[2,3] (MD), Christina Yoon[2,4] (MD), Rishi K Gupta (MRCP)[6], Mahdad Noursadeghi[7] (PhD), Robin M Warren[1] (PhD), Grant Theron[1†](PhD)

Affiliations:

[1]DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.

[2]Makerere University Joint AIDS Program (MJAP), Kampala, Uganda

[3]Department of Internal Medicine, Makerere University College of Health Sciences, Makerere, Kampala, Uganda

[4]Department of Medicine, Division of Pulmonary and Critical Care Medicine, Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, USA.

[5]Public Health Research Institute, New Jersey Medical School, Rutgers, The State University of New Jersey, Newark, NJ, USA

[6]Division of Infection and Immunity, University College London, London, UK

[7]Institute for Health Informatics, University College London, London, UK

[*]Equally contributing joint first authors

[†]Corresponding author: Prof Grant Theron, 2035 BMRI, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Francie Van Zijl Drive, Cape Town, South Africa, 7550. E-mail: gtheron@sun.ac.za. Tel.: +27 21 938 9693.

Keywords: tuberculosis, HIV, screening, diagnosis

Abstract

Background:

Tuberculosis (TB), a major cause of death in people living with HIV (PLHIV), remains challenging to diagnose. Diagnostic accuracy data are lacking for promising triage tests, such as C-reactive protein (CRP), and confirmatory tests, such as sputum and urine Xpert MTB/RIF Ultra (Ultra), and urine LAM, without prior symptom selection.

Methods:

897 PLHIV initiating antiretroviral therapy were consecutively recruited in settings with high TB incidence, irrespective of symptoms. Participants were offered sputum induction, with a liquid culture reference standard. First, we evaluated point-of-care CRP testing on blood, compared to the World Health Organization (WHO)-recommended four-symptom screen (W4SS) for triage (n=800). Second, we evaluated Xpert MTB/RIF Ultra (Ultra) versus Xpert MTB/RIF (Xpert) for sputum-based confirmatory testing (n=787), with or without sputum induction. Third, we evaluated Ultra and Determine LF-LAM for urine-based confirmatory testing (n=732).

Findings:

CRP and number of W4SS symptoms had areas under the receiver operator characteristic curve of 0.78 (95% confidence interval 0.73, 0.83) and 0.70 (0.64, 0.75), respectively. For triage, CRP (≥10 mg/l) has similar sensitivity to W4SS [77% (68, 85) vs. 77% (68, 85); p>0.999] but higher specificity [64% (61, 68) vs. 48% (45, 52); p<0.001]; reducing unnecessary confirmatory testing by 138 per 1000 people and the number-needed-to-test from 6.91 (6.25, 7.81) to 4.87 (4.41, 5.51). Using sputum, which required induction in 31% (24, 39) of people, Ultra had higher sensitivity than Xpert [71% (61, 80) vs. 56% (46, 66); p<0.001] but lower specificity [98% (96, 100) vs. 99% (98, 100); p<0.001]. The proportion of people with ≥1 positive confirmatory result detected by Ultra increased from 45% (26, 64) to 66% (46, 82)

when induction was done. Programmatically-done haemoglobin, triage test combinations, and urine tests showed comparatively worse performance.

Interpretation:

Among ART-initiators in a high burden setting, CRP is a more specific triage test than W4SS. Sputum induction improves yield. Sputum Ultra is a more accurate confirmatory test than Xpert.

Funding: SAMRC (MRC-RFA-IFSP-01-2013), EDCTP2 (SF1401, OPTIMAL DIAGNOSIS), NIH/NIAD (U01AI152087).

Research in context

Evidence before this study

Novel triage and confirmatory tests are urgently needed for TB, especially in key risk groups like PLHIV. Many TB cases do not meet World Health Organization (WHO)-recommended four-symptom screen (W4SS) criteria despite accounting for significant transmission and morbidity. W4SS also lacks specificity, which makes onward referral of triage-positive people for expensive confirmatory testing inefficient and hampers diagnostic scale-up. Alternative triage approaches like CRP have promise, but have comparatively little data in ART-initiators, especially when done without syndromic preselection and using point-of-care (POC) tools. After triage, confirmatory testing can be challenging due to sputum scarcity and paucibacillary early-stage disease. Next generation WHO-endorsed rapid molecular tests (including Xpert MTB/RIF Ultra; Ultra) are a standard-of-care for confirmatory testing. However, there are no supporting data in ART-initiators, among whom Ultra may offer large sensitivity gains over predecessors like Xpert MTB/RIF (Xpert). The added value of sputum induction to augment diagnostic sampling for confirmatory testing is also unclear. Lastly, the performance of urine tests (Ultra, Determine LF-LAM) in this population requires more data.

Added value of this study

We evaluated repurposed and new tests for triage and confirmatory testing using a rigorous microbiological reference standard in a highly vulnerable high-priority patient population (ART-initiators) regardless of symptoms and ability to naturally expectorate sputum. We showed POC CRP triage is feasible, performs better than W4SS, and that combinations of different triage approaches offer no advantages over CRP alone. Sputum Ultra has superior sensitivity to Xpert; often detecting W4SS-negative TB. Furthermore, without induction, confirmatory sputum-based testing would not be possible in a third of people. Urine tests had poor performance. This study contributed unpublished data to systematic reviews and meta-

analyses used by the WHO to inform global policy supporting use of CRP triage and Ultra in PLHIV.

Implication of all the available evidence

POC CRP triage testing is feasible and superior to W4SS and, together with sputum induction in people who triage CRP-positive should, after appropriate cost and implementation research, be considered for roll-out in ART-initiators in high burden settings. Such people should be offered Ultra, which outperforms Xpert.

Introduction

Tuberculosis (TB) is the single biggest cause of death among people living with HIV (PLHIV). Such individuals are most immunocompromised prior to antiretroviral therapy (ART) initiation and, at HIV diagnosis, are already captured within HIV care cascades. ART initiation clinics represent a valuable opportunity to rapidly screen and test for TB. However, the point-of-care (POC) feasibility, accuracy, and potential utility of new and repurposed tools is unclear, especially in the absence of symptomatic pre-selection[1].

The World Health Organization (WHO)-recommended four-symptom screen (W4SS) is used to identify people requiring confirmatory TB testing. However, in addition to being subjective, stigmatizing and poorly implemented[2,3], prevalence surveys show that most bacteriologically positive TB is in W4SS-negative individuals[4]. Consequently, the WHO emphasises that, as an alternative or adjunct to symptoms, an inexpensive and rapid triage test with ≥90% sensitivity and ≥70% specificity on accessible non-sputum specimens is needed to better focus relatively expensive confirmatory testing[5].

C-reactive protein (CRP) is a blood biomarker with evidence to assist in TB diagnosis in PLHIV[6-11]. CRP platforms are commercially-available[12], however, few are evaluated at POC[13] or alongside other biomarkers like haemoglobin[14]. For the implementation of CRP triage, POC testing is required to enable real-time decision making within the same encounter that symptoms are ascertained. A systematic review and individual participant data meta-analysis, which focussed on aggregate data in PLHIV but not specifically ART-initiators, W4SS-negatives, or POC feasibility, showed CRP to have similar sensitivity vs. W4SS but higher specificity[15]. This meta-analysis informed a WHO guideline update, where CRP was recommended as an adjunct to W4SS triage for HIV-associated TB[16].

CRP performance data in concert with the new Xpert MTB/RIF Ultra (Ultra) sputum test are also limited. Ultra has improved limit of detection for *Mycobacterium tuberculosis* complex (*Mtb*) versus its predecessor Xpert MTB/RIF (Xpert). Critically, this sensitivity increment is likely largest in people with paucibacillary early-stage disease[17]. Such individuals may not self-present to care due to TB symptoms (if any) but still do so, for example, to initiate ART. Ultra at this point could therefore represent a hitherto unavailable opportunity to detect TB early in a key risk group, however, there are limited data.

Importantly, individuals with early-stage TB disease and/or HIV may not be able to naturally expectorate sputum[18]. Sputum-scarce individuals are often excluded from rapid confirmatory testing – both programmatically and in research studies – because sputum induction is frequently unavailable in high burden primary care settings. Diagnostic reference standards are also challenging in people from whom a sputum was not retrieved and sputum-scarce people hence represent a source of missed cases not included in most sensitivity calculations. Furthermore, the proportion of people who would have needed induction is seldom reported, meaning that studies that include only sputum expectorators overestimate real-world test performance. Lastly, due to these and other challenges associated with sputum[19,] it is important to evaluate confirmatory tests on easily accessible fluids like urine.

We therefore sought to evaluate the performance of novel triage and confirmatory testing approaches among PLHIV who were unselected based on symptoms and visiting a facility in Cape Town, South Africa to start ART. For triage, we compared POC CRP to W4SS as the current standard-of-care. For confirmatory testing, we evaluated sputum tests (Xpert, Ultra), the incremental diagnostic yield of sputum induction, and urine tests (Ultra, LF-LAM).

Methods

Study population and flow

897 adults (≥18 years) newly diagnosed with HIV and referred to Kraaifontein Community Health Centre to start ART were prospectively enrolled (15 March 2017 - 17 March 2020) without symptom criteria and evaluated at a single visit (**Figure 1**). Participants were excluded if they had received any TB treatment within the last two months, were of unknown treatment status, or did not consent.

Ethics

This study was approved by the Stellenbosch University Faculty of Health Sciences Research Ethics Committee (N14/10/136) and the Western Cape Department of Health, South Africa (WC_2016RP38_944), and is on Clinicaltrials.gov (NCT03187964).

W4SS, CRP, and haemoglobin

People required at least one W4SS symptom (current cough, fever, night sweats or weight loss) to be positive[16]. Four-to-five drops of capillary blood were used by a health worker for POC CRP measurement (iChromaII platform, Boditech, South Korea) (**Supplement)**. CD4 counts were measured outside the study but captured (counts more than three months before or after recruitment were excluded) as well as haemoglobin (when medically indicated by routine staff per local guidelines[20]). CRP positivity thresholds of ≥5 or ≥10 mg/l (CRP_5, CRP_{10}) and <10 g/dl haemoglobin were pre-selected based on literature[10,15]. Different combinations of triage tests ("algorithms" defined in **Table 3**) were evaluated.

Specimen collection and bacteriological tests

Three sputa (≥1 ml each) were required from participants, who first attempted expectoration before induction was offered (whether specific individual sputa made were expectorated or induced was, due to a database error, only successfully recorded for the last 158 participants).

Induction used a nebuliser (Ultrasonic Hospital Grade WH-802, Hitech Therapy, South Africa) with 5% NaCl (Ysterplaat Medical Supplies, South Africa) for 7-10 min[18]. Two sputa were arbitrarily selected to each undergo 1% NaOH-NALC decontamination, Ziehl-Neelsen (ZN) smear microscopy, and Mycobacteria Growth Indicator Tube (MGIT) 960 culture. The other sputum underwent Ultra testing and Xpert was done on the sputum sediment remnant remaining after culture inoculation. The **Supplement** has further information on specimen storage, sputum, urine, and isolate testing. Study staff had access to all clinical, index test, and reference standard data. Sputum Ultra and microscopy (both available ≤48 h), culture and MTBDR*plus* results (typically available within 35 days) were reported for potential participant management to the programme. No adverse events occurred.

Reference standard

If at least one culture was *Mtb*-positive, participants were classified as having TB. Participants with negative culture(s) did not have TB.

Statistical analyses

Data were captured on RedCap[21]. Head-to-head analyses (only people with actionable results from the relevant tests) are presented unless stated otherwise. Sensitivity, specificity, and predictive values for tests and algorithms were calculated using two-by-two tables with 95% confidence intervals (binomial proportion method). Yield was calculated as, of the people who had an index test attempted, those with a positive result [sputum Ultra, sputum Xpert, urine Ultra (concentrated or unconcentrated), urine LF-LAM]. People who did not have a test attempted, due to human error for example, were excluded from yield calculations and non-actionable results (not positive or negative[22,23]; **Supplement**) were included. Statistical tests included McNemar's chi-squared, Mann-Whitney, two-sample proportion, Kruskal-Wallis and Spearman's coefficient using Stata (version 15; StataCorp) and GraphPad Prism (version 7.0;

GraphPad Software). Culture-negative people who were urine Ultra-positive had medical record information up to one-year post-recruitment requested to evaluate if they were later sputum confirmatory test-positive. We followed STARD analysis and reporting criteria[24].

Results

Patient characteristics

12% (107/897) of the total population was culture-positive [34% (36/107) positive by one culture]. This comprised 13% (104/800) of people in our comparison of CRP and W4SS triage tests, 13% (104/787) in our comparison of sputum confirmatory tests, and 13% (97/732) in our urine test comparison (**Figure 2A-C)**. Patient characteristics, including W4SS, TB, and smear statuses, are in **Table 1**. 25% (26/104) of people with TB were W4SS-negative.

Triage tests

Individual biomarker sensitivity and specificity

Cough (any), ≥2-week cough, and W4SS had a sensitivity and specificity of 53% [95% confidence interval (CI) 43, 63] and 73% (70, 76), 42% (33, 53) and 83% (81, 86), and 77% (68, 85) and 48 (45, 52), respectively **(Table 2)**. POC CRP testing was feasible: 1% (5/835) of tests were non-actionable, all resolved upon re-testing, and all prospective results were generated within three minutes of finger prick. Sensitivity and specificity for CRP_5 were 86% (78, 92) and 49% (45, 53), and for CRP_{10} were 77% (68, 85) and 64% (61, 68). Sensitivity at both thresholds was decreased in people with CD4 counts >350 vs. ≤350 cells/µl and specificity increased. When comparisons were limited to smear-negatives or included people that did not have head-to-head data, similar trends were observed (**Supplementary Tables 1-4**). Haemoglobin had 31% (22, 43) sensitivity and 88% (86, 91) specificity, displaying similar trends to CRP across CD4 count strata (**Supplementary Table 5**).

Individual biomarker receiver operating characteristic curves

W4SS and CRP receiver operating characteristic (ROC) curves are in **Figure 3** (haemoglobin in **Supplementary Figure 1**). Areas under the ROC curves (AUROCs) were 0.78 (0.73, 0.83) for CRP, 0.70 (0.64, 0.75) for W4SS, and 0.70 (0.64, 0.75) for haemoglobin. Among people with CD4 counts ≤350 cells/µl, higher AUROCs occurred for each biomarker vs. those with

>350 cells/µl. AUROCs were also higher for W4SS-positives vs. -negatives. We next assessed biomarker performance under different scenarios and identified corresponding thresholds. At a rule-out threshold with sensitivity prioritised (~95%), CRP (>3 mg/l; CRP_3) sensitivity was similar vs. the WHO-recommended CRP_5 [89% (83, 94) vs. 86% (78, 92)] and specificity diminished [39% (36, 42) vs. 49% (45, 53)], suggesting CRP_3 offers small sensitivity improvements at large specificity costs. CRP and haemoglobin had small variations in rule-out thresholds across CD4 cell count and W4SS strata. Sensitivities and specificities under rule-in and Youden's index scenarios are in **Supplementary Table 6**.

Test combinations

Each biomarker's sensitivity decreased in W4SS-negatives vs. -positives whereas specificity increased **(Supplementary Tables 1, 3)**. Triage test combinations were assessed as part of "algorithms" in **Table 3**. Algorithm 3 (triage positive if W4SS-positive and CRP_{10}-positive) had similar sensitivity [71% (60, 81)] to W4SS [77% (68, 85)] and CRP_{10} [77% (68, 85)] individually, but specificity improved vs. each biomarker [77% (74, 81) vs. 48% (45, 52) for W4SS or vs 64% (61, 68) for CRP_{10}] (**Table 2, 3**). Other algorithms, including CRP in W4SS-negatives, had similar sensitivities and generally diminished specificities versus their individual components **(Figure 4A)**. We did not include haemoglobin as it was in a programmatically-selected subset.

Comparisons to WHO TPP sensitivity and specificity criteria

No point estimate achieved the minimum WHO triage test sensitivity target (90%), however, the CRP_5 and Algorithm 2 95% CIs [86% (78, 92) and 92 (84, 98), respectively] overlapped 90%. The Algorithms 3 and 4 specificity estimate [77% (74, 81) and 70 (66, 74), respectively)] overlapped with the WHO minimum specificity estimate (70%).

Effect of triage tests and algorithms on confirmatory tests required

The effect of individual tests and algorithms on the number needing downstream confirmatory

testing (number needed to test, NNT) at different prevalences is in **Figure 5**. Haemoglobin resulted in the smallest number of unnecessary referrals or false-positives, but this was at the expense of few people with TB correctly referred (true-positives). CRP_{10} reduced unnecessary testing more than W4SS, and this was enhanced further by a combination strategy also involving W4SS (Algorithms 3 and 4) but this resulted in fewer people with TB correctly referred. Briefly, CRP_{10} use would, versus W4SS, reduce Ultra NNT from 6.91 (6.25, 7.81) to 4.87 (4.41, 5.51). Improvements offered by CRP10 vs. W4SS remained relatively consistent across prevalences, with the NNT of triage strategies plateauing beyond 30% prevalence.

<u>Sputum Ultra and Xpert diagnostic accuracy</u>

Overall

Ultra and Xpert non-actionable results rates were low [1% (5/811), 1% (2/883)]. Ultra had, vs. Xpert, higher sensitivity [71% (61, 80) vs. 56% (46, 66) overall, 67% (56, 77) vs. 50% (39, 61) in smear-negatives], and lower specificity [96% (94, 97) vs. 99% (98, 100)] (**Table 4**; **Supplementary Table 7**). The NNT for Ultra to detect a culture-positive person was less than Xpert [4.87 (4.41, 5.51) vs. 5.90 (5.34. 6.67) in CRP_{10}-positives, for example] Forest plots that include all confirmatory tests are in **Figure 4A** and Euler diagrams showing overlap between confirmatory tests are in **Supplementary Figure 2.**

By CD4 count

Ultra had decreased sensitivity in people with CD4 counts >350 vs. ≤350 cells/µl [42% (19, 68) vs. 77% (66, 86)] and similar specificity [99% (97, 100) vs. 97% (95, 99)] (**Supplementary Table 7**). Within CD4 count strata, Ultra sensitivity was higher than Xpert in people with counts ≤350 cells/µl [77% (66, 86) vs. 60% (48, 71)] and specificity was [97% (95, 99) vs. 99% (98, 100). Similar trends were observed in non-head-to-head data (**Supplementary Table 9**).

By W4SS

Ultra's sensitivity differed from Xpert's in W4SS-positives [80% (69, 88) vs. 66% (54, 76)] (**Table 4**). Trends in Ultra and Xpert comparative accuracy were similar in smear-negative people and non-head-to-head data (**Supplementary Table 8, 10**). Specificities were similar irrespective of W4SS status.

By previous TB

Ultra's specificity did not differ by previous TB status [97% (91, 99) vs. 98% (97, 99); p=0.50] (**Supplementary Table 11**).

Ultra "trace" recategorisation strategies

Reclassifying trace results from positive-to-negative resulted in similar sensitivity [71% (61, 80) vs. 66% (57, 76); Δ-5% (-10, 0); p=0.454] and increased specificity [98% (96,99) vs. 99% (99,100); +Δ1% (1, 3); p=0.011] **(Supplementary Table 12)**. Trace exclusion had a similar effect. Conclusions were unchanged when stratified by previous TB status.

<u>Urine Ultra and LF-LAM</u>

Ultra non-actionable results

Ultra results on concentrated urine were frequently non-actionable [16% (118/732)]. When an unconcentrated aliquot was tested, 99% (117/118) of non-actionables resolved and 11% (13/117) were positive (**Supplementary text, Supplementary Figure 3A**).

Sensitivity and specificity

Of the concentrated urine Ultra-positives, 73% (23/33) had a corresponding unconcentrated urine test Ultra-positive (**Supplementary text, Supplementary Figure 3B)**. Amongst the 732 PLHIV in the confirmatory urine test analysis, Ultra and LF-LAM had low sensitivities [25% (17, 35) vs. 15% (9, 24); p=0.070] and high specificities [99 (98, 100) vs. 99 (98, 100); p=0.248] (**Figure 2C, Supplementary Table 13**).

Overlap with other tests

Ultra and LF-LAM overlap was high, with each test infrequently giving the only positive non-sputum result [3% (5/129) and 7% (9/129), respectively] (**Supplementary Figure 2**). Five culture-negative people were urine Ultra-positive and, of these, four had a later sputum-based programmatic diagnosis of TB within a year of recruitment (one had no data).

<u>Sputum scarcity and sputum and urine test yield</u>

Sputum scarcity

Amongst people in whom it was known whether sputum induction was necessary, 31% [49/158 (24, 39)] were sputum-scarce (11 W4SS-negative, 38 W4SS-positive), of which 20% [10/49 (10, 34) were culture-positive (all W4SS-positive); comparable to the culture-positivity rate among people who could expectorate [14% (15/109; 7, 22)] **(Supplementary Table 14)**.

Overall yield

17% (152/897; 15, 20) of people had at least one positive sputum (culture, Xpert, Ultra) or urine (Ultra, LF-LAM) result. When limited to people in which each test was available and the absence of a result was not due to human error or stock outs, the proportion any-test-positive was 18% (145/804; 15, 21). Sputum tests had highest yields: 71% (103/145; 63, 78) for culture [34% (36/107; 25, 43) single culture-positive], 61% (89/145; 53, 69) for Ultra and 44% (64/145; 36, 53) for Xpert. Urine tests had lower yields with 30% (43/145; 22, 38) and 17% (25/145; 12, 24) positive by Ultra and LF-LAM, respectively.

Yield in sputum-scarce people

Of the 158 people in whom it was known whether expectorated sputum could be produced before induction was done, 126 had all the above tests attempted. 23% (29/126; 16, 21) had ≥1 positive result. Of these, 83% (24/29; 64, 95), 66% (19/29; 46, 82) and 59% (17/29; 39, 76) had sputum (expectorated or induced) positive by culture, Ultra, and Xpert respectively. 34% (10/29; 18, 54) and 24% (7/29; 10, 44) had urine positive by Ultra and LF-LAM, respectively. 33% (42/126; 25, 42) were unable to naturally expectorate and induction was their only source

of sputum. If induction was done, yields of the aforementioned sputum tests would decrease to 55% (16/29; 36, 74), 45% (13/29; 26, 64), and 38% (11/29; 21, 58), respectively. The aforementioned urine tests' yields were 17% (5/29; 6, 36) and 10% (3/29; 2, 27), respectively. Comparisons of tests did not show yield and sensitivity and specificity differences when results from expectorated or induced sputum were compared **(Supplementary Table 15)**.

Discussion

Our key findings are: 1) CRP testing is feasible at POC and superior to W4SS for triaging ART-initiators, reducing unnecessary referrals, improving NNT, and approaching but not capable of meeting the WHO TPP minimum sensitivity and specificity benchmarks, 2) triage algorithms combining W4SS and CRP approach the WHO-recommended optimal specificity target but result in more missed TB than CRP alone, 3) sputum Ultra is more sensitive than Xpert for culture-confirmed TB which is common, both in W4SS-negative and sputum-scarce people, 4) offering sputum induction enhances yield (beyond that possible using urine tests), and 5) urine Ultra and LF-LAM have similar performance and urine Ultra is hampered by high non-actionable result rates. These data inform triage and confirmatory testing strategies, including specimen acquisition, in ART-initiators.

PLHIV should be screened for TB, a process most efficient at ART initiation when patients are immunosuppressed and have relatively high pre-test probability of TB. ART initiators are already within HIV treatment cascades, representing a captured population in a setting in which that we now show CRP measurement at POC is highly feasible and a better alternative to symptoms, significantly reducing unnecessary onward referrals (from 452 to 314 people per 1000 using CRP_{10}, translating into a NNT reduction of ~7 to ~5). In line with recent WHO guidance[16], our data also informs the use of CRP at this threshold of ≥10 mg/l, with lower thresholds potentially negating benefits CRP offers over W4SS. Importantly, and in contrast to recent community-based evaluations of CRP[25], we show little added benefit of combining W4SS with CRP in ART initiators, unless large TB case detection reductions are acceptable, and that haemoglobin should not be considered further due to low sensitivity.

Ultra, despite being WHO-recommended, is unevaluated in ART-initiators without syndromic preselection. We now provide the first data to support its use in this diagnostically-challenging

population versus the previous generation Xpert, where sensitivity was higher (71% vs. 56%), including in people who did not meet the W4SS and in people with advanced HIV (CD4 counts <350 cells/µl). This supports Ultra's use as part of recent efforts to move beyond symptom criteria in high burden settings[26]. Encouragingly, we did not observe dramatically reduced Ultra specificity in contrast to previous research in our setting, however, such research in self-reporting people with presumptive TB, who have a higher rate of previous TB that unselected ART initiators[23].

Although non-sputum tests are a priority for TB control, there are, unfortunately, unlikely to be alternatives available in the short-term for bacteriological confirmatory testing[1]. Furthermore, countries are heavily invested in sputum-based testing infrastructure and may wish to consider induction, which our data support the use of: 31% required induction to make at least one sputum and without induction the proportion of people with a positive confirmatory test result detected by Ultra would reduce from 65% to 45%. We therefore recommend that programmes consider making sputum induction available at ART-initiation sites, which is currently not standard-of-care in most high burden settings including South Africa. This is especially important given the relatively poor performance of urine tests (although LF-LAM has low yield, it should, due to its low cost, POC nature and, unlike urine Ultra, absence of non-actionable results, still be available in line with current guidance[27]). We also suggest that the utility of sputum induction as a diagnostic intervention, as well as associated implementation challenges like biosafety, requires further evaluation as, based on our data, induction improved case detection more than a new diagnostic test (i.e., Ultra vs. Xpert).

This study has strengths and limitations. We included people irrespective of symptoms, did CRP at POC, utilised a two-sputa culture reference standard, and offered all people sputum induction, however, due to database error, could only confidently establish if induction was

necessary in a consecutive subset. Such data are generally limited and our study is still one of only a few to describe sputum production needs in detail in PLHIV[18,28], which informs on the potential impact of induction facilities and non-sputum tests. Other limitations include haemoglobin was done in a programmatically-selected subset, which may bias accuracy estimates. We did not analyse the cost, cost-effectiveness, or affordability and these require study. Furthermore, shifting away from decades of entrenched W4SS-based triage in this population will present significant implementation challenges. Barriers to POC CRP implementation require identification, including detailing its impact on patient important outcomes in clinical trials.

In summary, POC CRP is an alternative TB triage tool that has improved specificity compared to self-reported W4SS in PLHIV initiating ART that reduces unnecessary testing. Ultra has improved sensitivity compared to Xpert and detects W4SS-negative TB, and Ultra's yield is significantly enhanced through sputum induction provision. These tools should be pursued for implementation in this key risk group, together with further evaluations in different populations and settings.

Figure legends

Figure 1. Study profile showing participant enrolment, specimen collection and processing, and tests done. Abbreviations: ART, antiretroviral treatment; CRP, C-reactive protein; LF-LAM, Determine TB LAM Ag test (LF-LAM); MGIT960, Mycobacterial Growth Indicator Tube; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF; ZN, Ziehl–Neelsen.

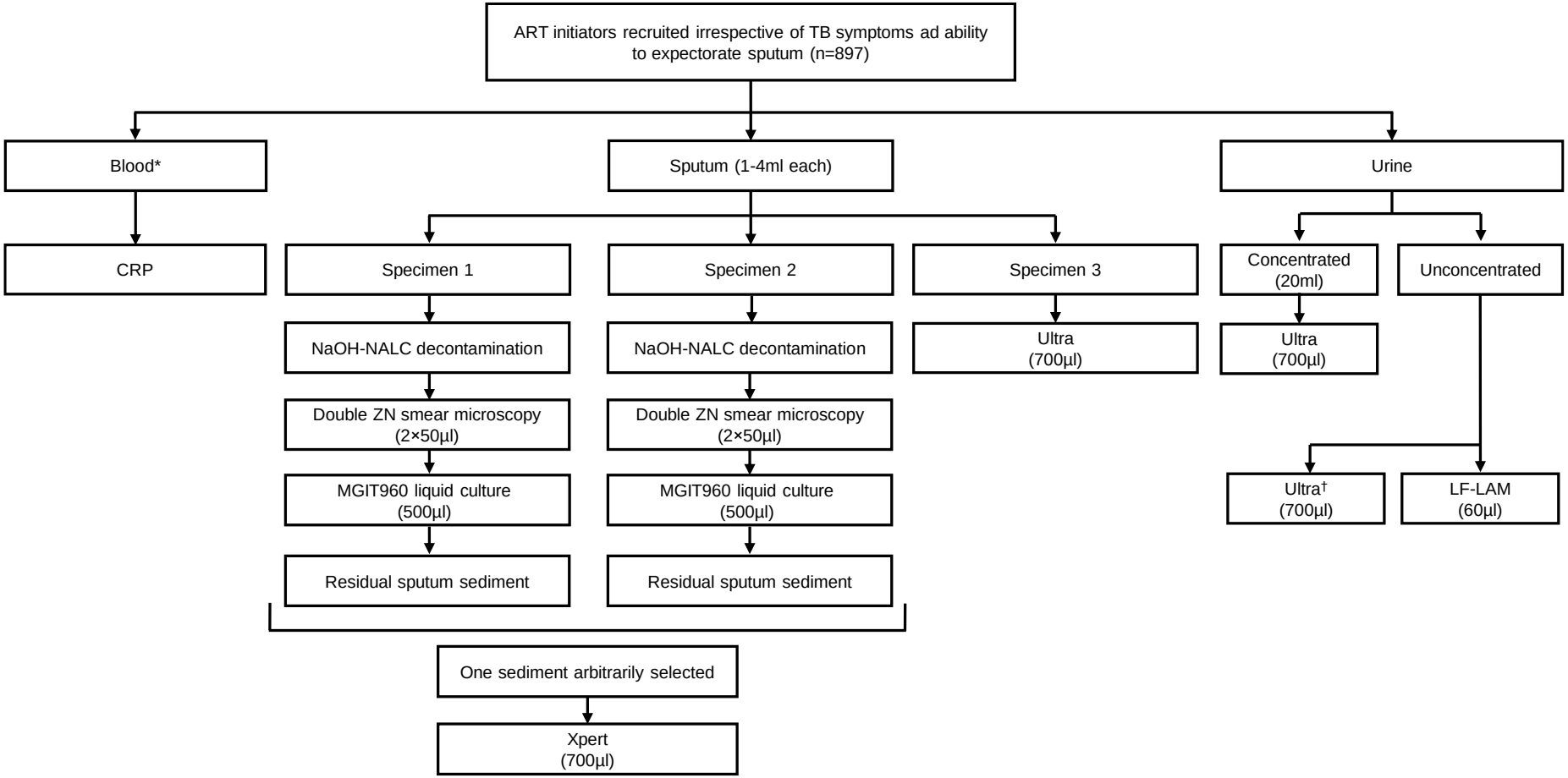
Figure 2. Flow diagrams showing the number of people in head-to-head comparisons of (**A**) triage (CRP, W4SS), (**B**) confirmatory (Xpert, Ultra), and (**C**) urine (concentrated Ultra, LF-LAM) tests. CRP correctly classified more people without TB compared to W4SS, Ultra detected more TB than Xpert on sputum and, on concentrated urine, more TB than LF-LAM. Abbreviations: CRP, C-reactive protein; LF-LAM, Determine TB LAM Ag ; POC, point of care; RIF, rifampicin; Ultra, Xpert MTB/RIF Ultra; W4SS, WHO-recommended four-symptom screen; Xpert; Xpert MTB/RIF.

Figure 3. ROCs of CRP and number of W4SS in (**A**) all ART initiators, (**B, C**) stratified by W4SS-status, and (**D, E**) by CD4 cell counts. CRP had higher AUCROC in W4SS-positives than -negatives and in patients with lower than higher CD4 cell counts (as did W4SS). CRP was the only test for which CI overlapped with the WHO TPP minimal sensitivity and specificity target, and this was only in people with advanced disease (W4SS-positive or low CD4 count). AUROCs with 95% CIs with double and single ticks approximating the optimum and minimum WHO TPP sensitivity and specificity, respectively are shown. The number of symptoms are shown. Abbreviations: CRP, C-reactive protein; Opt., optimal WHO TPP; Min., minimum; WHO TPP; WHO TPP, World Health Organization target product profile; W4SS, WHO-recommended four-symptom screen.

Figure 4. Summaries of triage and confirmatory test performance. **(A)** Forrest plots comparing sensitivity and specificity point estimates (with 95% CIs) of triage tests and algorithms as well as sputum and urine confirmatory tests. Black dashed vertical lines indicate WHO TPP estimates. **(B)** Effect of different triage tests and algorithms (defined in **Table 3**) on participant classification, showing CRP_{10} alone or in combination with W4SS (Algorithms 3 and 4) to result in fewer onward unnecessary referrals (false-positives). Abbreviations: CRP, C-reactive protein; Hb, haemoglobin; LF-LAM, Determine TB LAM Ag test; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; W4SS, WHO-recommended four-symptom screen; WHO, World Health Organization; WHO TTP, WHO target product profile; Xpert, Xpert MTB/RIF.

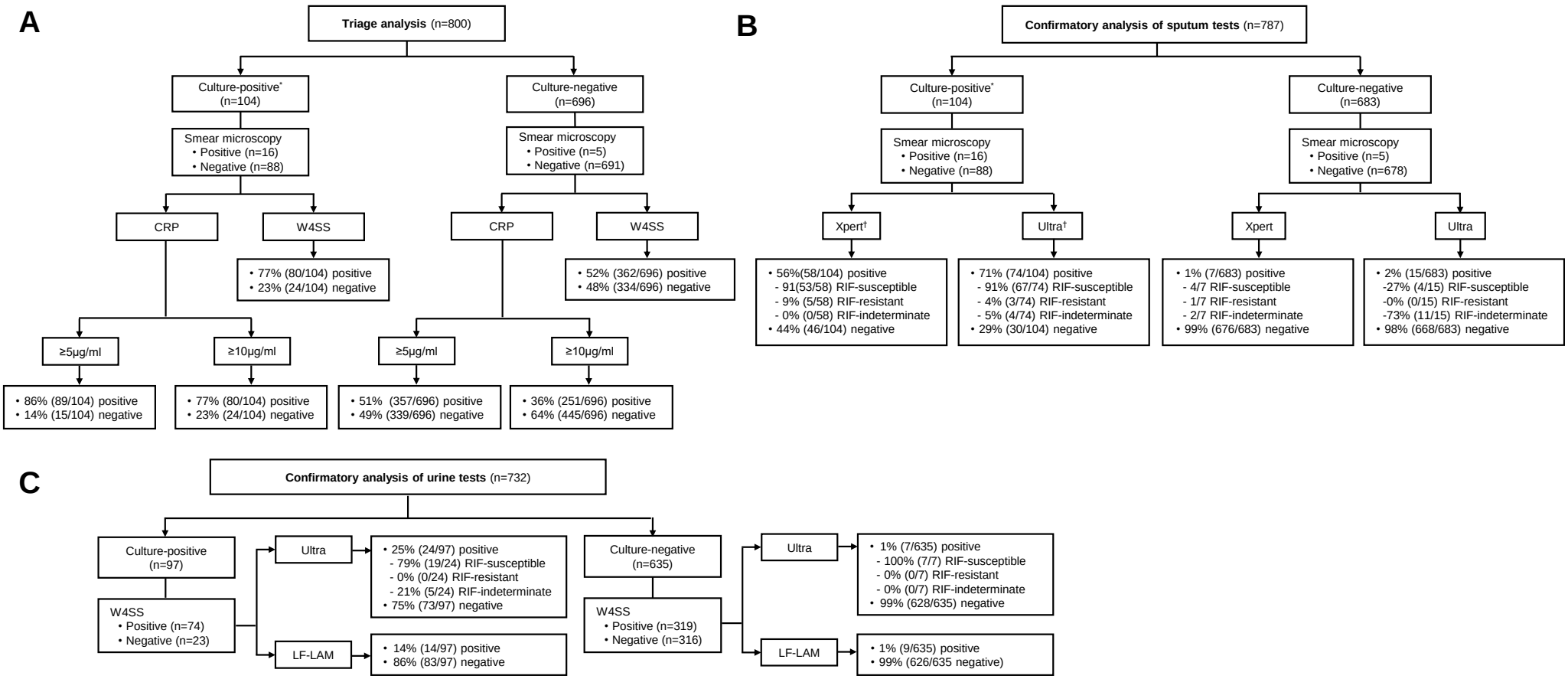
Figure 5. NNT for Ultra and Xpert to detect one culture-positive TB case when different triage methods are used in **(A)** our cohort and **(B)**, for Ultra only, modelled against different TB prevalences vs. an Ultra-in-all scenario. Use of CRP_{10} would result in a TB case correctly detected every five rather than, as for W4SS, every seven people. Other triage methods had lower NNTs but these would be offset by diminished sensitivities. In **B**, not all triage methods are shown and the grey column shows our prevalence. Abbreviations: CRP, C-reactive protein; Hb, haemoglobin; NNT, number needed to test; Ultra, Xpert MTB/RIF Ultra; W4SS, World Health Organization-recommended four symptom screen; Xpert, Xpert MTB/RIF.

Figure 1.



*Haemoglobin was done when programmatically indicated outside of the study.
†Unconcentrated Ultra done if concentrated Ultra non-actionable or positive.

Figure 2.



Reasons people excluded from head-to-head analyses: (**A**) no culture [n=42; n=22 no sputum available, n=20 contaminated] and/or no CRP (n=55) results, (**B**) no culture (n=43; n=22 no specimen, n=21 contaminated), no actionable Xpert (n=4 "no result", n=1 "error", n=60 no or under volume specimen) and/or results (n=2 "error"), (**C**) no culture (n=46; n=23 no specimen, n=23 contaminated), non-actionable concentrated urine Ultra (n=113; n=3 "no result", n=22 invalid, n=88 invalid), and/or insufficient urine (n=6; n=3 for both LF-LAM and concentrated Ultra, n=3 for concentrated Ultra only).

*69/104 (66%) people were culture-positive for both sputa; 30/104 (29%) positive for one sputum and negative for the other; 4/104 (4%) positive for one sputum and the other contaminated; 1/104 (1%) positive for one sputum and culture not done for the other.

†Xpert detected three Ultra-negative cases [10% 3/30 (2, 27)]. Ultra detected 19 Xpert-negative cases [41% 19/46 (27, 57)].

Figure 3.

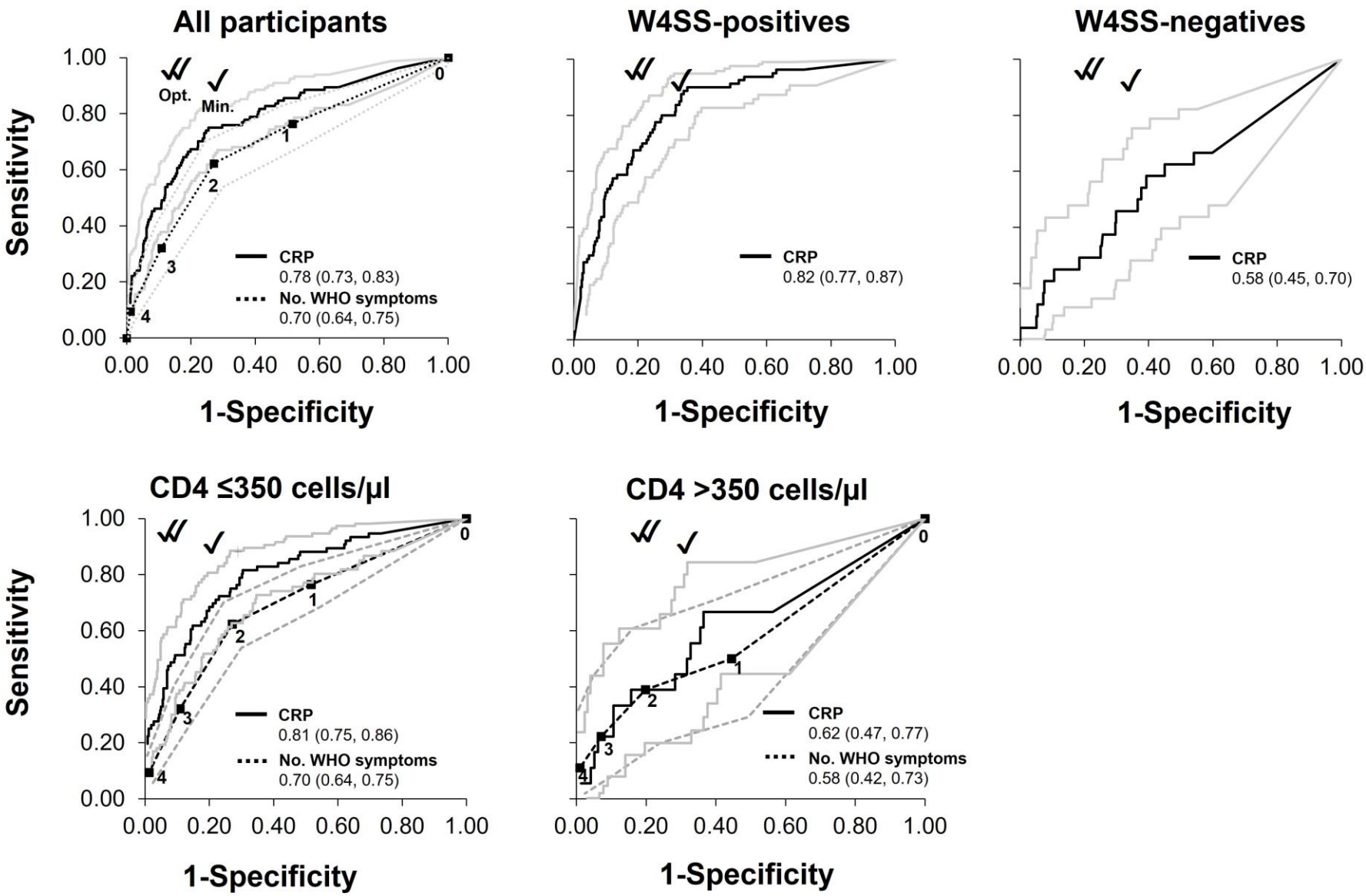

All participants
W4SS-positives
W4SS-negatives
CD4 ≤350 cells/µl
CD4 >350 cells/µl
Sensitivity
1-Specificity
Opt.
Min.
CRP
0.78 (0.73, 0.83)
No. WHO symptoms
0.70 (0.64, 0.75)
CRP
0.82 (0.77, 0.87)
CRP
0.58 (0.45, 0.70)
CRP
0.81 (0.75, 0.86)
No. WHO symptoms
0.70 (0.64, 0.75)
CRP
0.62 (0.47, 0.77)
No. WHO symptoms
0.58 (0.42, 0.73)

Figure 4.

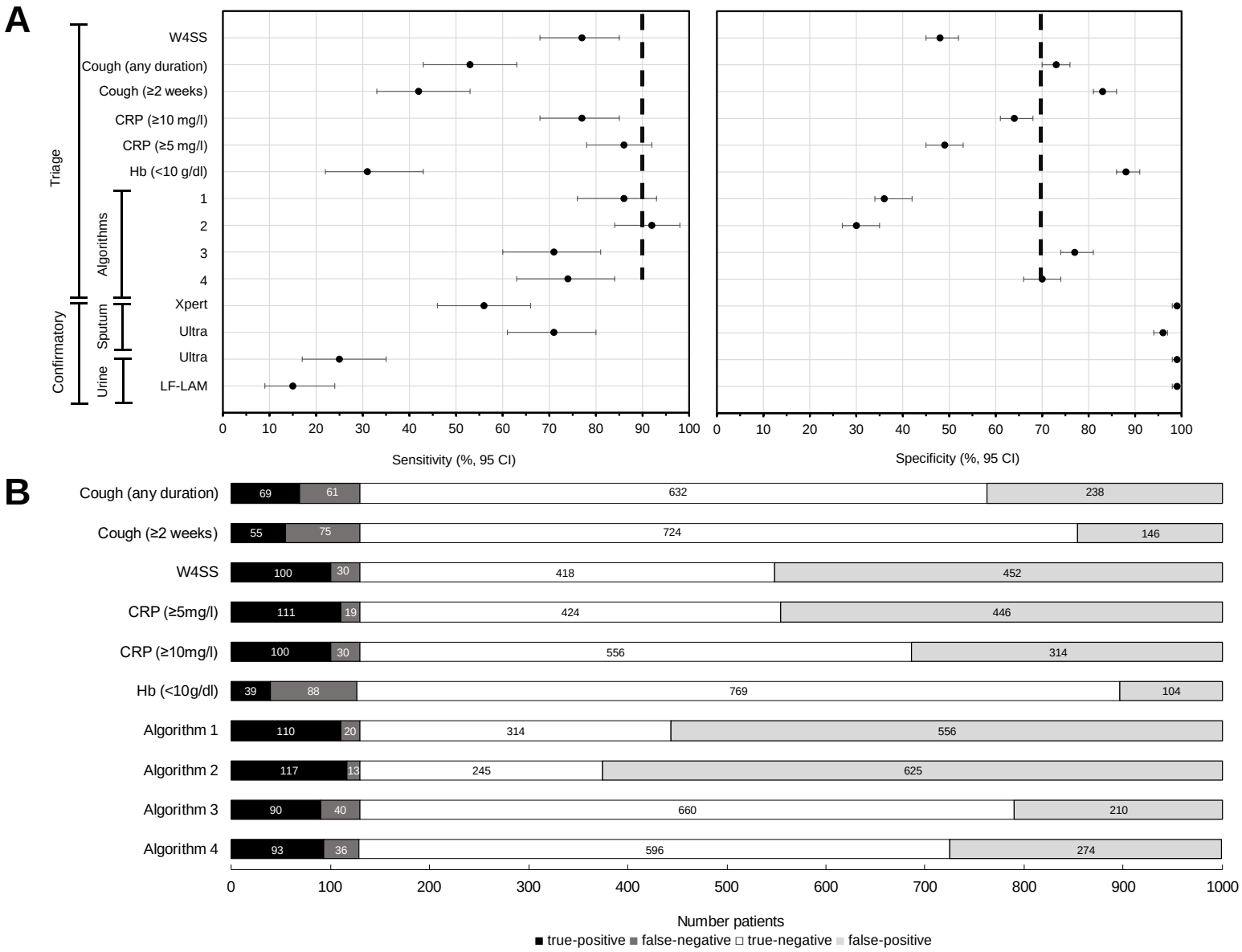

A
Triage
Algorithms
Confirmatory
Sputum
Urine
W4SS
Cough (any duration)
Cough (≥2 weeks)
CRP (≥10 mg/l)
CRP (≥5 mg/l)
Hb (<10 g/dl)
1
2
3
4
Xpert
Ultra
Ultra
LF-LAM
0 10 20 30 40 50 60 70 80 90 100
Sensitivity (%, 95 CI)
Specificity (%, 95 CI)
B
Cough (any duration) 69 61 632 238
Cough (≥2 weeks) 55 75 724 146
W4SS 100 30 418 452
CRP (≥5mg/l) 111 19 424 446
CRP (≥10mg/l) 100 30 556 314
Hb (<10g/dl) 39 88 769 104
Algorithm 1 110 20 314 556
Algorithm 2 117 13 245 625
Algorithm 3 90 40 660 210
Algorithm 4 93 36 596 274
0 100 200 300 400 500 600 700 800 900 1000
Number patients
true-positive false-negative true-negative false-positive

Figure 5.

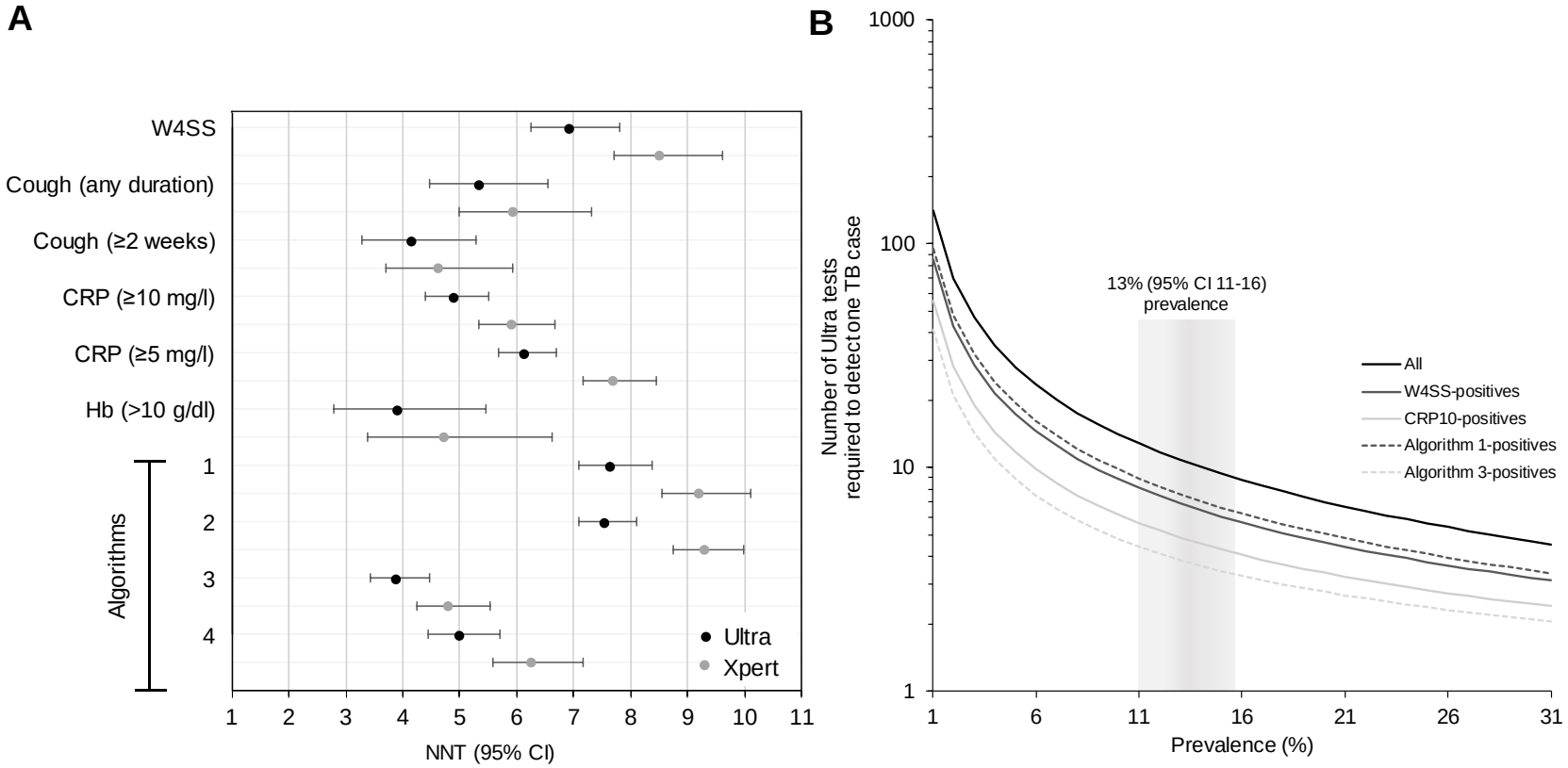

A
W4SS
Cough (any duration)
Cough (≥2 weeks)
CRP (≥10 mg/l)
CRP (≥5 mg/l)
Hb (>10 g/dl)
Algorithms
1
2
3
4
Ultra
Xpert
1 2 3 4 5 6 7 8 9 10 11
NNT (95% CI)
B
1000
100
10
1
Number of Ultra tests required to detect one TB case
13% (95% CI 11-16) prevalence
All
W4SS-positives
CRP10-positives
Algorithm 1-positives
Algorithm 3-positives
1 6 11 16 21 26 31
Prevalence (%)

Table 1. Demographic and clinical characteristics by W4SS, culture, and smear statuses. Compared to W4SS-positives, W4SS-negatives were more likely to be younger, female, non-smokers, culture-negative, have lower CRP but higher CD4 and haemoglobin levels. W4SS-positives were also more likely to have TB, be able to naturally expectorate sputum and, in those with TB, greater sputum mycobacterial load. Culture-positives were, compared to culture-negatives, more likely to be older, of mixed ancestry, smokers, have lower CD4 counts and haemoglobin, and higher CRP levels. Data are median (IQR) or n (%).

	Overall [a]n=787	**W4SS status**		**Culture and smear status**			
				Culture-positive			
		Positive 430/787 (55)	**Negative** 357/787 (45)	**All** 104/787 (13)	**Smear-positive** 16/104 (15)	**Smear-negative** 88/104 (85)	**Culture-negative** 683/787 (87)
Demographic							
Age, years	32 (26-39)	33 (28-40)	31 (25-37) **p<0.001***	34 (29-39)	37 (33-42)	34 (29-39) p=0.170ǂ	31 (26-38) **p=0.010****
Female	463/787 (59)	233/430 (54)	230/357 (64) **p<0.001***	43/104 (41)	4/16 (25)	39/88 (44) p=0.149ǂ	420/683 (61) **p<0.001****
Ethnicity‡							
Mixed ancestry	158/787 (21)	93/430 (22)	70/357 (20) p=0.486*	31/104 (30)	4/16 (25)	27/88 (31) p=0.648ǂ	132/683 (19) **p=0.014****
Black	618/787 (79)	334/430 (78)	284/357 (80) **p=0.009***	72/104 (69)	11/16 (69)	61/88 (69) p=0.964ǂ	546/683 (80) **p=0.013****
Tobacco smoker (past or current)	372/787 (47)	239/430 (56)	133/357 (37) **p<0.001***	64/104 (62)	9/16 (56)	55/88 (63) p=0.636ǂ	308/683 (45) **p<0.001****
Clinical							
Number of W4SS symptoms (0-4)	1 (0-2)	2 (1-2)	-	2 (1-3)	3 (2-3)	2 (1-3) p=0.171ǂ	1 (0-2) **p<0.001****
Able to expectorate at least one ≥1 ml sputum	80/121† (66)	47/79 (59)	33/42 (79) **p=0.035***	16/24 (67)	3/5 (60)	13/19 (68) p=0.722ǂ	64/97 (66) p=0.949**
Culture-positive	104/787 (13)	78/430 (18)	26/357 (7) **p<0.001***	104/104 (100)	16/16 (100)	88/88 (100) p>0.999ǂ	-
TTP (days)	13 (9-18)	12 (9-15)	15 (12-21) **p=0.013***	13 (9-18)	6 (5-7)	13 (11-19) **p<0.001**ǂ	-
Previous TB	112/787 (14)	69/430 (16)	43/357 (12) p=0.110*	16/104 (15)	3/16 (19)	13/88 (15) p=0.685ǂ	96/683 (14) p=0.718**
CD4 (cells/µl)	299 (170-483)	263 (110-428)	348 (227-521) **p<0.001***	193 (60-285)	199 (45-257)	192 (61-285) p=0. 896ǂ	326 (197-506) **p<0.001****
CRP (mg/l)	7 (3-34)	14 (3-77)	4 (3-11) **p<0.001***	72 (14-150)	125 (101-286)	45 (10-139) **p<0.001**ǂ	6 (3-20) **p<0.001****
Haemoglobin (g/dl)	13 (12-14)	13 (11-14)	14 (12-15) **p<0.001***	12 (10-13)	10 (10-13)	12 (10-13) p=0.311ǂ	13 (12-14) **p<0.001****

[a]N are shown for people included in the Ultra and Xpert head-to-head comparison.
Within row p-values are for comparisons by *W4SS, ǂsmear, or **culture statuses.
‡Eight people self-reported as "other ethnicity".
†121 people had whether sputum was expectorated or induced recorded.
§Negative values are days before recruitment, positives days after.
Missing or not done: CRP (n=171; n=54 not done; n=117 outside detection range, see **Supplement**), CD4 (n=4 not done; n=47 older than three months), haemoglobin (n=50 not captured).
Abbreviations: CRP, C-reactive protein; IQR, interquartile range; TB, tuberculosis; TTP, time-to-positivity; W4SS, WHO-recommended four-symptom screen.

Table 2. Head-to-head diagnostic accuracy of cough, W4SS, and CRP stratified by CD4 count. Cough (any or ≥2 weeks) had decreased sensitivity vs. W4SS but W4SS had sub-optimal specificity. In people with CD4 count >350 cells/µl, only the ≥ 10 mg/l CRP (CRP_{10}) threshold (not ≥ 5 mg/l; CRP_5) had increased specificity vs. W4SS whereas in people CD4 counts ≤350 cells/µl CRP_5 and not CRP_{10} had increased specificity vs. W4SS. CRP_{10} had increased specificity CRP_5 within each CD4 count stratum without a sensitivity decrement. Data are %, 95% CI, and n/N.

	All people (n=800)				**CD4 count >350 cells/µl (n=312)**				**CD4 count ≤350 cells/µl (n=441)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Cough (any duration)	53 (43, 63) 55/104	73 (70, 76) 506/696	22 (18, 29) 55/245	91 (89, 94) 506/555	39 (18, 65) 7/18	80 (75, 85) 235/294	11 (5, 21) 7/66	96 (93, 98) 235/246	54 (43, 66) 41/76 p=0.251**	67 (63, 72) 245/365 **p<0.001****	25 (19, 33) 41/161 **p=0.013****	88 (84, 92) 245/280 **p<0.001****
≥2 weeks	42 (33, 53) 44/104 p=0.127*	83 (81, 86) 579/696 **p<0.001***	27 (21, 35) 44/161 p=0.263*	91 (89, 93) 579/639 p=0.737*	28 (10, 54) 5/18 p=0.480*	88 (84, 92) 258/294 **p=0.010***	12 (5, 27) 5/41 p=0.800*	95 (92, 98) 258/271 p=0.861*	43 (33, 56) 33/76 p=0.224** p=0.194*	81 (77, 85) 294/365 **p=0.013**** **p<0.001***	32 (23, 42) 33/104 **p=0.016**** p=0.267*	87 (84, 91) 294/337 **p=0.001**** p=0.923*
W4SS	77 (68, 85) 80/104	48 (45, 52) 334/696	18 (15, 23) 80/442	93 (91, 96) 334/358	50 (27, 74) 9/18	54 (49, 61) 160/294	6 (3, 12) 9/143	95 (91, 98) 160/169	82 (72, 90) 62/76 **p=0.005****	44 (39, 50) 161/365 **p=0.009****	23 (19, 29) 62/266 **p<0.001****	92 (87, 96) 161/175 p=0.321**
CRP_5	86 (78, 92) 89/104 **p<0.001**†** p=0.101‡	49 (46, 53) 339/696 **p<0.001**†** p=0.789‡	20 (17, 24) 89/446 p=0.440† p=0.481‡	96 (94, 98) 339/354 **p=0.008**†** p=0.148‡	67 (41, 87) 12/18 p=0.095† p=0.311‡	60 (55, 66) 176/294 **p<0.001**†** p=0.182‡	9 (5, 16) 12/130 p=0.758† p=0.363‡	97 (93, 99) 176/182 p=0.538† p=0.348‡	89 (81, 96) 68/76 p=0.089** **p<0.001**†** p=0.667‡	41 (36, 46) 148/365 p=0.749** **p<0.001**†** **p<0.001**‡**	24 (20, 30) 68/285 **p<0.001**** p=0.705† p=0.109‡	95 (91, 98) 148/156 p=0.325** **p=0.013**†** p=0.278‡
CRP_{10}	77 (68, 85) 80/104 **p<0.001**†** p>0.999‡ p=0.394§	64 (61, 68) 445/696 **p<0.001**†** **p<0.001**‡** **p<0.001**§**	24 (20, 30) 80/331 p=0.630† **p=0.039**‡** p=0.159§	95 (93, 97) 445/469 **p=0.021**†** p=0.334‡ p=0.556§	39 (18, 65) 7/18 p>0.999† p=0.502‡ p=0.095§	72 (67, 77) 211/294 **p=0.021**†** **p<0.001**‡** **p=0.002**§**	8 (4, 16) 7/90 p=0.541† p=0.663‡ p=0.706§	95 (92, 98) 211/222 p=0.805† p=0.869‡ p=0.409§	84 (75, 92) 64/76 **p<0.001**** **p<0.001**†** p=0.167‡ p=0.337§	59 (54, 64) 214/365 **p<0.001**** **p=0.018**†** p=0.330‡ **p<0.001**§**	30 (24, 37) 64/215 **p=0.001**** p=0.358† p=0.879‡ p=0.138§	95 (91, 98) 214/226 p=0.940** **p=0.006**†** p=0.295‡ p=0.938§

**Within row p-values comparing data across CD4 count strata.
Within column p-values: *Cough (any) vs. ≥2 weeks; †Cough (any) vs. CRP; ‡W4SS vs. CRP; §CRP_5 vs CRP_{10}.
Missing data: no CD4 within three months (n=47).
Abbreviations: CRP, C-reactive protein; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; W4SS, WHO-recommended four-symptom screen.

Table 3. Diagnostic accuracy of parallel and sequential algorithms involving W4SS and CRP (stratified by CD4 cell count). A sequential W4SS-CRP approach (Algorithm 3) resulted in a more balanced sensitivity and specificity whereas a parallel approach (Algorithm 1) resulted in higher sensitivity but lower specificity. Each algorithm's sensitivity increased at lower CD4 counts compared to higher counts while specificity decreased. People who, in the second column, do not meet the definition of triage-positive are triage-negative. Data are %, 95% CI, and n/N.

	Triage positive if*:	**All people (n=603)**				**CD4 count >350 cells/µl (n=243)**				**CD4 count ≤350 cells/µl (n=360)**			
		Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
W4SS or CRP first	Algorithm 1: W4SS- and/or CRP_{10}-positive	86 (76, 93) 65/76	36 (34, 42) 196/527	16 (13, 21) 65/396	95 (91, 98) 196/207	69 (39, 91) 9/13	42 (36, 49) 97/230	6 (3, 12) 9/142	96 (91, 99) 97/101	89 (79, 96) 56/63	33 (28, 40) 99/297	22 (18, 28) 56/254	93 (87, 98) 99/106
	Algorithm 2: W4SS- and/or CRP_5-positive	92 (84, 98) 70/76	30 (27, 35) 158/527	16 (13, 20) 70/439	96 (93, 99) 158/164	85 (55, 99) 11/13	36 (30, 43) 83/230	7 (4, 13) 11/155	98 (92, 100) 83/85	94 (85, 99) 59/63	25 (21, 31) 75/297	21 (17, 27) 59/281	95 (88, 99) 75/79
W4SS first	Algorithm 3: If W4SS-positive, CRP_{10}-positive	71 (60, 81) 54/76	77 (74, 81) 406/527	31 (25, 39) 54/175	95 (93, 97) 406/428	38 (14, 69) 5/13	83 (79, 89) 192/230	12 (4, 26) 5/43	96 (93, 99) 192/200	78 (66, 88) 49/63	72 (67, 78) 214/297	37 (29, 46) 49/132	94 (90, 97) 214/228
	Algorithm 4: If W4SS-positive, CRP_5-positive	74 (63, 84) 56/76	70 (66, 74) 367/527	26 (21, 33) 56/216	95 (93, 97) 367/387	54 (26, 81) 7/13	79 (73, 84) 181/230	13 (6, 25) 7/56	97 (94, 99) 181/187	78 (66, 88) 49/63	63 (57, 69) 186/297	31 (24, 39) 49/160	93 (89, 97) 186/200

*Algorithms 5 [W4SS-positive or, if W4SS-negative, CRP_{10}-positive] and 7 [W4SS-positive or, if W4SS-negative, CRP_5-positive] had the same performances as Algorithm 1, whereas Algorithms 6 [W4SS-positive or, if W4SS-negative, CRP_5-positive] and 8 [CRP_5-positive or, if CRP_5-negative, W4SS-positive] had the same performance as Algorithm 2.
Missing or CD4 data ≥3-month-old months (n=31).
Abbreviations: CRP, C-reactive protein; CRP_5, CRP ≥ 5mg/l; CRP_5, CRP ≥10mg/l; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; W4SS; WHO-recommended four-symptoms screen.

Table 4. Head-to-head diagnostic accuracy of sputum Xpert and Ultra for TB stratified by W4SS status. Ultra's sensitivity was higher than Xpert, with Ultra detecting W4SS-negatives missed by Xpert. Ultra specificity decreased compared to Xpert. Sensitivities were higher in W4SS-positives than -negatives. Data are %, 95% CI, and n/N.

		All people (n=787)				**W4SS-negative (n=357)**				**W4SS-positive (n=430)**			
		Sensitivity	**Sensitivity**	**PPV**	**NPV**	**Sensitivity**	**Sensitivity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Smear microscopy		15 (9, 24) 16/104	99 (98, 100) 678/683	76 (53, 92) 16/21	89 (86, 91) 678/766	12 (3, 31) 3/26	100 (99, 100) 331/331	100 (30, 100) 3/3	94 (91, 96) 331/354	17 (10, 27) 13/78 p=0.530*	99 (97, 100) 347/352 **p=0.030***	73 (47, 91) 13/18 p=0.296*	85 (81, 88) 347/412 **p<0.001***
Xpert		56 (46, 66) 58/104 **p<0.001**†	99 (98, 100) 676/683 p=0.562†	89 (79, 96) 58/65 p=0.134†	94 (92, 95) 676/722 **p=0.001**†	27 (12, 48) 7/26 p=0.159†	100 (98, 100) 328/331 p=0.083†	70 (35, 94) 7/10 p=0.279†	95 (92, 97) 328/347 p=0.569†	66 (54, 76) 51/78 **p=0.001*** **p<0.001**†	99 (98, 100) 348/352 p=0.766* p=0.737†	93 (83, 98) 51/55 **p=0.033*** **p=0.022**†	93 (90, 96) 348/375 p=0.343* **p<0.001**†
	Ultra-negative	10 (2, 27) 3/30	99 (98, 100) 663/668	38 (9, 76) 3/8	96 (94, 97) 663/690	0 (0, 24) 0/14	100 (98, 100) 324/327	0 (0, 71) 0/3	96 (94, 98) 324/338	19 (5, 46) 3/16 p=0.088*	100 (98, 100) 339/341 p=0.620*	60 (15, 95) 3/5 p=0.090*	97 (94, 99) 339/352 p=0.761*
Ultra		71 (61, 80) 74/104 **p<0.001**‡	98 (96, 99) 668/683 **p<0.001**‡	83 (74, 90) 74/89 p=0.287‡	96 (94, 97) 668/698 p=0.083‡	47 (27, 67) 12/26 p=0.150‡	99 (97, 100) 327/331 p=0.704‡	75 (48, 93) 12/16 p=0.780‡	96 (94, 98) 327/341 p=0.401‡	80 (69, 88) 62/78 **p=0.001*** **p=0.049**‡	97 (95, 99) 341/352 p=0.088* p=0.435‡	85 (75, 93) 62/73 p=0.337* p=0.175‡	96 (93, 98) 341/357 p=0.807* p=0.118‡
	Xpert-negative	41 (27, 57) 19/46	98 (97, 99) 663/676	59 (41, 76) 19/32	96 (94, 97) 663/690	27 (10, 52) 5/19	99 (97, 100) 324/328	56 (22, 87) 5/9	96 (94, 98) 324/338	52 (32, 72) 14/27 p=0.083*	98 (96, 99) 339/348 p=0.196*	61 (39, 81) 14/23 p=0.783*	97 (94, 99) 339/352 p=0.761*

*Within row p-values comparing data across W4SS strata.
Within column p-values: †smear vs. Xpert; ‡Xpert vs. Ultra.
Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis; W4SS, WHO-recommended four-symptom screen; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Acknowledgements

The authors thank the participants, Ms Fikiswa Seti, Sr Charmaine Van der Walt, Dr Kim Stanley, and A/Prof Gian van der Spuy. The authors thank the Department of Health, Western Cape Data Warehouse team for providing TB testing and treatment data for patients after the study visit.

Funding

GT acknowledges funding from South African Medical Research Council (SAMRC Flagship Project MRC-RFA-IFSP-01-2013), the European and Developing Countries Clinical Trials Partnership (EDCTP), and the Faculty of Medicine and Health Sciences, Stellenbosch University. BWPR and HM acknowledge funding from the Faculty of Medicine and Health Sciences, Stellenbosch University. GN acknowledges funding from the SAMRC. The work reported herein was made possible through funding by the South African Medical Research Council through its Division of Research Capacity Development under the SAMRC Internship Scholarship Programme. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the SAMRC. GT had full access to all study data and final responsibility for the decision to submit for publication.

Contributions of authors

GT conceived and designed the study. GT, RW, and PVH acquired funding. Clinical staff collected sample and curated clinical data. Laboratory analysis was done by GN, LR, HT, DM, and ZP. BWPR, GN, HM and GT did analyses and wrote the first draft. All authors critiqued analyses and revised the manuscript.

Declarations of interest

GT received in-kind donations from Cepheid and Boditech. Cepheid and Boditech had no role in study design or interpretation of results. BWPR received travel support from Cepheid to

attend a conference and present unrelated data. The authors have no financial involvement with any organisation or entity with a financial interest in, or financial conflict with, the subject matter or materials discussed in the manuscript apart from those disclosed.

References

1. Abdulgader SM, Okunola AO, Ndlangalavu G, et al. Diagnosing Tuberculosis: What Do New Technologies Allow Us to (Not) Do? *Respiration* 2022; **101**(9): 797-813.
2. Christian CS, Gerdtham U-G, Hompashe D, Smith A, Burger R. Measuring quality gaps in TB screening in South Africa using standardised patient analysis. *International journal of environmental research and public health* 2018; **15**(4): 729.
3. Yoon C, Dowdy DW, Esmail H, MacPherson P, Schumacher SG. Screening for tuberculosis: time to move beyond symptoms. *The Lancet Respiratory Medicine* 2019; **7**(3): 202-4.
4. Wong EB. It is time to focus on asymptomatic tuberculosis. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 2021; **72**(12): e1044.
5. World Health Organization. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. 2014 (Accessed December 2019).
6. Balcha T, Skogmar S, Sturegård 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 infectious diseases; 2014: Oxford University Press; 2014.
7. 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* 2012; **16**(9): 1252-9.
8. 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 infectious diseases; 2019: Oxford University Press US; 2019. p. ofz233.
9. Shapiro AE, Ting H, 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.
10. 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. *The Lancet Infectious diseases* 2017; **17**(12): 1285-92.
11. Calderwood CJ, Reeve BW, Mann T, et al. Clinical utility of C-reactive protein-based triage for presumptive pulmonary tuberculosis in South African adults. 2023; **86**(1): 24-32.
12. 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 infectious diseases; 2020: Oxford University Press US; 2020. p. ofaa260.
13. 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. *The Lancet Infectious Diseases* 2017; **17**(12): 1285-92.
14. Gelaw Y, Getaneh Z, Melku MJEH, Medicine P. Anemia as a risk factor for tuberculosis: a systematic review and meta-analysis. 2021; **26**(1): 1-15.
15. 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. *The Lancet Infectious Diseases*.
16. World Health Organization. WHO consolidated guidelines on tuberculosis: module 2: screening: systematic screening for tuberculosis disease. 2021.
17. Chakravorty S, Simmons AM, Rowneki M, et al. The new Xpert MTB/RIF Ultra: improving detection of Mycobacterium tuberculosis and resistance to rifampin in an assay suitable for point-of-care testing. *mBio* 2017; **8**(4): e00812-17.
18. Peter JG, Theron G, Singh N, Singh A, Dheda K. Sputum induction to aid diagnosis of smear-negative or sputum-scarce tuberculosis in adults in HIV-endemic settings. *The European respiratory journal* 2014; **43**(1): 185-94.
19. Nathavitharana RR, Garcia-Basteiro AL, Ruhwald M, Cobelens F, Theron G. Reimagining the status quo: How close are we to rapid sputum-free tuberculosis diagnostics for all? *EBioMedicine* 2022: 103939.

20. Provincial Government of the Western Cape. The Western Cape consolidated guidelines for HIV treatment: prevention of mother-to-child transmission of HIV (PMTCT), children, adolescents and adults (amended version 2018). Cape Town, South Africa: Department of Health; 2020.
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**(2): 377-81.
22. Pillay S, de Vos M, Derendinger B, et al. Non-actionable results, accuracy and effect of the first-and second-line line probe assays for diagnosing drug resistant tuberculosis, including on smear-negative specimens, in a high-volume laboratory. 2022.
23. Mishra H, Reeve BW, 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. 2020; **8**(4): 368-82.
24. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ open* 2016; **6**(11): e012799.
25. Ruperez M, Shanaube K, Mureithi L, et al. Use of point-of-care C-reactive protein testing for screening of tuberculosis in the community in high-burden settings: a prospective, cross-sectional study in Zambia and South Africa. 2023; **11**(5): e704-e14.
26. Martinson NA, Nonyane BA, Genade LP, et al. A cluster randomized trial of systematic targeted universal testing for tuberculosis in primary care clinics of South Africa (The TUTT Study). 2022.
27. World Health Organization. Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV: policy update 2019: World Health Organization, 2019.
28. Peter JG, Theron G, Pooran A, Thomas J, Pascoe M, Dheda KJTLRM. Comparison of two methods for acquisition of sputum samples for diagnosis of suspected tuberculosis in smear-negative or sputum-scarce people: a randomised controlled trial. 2013; **1**(6): 471-8.