THE LANCET
Global Health

Supplementary appendix 1

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Reeve BWP, Ndlangalavu G, Mishra H, et al. Point-of-care C-reactive protein and Xpert MTB/RIF Ultra for tuberculosis screening and diagnosis in unselected antiretroviral therapy initiators: a prospective, cross-sectional, diagnostic accuracy study. *Lancet Glob Health* 2024; published online April 4. https://doi.org/10.1016/S2214-109X(24)00052-4.

Appendix

Table of Contents:

Methods

Setting

The study was done at Kraaifontein Community Health Centre, a secondary level health facility.

Definitions

Time-to-positivity

Days for a culture to become *Mycobacterium tuberculosis* complex (*Mtb*) positive in the absence of contamination. If both cultures were positive, the shorter time-to-positivity (TTP) was used.

Ultra "trace reclassified" and "trace excluded"

"Trace reclassified" treats Ultra-positives with a trace result as Ultra-negative. "Trace excluded" excludes participants with trace results from a particular analysis.

Non-actionable result

Any result that is not positive or negative and hence does not provide potentially actionable information to aid clinical decision making. Actionable results are therefore positive or negative or, for a drug susceptibility test, resistant or susceptible.

Specimen processing and diagnostic testing

C-reactive protein

iChromaII has a detection range of 2.5-300 mg/l[1]. Values reported by the machine as <2.5 mg/l (n=113) were classified as index test negative. For continuous comparisons involving the n=113, a midpoint value of 1.25 mg/l was arbitrarily assigned. C-reactive protein (CRP) testing was not done for 62 participants (no blood collected) enrolled in the study. A further 150

participants were not tested prospectively due to reagent shortages, but plasma stored at -80°C was tested retrospectively.

Sputum Xpert and Ultra

700 µl of one sputum sediment remaining post-culture inoculation was, following storage at -20°C for a maximum of two weeks, arbitraily selected for Xpert[2] [sediments ≤0.7ml were made up to volume with phosphate buffered saline (Sigma Aldrich, Modderfontein, South Africa)]. The sediment used for Xpert was arbitrarily selected by the technician, who had no knowledge of which sediment was, per Figure 1, from sputum 1 or 2. Ultra (version 4) and Xpert (version 2) were done per manufacturer specifications[3,4] (Cepheid, United States) and, if either result was non-actionable (see above **Definitions**), specimen-sample reagent mixes were re-tested (if insufficient volume existed the initial result was reported).

Speciation and drug susceptibility testing

MTBDR*plus* (Hain Lifesciences, Nehren, Germany) was done on Auramine-O acid fast-positive culture growth (when two cultures were positive, the one with the shortest TTP was tested) for *Mtb*, and rifampicin- and isoniazid-susceptibility detection. MTBDR*plus* on the culture isolate served as a reference standard for rifampicin resistance.

Urine testing

For Ultra, most urine was tested prospectively [98% (716/732); proportions based on urine included in the head-to-head analysis]. 20 ml was tested or, if less was available, the entire volume used [99% (729/732) of participants had at least 20 ml available]. After centrifugation (4000 rpm, 10 min), supernatant was decanted until a ~700 µl pellet-supernatant mixture was retained prior to resuspension in Ultra buffer (Cepheid, Sunnyvale, CA, USA) and testing[4]. If the concentrated result was non-actionable or positive, Ultra was repeated on 700 µl unconcentrated urine. For Determine TB LAM (LF-LAM; Abbott, South Africa),

unconcentrated urine (60 µl) was tested as recommended[5]. 96% (701/732) participants were tested prospectively, and the remainder tested retrospectively due to stock shortages.

Sample sizes

We did post-hoc sample size calculations for sensitivity and specificity by measuring the precision of accuracy estimates (95% CI calculated using the binomial method) as a function of sample size, as recommended by TB diagnostic test evaluation guidance[6,7]. For triage tests (W4SS, CRP10), we tested until we attained at least 95% CI widths on either side of the point estimate of <10% for sensitivity and <5% for specificity. We did sputum testing until we had the same level of precision for each estimate. We did not do sample size calculations for urine tests, which was an exploratory evaluation and included, in addition to LF-LAM, multiple forms of Ultra testing (concentrated, selected unconcentrated testing).

Results

Rifampicin resistance

Ultra

Amongst Ultra-positives in the head-to-head sputum comparison with Xpert, 3% (3/89), 80% (71/89) and 17% (15/89) were Ultra rifampicin-resistant, -susceptible, and -indeterminate (all indeterminates trace), respectively. All three Ultra rifampicin-resistant people were MTBDR*plus* rifampicin-resistant and one MTBDR*plus* rifampicin-resistant person was Ultra rifampicin-susceptible [Ultra resistance sensitivity hence 75% (3/4)]. Of the Ultra rifampicin-susceptibles, 6% (4/71) were culture-negative. Ultra's specificity for rifampicin resistance was therefore 99% (66/67). Of the Ultra rifampicin-indeterminate people, 73% (11/15) were culture-negative and the four culture-positives were all MTBDR*plus* rifampicin-susceptible.

Xpert

Amongst Xpert-positives in the head-to-head sputum comparison with Ultra, 9% (6/65) were Xpert rifampicin-resistant, 88% (57/65) were Xpert rifampicin-susceptible and 3% (2/65) were rifampicin indeterminate. Of the Xpert resistant cases, 83% (5/6) were culture-positive and MTBDR*plus*-resistant [Xpert sensitivity for rifampicin resistance 100% (5/5)]. Ultra detected 50% (3/6) of these people as rifampicin resistant (of the others, two were Ultra-negative and the other Ultra rifampicin-susceptible). Of the Xpert rifampicin-susceptibles, 7% (4/57) MTBDR*plus* were culture-negative and, of the remaining culture-positives, 100% (53/53) were MTBDR*plus*-susceptible [Xpert specificity for resistance therefore 100% (53/53)]. Two Xpert rifampicin-indeterminates occurred and both were culture-negative.

Urine testing

Sensitivity and specificity

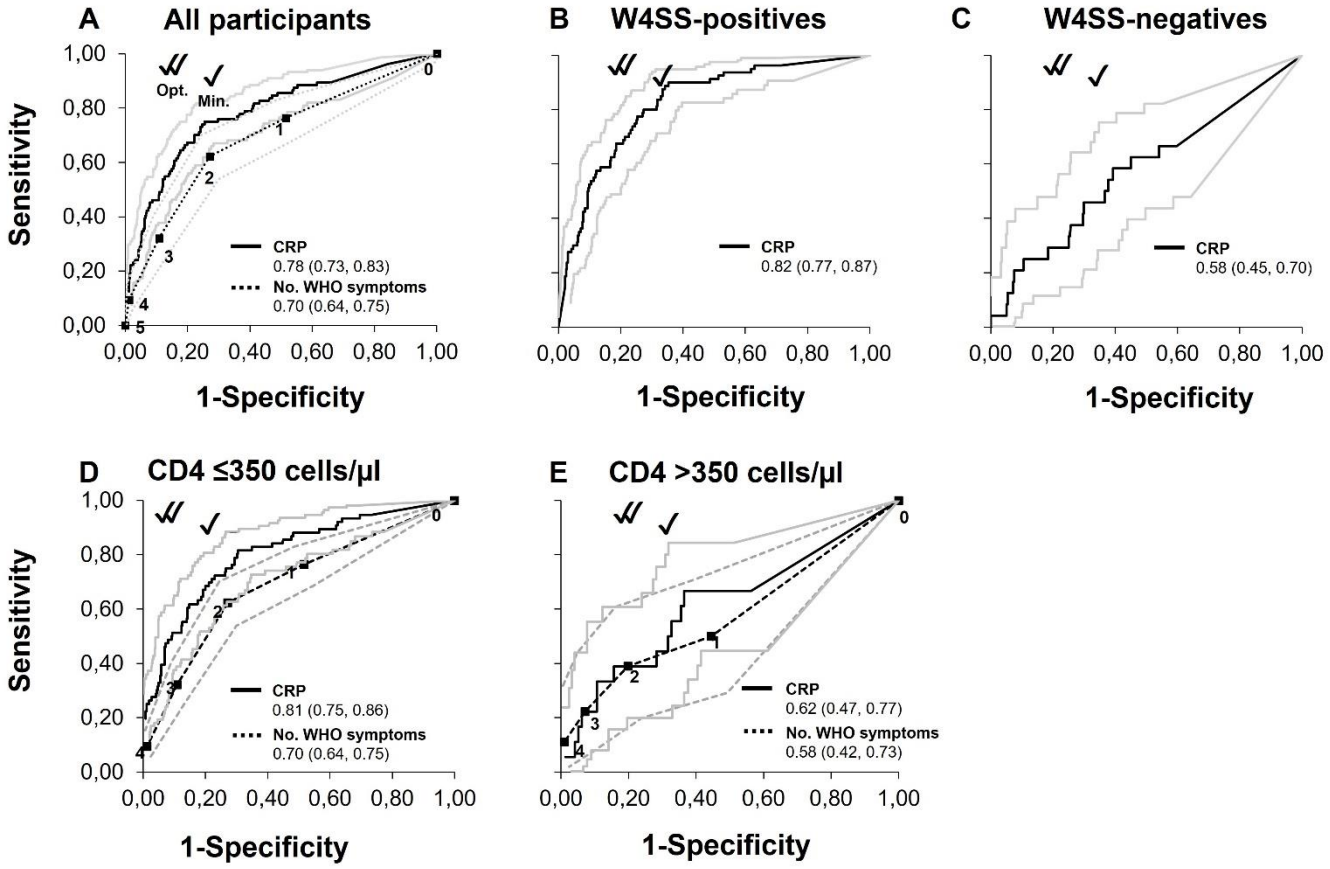
No differences were observed between tests when stratified by W4SS-positive or -negatives nor between tests after W4SS stratification. Similar trends occurred for non-head-to-head data.

Concentrated vs. unconcentrated Ultra

Concentrated Ultra initially non-actionable: The unconcentrated urine Ultra-positivity rate in people with a non-actionable result from concentrated urine was more than that in people whose concentrated urine was positive [11% (13/118) vs. 5% (33/732); p=0.001], indicating that, in the event of non-actionable concentrated urine Ultra, that participant has increased odds of Ultra-positivity on a corresponding unconcentrated urine **(Appendix Figure 3A)**.

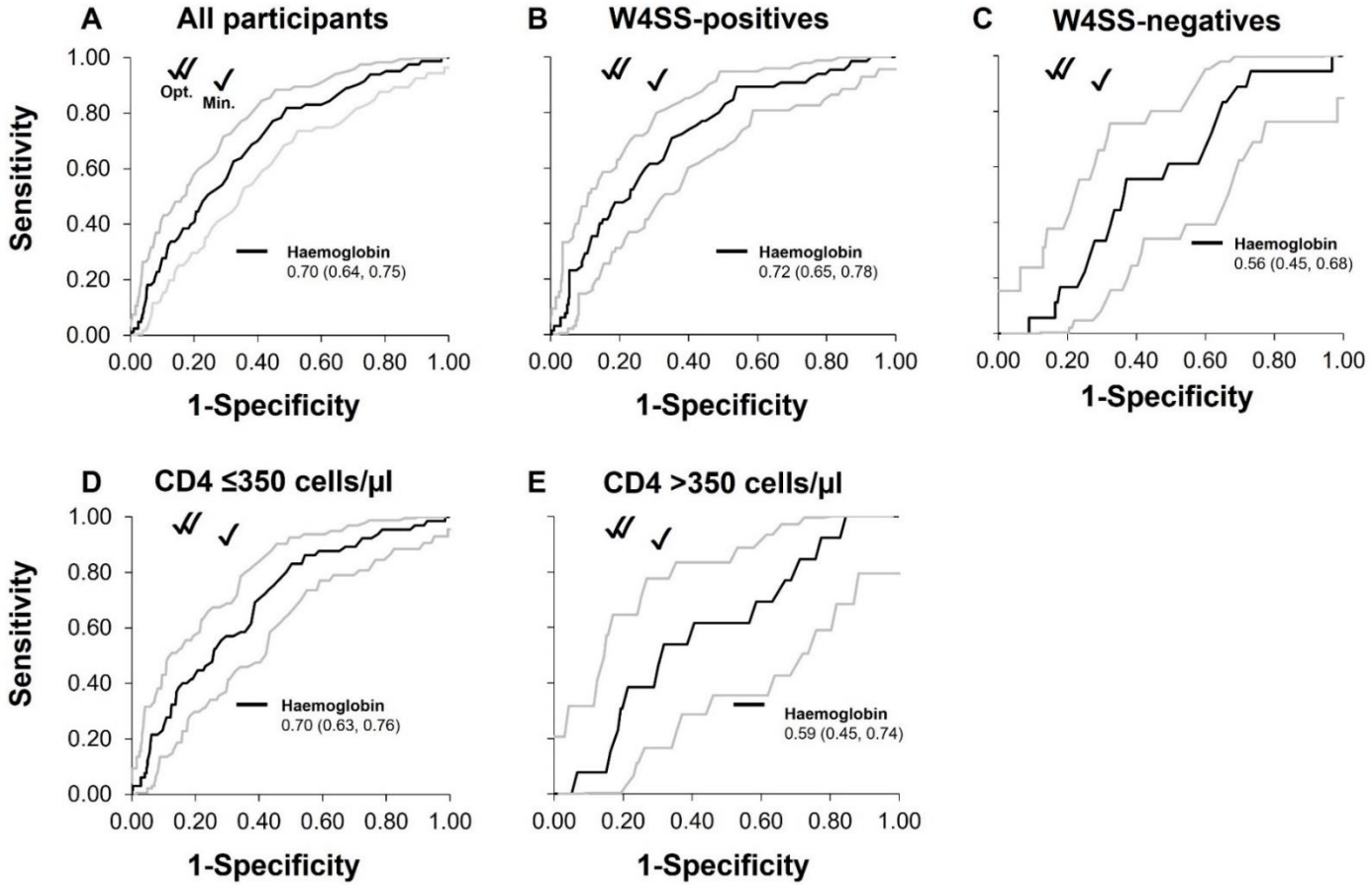
Concentrated Ultra initially positive: 890/897 urines were concentrated for Ultra testing. Of these, 772 had actionable results (13% [118/890] were non-actionable), and 4% (33/772) had a positive result [27% (9/33) negative when unconcentrated urine was used, indicating concentration increases yield] (**Appendix Figure 3B**, remainder of unconcentrated results negative). 100% (33/33) of the concentrated Ultra-positives received unconcentrated Ultra: 97% (32/33) were actionable (one non-actionable) and 72% (23/32) were positive [83% (19/23) rifampicin-susceptible, 4% (1/23) rifampicin-resistant, 9% (2/23) rifampicin-indeterminate, 4% (1/23) rifampicin "no result"].

Appendix Figure 1. 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 AUROC in W4SS-positives than -negatives and in patients with lower CD4 cell counts than higher (as did W4SS). CRP was the only test for which Cis overlapped with the WHO TPP minimal sensitivity and specificity target, and this was only in people with advanced disease (W4SS-positive or low CD4 cell 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.



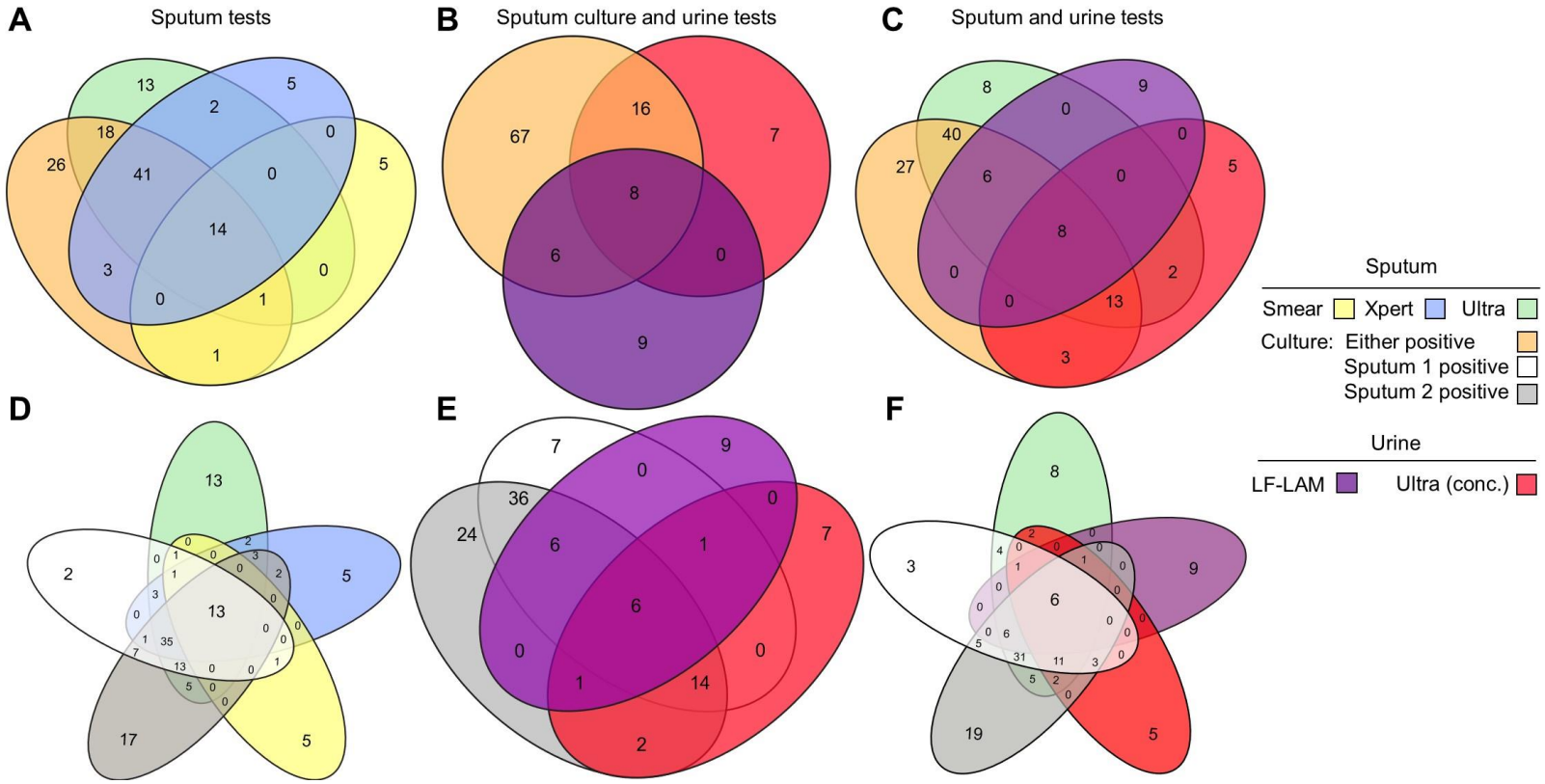
Abbreviations: AUROC, area under the ROC curve; Cis, confidence intervals; CRP, C-reactive protein; Min., minimum; Opt., optimal WHO TPP; ROC, receiver operator characteristics; WHO, World Health Organization; WHO TPP, WHO target product profile; W4SS, WHO-recommended four-symptom screen.

Appendix Figure 2. ROC curves of haemoglobin among all participants (**A**) and after W4SS (**B, C**) or CD4 count stratification (**D, E**). AUCROCs showed a trend towards being reduced in W4SS-negatives (vs. -positives) and in people with a CD4 count >350 cells/µl (vs. ≤350 cells/µl), but no significant differences were detected. Minimal and optimal sensitivity and specificity for the triage test TPP are represented by single and double ticks, respectively. Grey lines are 95% CIs.



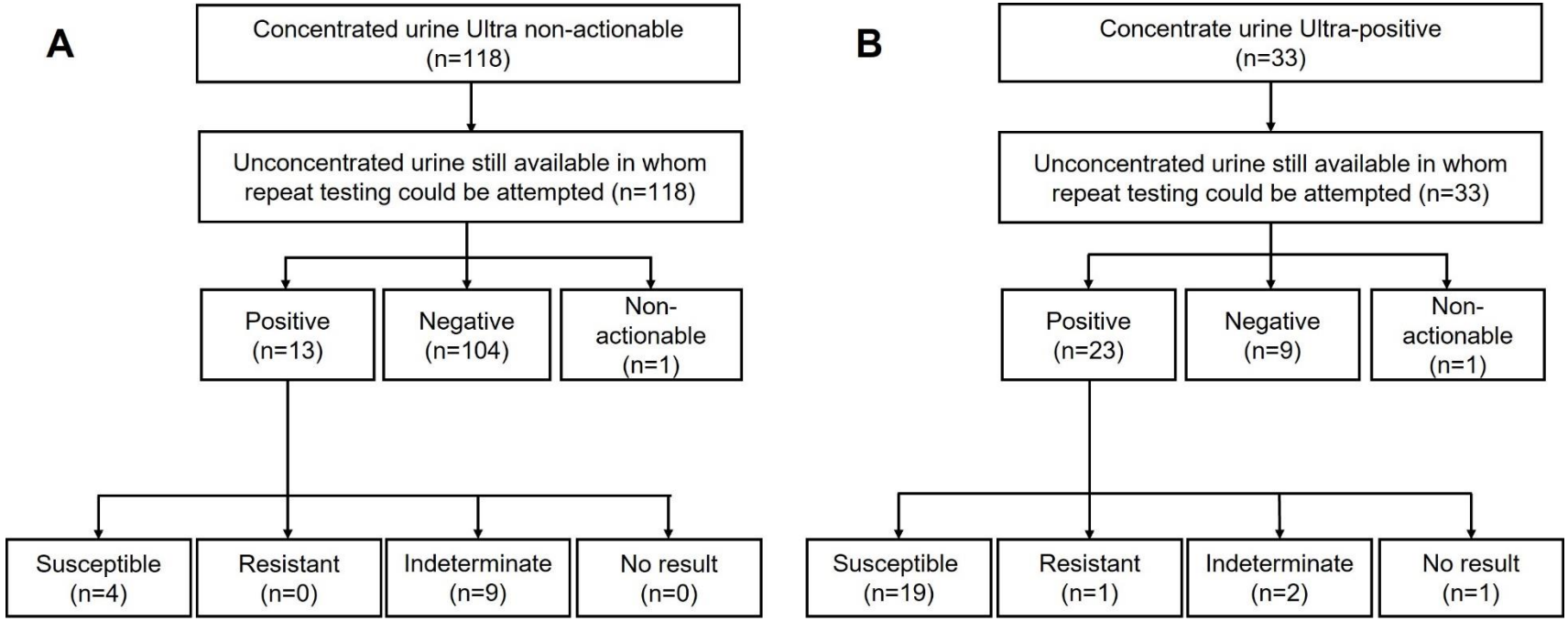
Abbreviations: AUCROC, area under the receiver operator characteristic curve; CI, confidence interval; Min., minimal; Opt., optimal; ROC, receiver operator characteristic; TPP, target product profile; W4SS, World Health Organization-recommended four-symptom screen; WHO, World Health Organization.

Appendix Figure 3. Euler diagrams showing overlap between sputum- and non-sputum tests against composite (**A-C**) and individual (**D-F**) culture results. Numbers represent people positive for each test on sputum (first column), urine (with the sputum reference standard; second column), or sputum and urine (third column). (**A**) Sputum Ultra detected most culture-positives missed by Xpert and, in 10% (13/129) of people with a positive sputum test result, was the only positive result. (**B**) Urine Ultra and LF-LAM had poor yield and missed many cases but detected people missed by other tests (including culture). (**C**) When comparing tests on sputum and urine, each test exclusively detected some people missed by other tests. **D-F** show many people were positive only by a single culture, and a two-culture reference standard's importance. Note people require a positive or negative result to be included here, which is the why total n varies slightly across panels and versus yield calculations (**Results).**



Abbreviations: conc., concentrated. LF-LAM, Determine TB LAM Ag test; MGIT, Mycobacteria Growth Indicator Tube; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Appendix Figure 4. Unconcentrated urine Ultra results in participants who had, on a paired concentrated urine, a (**A**) non-actionable or (**B**) positive result. 27% (9/33) of concentrated Ultra-positive urines were negative when unconcentrated urine was tested, indicating concentration increases yield. Of the concentrated urine non-actionable results, 99% (117/118) resolved when unconcentrated urine was tested, and 11% (13/117) of these were Ultra-positive upon retesting (for two people, this unconcentrated Ultra result was the only positive confirmatory test result). The unconcentrated urine Ultra-positivity rate in people with a non-actionable result from concentrated urine was more than double that in people whose concentrated urine was positive [11% (13/118) vs. 5% (33/732); p=0.001], indicating that, in the event of non-actionable concentrated urine Ultra, the participant has increased odds of Ultra-positivity on a corresponding unconcentrated urine.



Abbreviations: RIF, rifampicin; Ultra, Xpert MTB/RIF Ultra.

Appendix Table 1. Head-to-head diagnostic accuracy of cough and CRP stratified by W4SS status, shown overall and in smear-negative people only. CRP had decreased sensitivity in W4SS-negatives vs. -positives but higher specificity. When comparing CRP thresholds within people of the same W4SS status, ≥10 mg/l had higher specificity than ≥5 mg/l. Trends were similar in smear-negative people. The "All" column (for the "irrespective of smear status" section) is the same as in **Table 2**. Data are %, 95% CI, and n/N.

	All (n=800)				**W4SS-positives (n=442)**				**W4SS-negatives (n=358)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Irrespective of smear status												
Cough (any duration)	53 (43, 63) 55/104	73 (70, 76) 506/696	22 (18, 29) 54/226	91 (89, 94) 524/574	69 (58, 79) 55/80	48 (43, 53) 172/362	22 (18, 29) 55/245	87 (82, 92) 172/197	0 (0, 15) 0/24 **p<0.0001****	100 (99, 100) 334/334 **p<0.0001****	Non-calculable -	93 (91, 96) 334/358 **p=0.017****
≥2 weeks	42 (33, 53) 44/104 p=0.13*	83 (81, 86) 579/696 **p<0.0001***	27 (21, 35) 44/161 p=0.26*	91 (89, 93) 579/639 p=0.74*	55 (44, 67) 44/80 p=0.073*	68 (63, 73) 245/362 **p<0.0001***	27 (21, 35) 44/161 p=0.26*	87 (83, 91) 245/281 p=0.97*	0 (0, 15) 0/24 **p<0.0001**** -	100 (99, 100) 334/334 **p<0.0001**** -	Non-calculable - -	93 (91, 96) 334/358 **p=0.0086**** p>0.99*
CRP ≥5 mg/l	86 (78, 92) 89/104 **p<0.0001**†	49 (45, 53) 339/696 **p<0.0001**†	20 (17, 24) 89/446 p=0.44†	96 (94, 98) 339/354 **p=0.0095**†	94 (87, 98) 75/80 **p<0.0001**†	40 (35, 45) 143/362 **p=0.030**†	26 (21, 31) 75/294 p=0.41†	97 (93, 99) 143/148 p=0.97†	58 (37, 78) 14/24 **p<0.0001**** **p<0.0001**†	59 (54, 65) 196/334 **p<0.0001**** **p<0.0001**†	9 (6, 15) 14/152 **p<0.0001**** -	95 (92, 98) 196/206 p=0.49** p=0.37†
≥10 mg/l	77 (68, 85) 80/104 **p<0.0001**† p=0.39‡	64 (61, 68) 445/696 **p<0.0001**† **p<0.0001**‡	24 (20, 30) 80/331 p=0.63† p=0.16‡	95 (93, 97) 445/469 **p=0.021**† p=0.56‡	90 (82, 96) 72/80 **p=0.001**† p=0.39‡	54 (49, 59) 194/362 p=0.10† **p<0.0001**‡	30 (25, 37) 72/240 p=0.059† p=0.25‡	96 (93, 99) 194/202 **p=0.0015**† p=0.78‡	33 (16, 56) 8/24 **p<0.0001**** **p=0.0019**† p=0.082‡	75 (71, 80) 251/334 **p<0.0001**** **p<0.0001**† **p<0.0001**‡	9 (4, 17) 8/91 **p<0.0001**** - p=0.91‡	94 (91, 97) 251/267 p=0.32** p=0.72† p=0.59‡
Smear-negatives only												
Cough (any duration)	49 (39, 60) 43/88	73 (70, 77) 504/691	19 (14, 25) 43/230	92 (90, 95) 504/549	64 (52, 76) 43/67	4 (43, 53) 170/357	19 (14, 25) 43/230	88 (83, 92) 170/194	0 (0, 17) 0/21 **p<0.0001****	100 (99, 100) 334/334 **p<0.0001****	Non-calculable -	94 (92, 97) 334/355 **p=0.0084****
≥2 weeks	38 (28, 49) 33/88 p=0.13*	84 (81, 87) 576/691 **p<0.0001***	22 (16, 30) 33/148 p=0.39*	91 (89, 94) 576/631 p=0.75*	49 (37, 62) 33/67 p=0.081*	68 (63, 73) 242/357 **p<0.0001***	22 (16, 30) 33/148 p=0.39*	88 (84, 92) 242/276 p=0.99*	0 (0, 17) 0/21 **p<0.0001**** -	100 (99, 100) 334/334 **p<0.0001**** -	Non-calculable - -	94 (92, 97) 334/355 **p=0.0047**** p>0.99*
CRP ≥5 mg/l	84 (75, 92) 74/88	49 (46, 54) 338/691	17 (14, 22) 74/427	96 (94, 98) 338/352	93 (84, 98) 62/67	40 (35, 46) 142/357	22 (18, 28) 62/277	97 (93, 99) 142/147	57 (35, 79) 12/21 **p<0.0001****	59 (54, 65) 196/334 **p<0.0001****	8 (5, 14) 12/150 **p<0.0001****	96 (92, 98) 196/205 p=0.64**

	p<0.0001[†]	**p<0.0001**[†]	p=0.66[†]	**p=0.012**[†]	**p<0.0001**[†]	**p<0.0001**[†]	p=0.25[†]	**p=0.003**[†]	**p<0.0001**[†]	**p<0.0001**[†]	-	p=0.44[†]
≥10 mg/l	74 (64, 83) 65/88 **p<0.0001**[†] p=0.096[‡]	64 (61, 68) 444/691 **p<0.0001**[†] **p<0.0001**[‡]	21 (17, 26) 65/312 p=0.54[†] p=0.23[‡]	95 (93, 97) 444/467 **p=0.038**[†] p=0.52[‡]	88 (78, 95) 59/67 **p=0.001**[†] p=0.38[‡]	54 (49, 60) 193/357 p=0.085[†] **p<0.0001**[‡]	26 (21, 33) 59/223 **p=0.048**[†] p=0.29[‡]	96 (93, 99) 193/201 **p=0.0022**[†] p=0.78[‡]	29 (12, 53) 6/21 **p<0.0001**[**] **p=0.008**[†] p=0.061[‡]	75 (71, 80) 251/334 **p<0.0001**[**] **p<0.0001**[†] **p<0.0001**[‡]	7 (3, 15) 6/89 **p<0.0001**[**] - p=0.72[‡]	94 (91, 97) 251/266 p=0.41[**] p=0.88[†] p=0.51[‡]

Within row p-values: [**]W4SS-positive vs. -negative.
Within column p-values: [*]Any vs. ≥2-week cough; [†]Cough (any) vs. CRP; [‡]CRP (≥5 vs ≥10 mg/l).
Abbreviations: CRP, C-reactive protein; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; W4SS, WHO-recommended four-symptom screen.

Appendix Table 2. Head-to-head diagnostic accuracy of W4SS and CRP stratified by CD4 count in smear-negative people. Trends were like those in all people irrespective of smear status (**Table 2**). Data are %, 95% CI, and n/N.

	All (n=779)				**CD4 count >350 cells/µl (n=309)**				**CD4 count ≤350 cells/µl (n=426)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Cough (any duration)	49 (39, 60) 43/88	73 (70, 77) 504/691	19 (14, 25) 43/230	92 (90, 95) 504/549	40 (17, 68) 6/15	80 (75, 85) 235/294	9 (4, 20) 6/65	96 (94, 99) 235/244	50 (38, 63) 33/66 p=0.48**	68 (63, 73) 243/360 **p<0.0001****	22 (16, 30) 33/150 **p=0.026****	88 (84, 92) 243/276 **p=0.0010****
≥2 weeks	38 (28, 49) 33/88 p=0.13*	84 (81, 87) 576/691 **p<0.0001***	22 (16, 30) 33/148 p=0.39*	91 (89, 94) 576/631 p=0.75*	27 (8, 56) 4/15 p=0.44*	88 (84, 92) 258/294 p=0.010*	10 (3, 24) 4/40 p=0.89*	50 (93, 98) 258/269 p=0.815*	39 (28, 53) 26/66 p=0.36** p=0.22*	81 (77, 85) 291/360 **p=0.017**** **p<0.0001***	27 (19, 38) 26/95 **p=0.027**** p=0.34*	88 (84, 92) 291/331 **p=0.0010**** p=0.96*
W4SS	76 (66, 85) 67/88	48 (46, 53) 334/691	16 (13, 20) 67/424	94 (92, 97) 334/355	47 (22, 74) 7/15	54 (49, 61) 160/294	5 (3, 10) 7/141	95 (91, 98) 160/168	82 (71, 91) 54/66 **p=0.0044****	45 (40, 51) 161/360 **p=0.014****	21 (17, 27) 54/253 **p<0.0001****	93 (89, 97) 161/173 p=0.39**
CRP ≥5 mg/l	84 (75, 92) 74/88 p=0.19†	49 (46, 54) 338/691 p=0.83†	17 (14, 22) 74/427 p=0.55†	96 (94, 98) 338/352 p=0.24†	71 (45, 90) 10/15 p=0.27†	60 (55, 66) 176/294 p=0.18†	9 (5, 16) 10/128 p=0.34†	97 (94, 100) 176/181 p=0.324†	88 (78, 95) 58/66 **p=0.043**** p=0.33†	41 (36, 47) 147/360 **p=0.001**** p=0.29†	21 (17, 27) 58/271 **p=0.001**** p=0.99†	95 (91, 98) 147/155 p=0.26** p=0.50†
≥10 mg/l	74 (64, 83) 65/88 p=0.73† p=0.096‡	64 (61, 68) 444/691 **p<0.0001†** **p<0.0001‡**	21 (17, 26) 65/312 p=0.079† p=0.23‡	95 (93, 97) 444/467 p=0.53† p=0.52‡	33 (12, 62) 5/15 p=0.46† p=0.068‡	72 (67, 77) 211/294 **p<0.0001†** **p=0.0023‡**	6 (2, 13) 5/88 p=0.81† p=0.55‡	95 (92, 98) 211/221 p=0.912† p=0.354‡	82 (71, 91) 54/66 **p<0.0001**** p>0.99† p=0.33‡	59 (54, 65) 213/360 **p=0.0010**** **p<0.0001†** **p<0.0001‡**	27 (21, 34) 54/201 **p<0.0001**** p=0.17† p=0.17‡	95 (91, 98) 213/225 p=0.69** p=0.50† p=0.94‡

Within row p-values: **CD4 count >350 vs. ≤350 cells/µl.
Within column p-values: *Any vs. ≥2-week cough; †W4SS vs. CRP; ‡CRP (≥5 vs. ≥10 mg/l).
Missing data: CD4 within three months (n=44).
Abbreviations: CRP, C-reactive protein; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; W4SS, WHO-recommended four-symptom screen.

Appendix Table 3. Non-head-to-head diagnostic accuracy of cough and CRP stratified by W4SS status. Findings in this table are like head-to-head data in **Appendix Table 2**. Data are %, 95% CI, and n/N.

	All				W4SS-positives				W4SS-negatives			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Irrespective of smear status												
Cough (any duration)	52 (43, 63) 56/107	73 (70, 76) 544/748	22 17, 28) 56/260	91 (89, 94) 544/595	65 (53, 76) 44/68	47 (43, 53) 180/381	18 (14, 24) 44/245	88 (84, 93) 180/204	0 (0, 15) 0/23 **p<0.0001****	100 (99, 100) 362/362 **p<0.0001****	Non-calculable -	94 (92, 97) 362/385 **p=0.014****
≥2 weeks	42 (33, 52) 45/107 p=0.13*	83 (81, 86) 621/748 **p<0.0001***	26 (20, 34) 45/172 p=0.27*	91 (89, 93) 621/683 p=0.75*	50 (38, 63) 34/68 p=0.083*	67 (63, 72) 256/381 **p<0.0001***	21 (16, 29) 34/159 p=0.39*	88 (85, 92) 256/290 p=0.99*	0 (0, 15) 0/23 **p<0.0001**** -	100 (99, 100) 362/362 **p<0.0001**** -	Non-calculable - -	94 (92, 97) 362/385 p=0.0078* * -
CRP ≥5 mg/l	86 (78, 92) 89/104 **p<0.0001**†	49 (45, 53) 339/696 **p<0.0001**†	20 (17, 24) 89/446 p=0.62†	96 (94, 98) 339/354 **p=0.011**†	93 (84, 98) 62/67 **p<0.0001**†	40 (35, 46) 142/357 **p=0.041**†	22 (18, 28) 62/277 p=0.21†	97 (93, 99) 142/147 **p=0.0050**†	57 (35, 79) 12/21 **p<0.0001**** **p<0.0001**†	59 (54, 65) 196/334 **p<0.0001**** **p<0.0001**†	8 (5, 14) 12/150 **p<0.0001**** -	96 (92, 98) 196/205 p=0.64** p=0.42†
≥10 mg/l	77 (68, 85) 80/104 **p<0.0001**†	64 (61, 68) 445/696 **p<0.0001**†	24 (20, 30) 80/331 p=0.45†	95 (93, 97) 445/469 **p=0.029**†	88 (78, 95) 59/67 **p=0.001**†	54 (49, 60) 193/357 p=0.064†	26 (21, 33) 59/223 **p=0.027**†	96 (93, 99) 193/201 **p=0.0037**†	29 (12, 53) 6/21 **p<0.0001**** **p=0.0058**†	75 (71, 80) 251/334 **p<0.0001**** **p<0.0001**†	7 (3, 15) 6/89 **p<0.0001**** -	94 (91, 97) 251/266 p=0.41** p=0.86†
Smear-negatives only												
Cough (any duration)	48 (38, 60) 44/91	73 (70, 77) 542/743	18 (14, 24) 44/245	92 (90, 95) 542/589	65 (53, 76) 44/68	47 (43, 53) 180/381	18 (14, 24) 44/245	88 (84, 93) 180/204	0 (0, 15) 0/23 **p<0.0001****	100 (99, 100) 362/362 **p<0.0001****	Non-calculable -	94 (92, 97) 362/385 **p=0.014****
≥2 weeks	37 (28, 49) 34/91 p=0.13*	83 (81, 86) 618/743 **p<0.0001***	21 (16, 29) 34/159 p=0.39*	92 (90, 94) 618/675 p=0.76*	50 (38, 63) 34/68 p=0.083*	67 (63, 72) 256/381 **p<0.0001***	21 (16, 29) 34/159 p=0.39*	88 (85, 92) 256/290 p=0.99*	0 (0, 15) 0/23 **p<0.0001**** -	100 (99, 100) 362/362 **p<0.0001**** -	Non-calculable - -	94 (92, 97) 362/385 **p=0.0078*** * p>0.99*
CRP ≥5 mg/l	84 (75, 92) 74/88 **p<0.0001**†	49 (46, 53) 338/691 **p<0.0001**†	17 (14, 22) 74/427 p=0.84†	96 (94, 98) 338/352 **p=0.016**†	93 (84, 98) 62/67 **p<0.0001**†	40 (35, 46) 142/357 **p<0.0001**†	22 (18, 28) 62/277 p=0.21†	97 (93, 99) 142/147 **p=0.0050**†	57 (35, 79) 12/21 **p<0.0001**** **p<0.0001**†	59 (54, 65) 196/334 **p<0.0001**** **p<0.0001**†	8 (5, 14) 12/150 **p<0.0001**** -	96 (92, 98) 196/205 p=0.64** p=0.42†
≥10 mg/l	74	64	21	95	88	54	26	96	29	75	7	94

	(64, 83) 65/88 **p<0.0001**[†]	(61, 68) 444/691 **p<0.0001**[†]	(17, 26) 65/312 p=0.39[†]	(93, 97) 444/467 **p=0.048**[†]	(78, 95) 59/67 **p=0.0014**[†]	(49, 60) 193/357 p=0.064[†]	(21, 33) 59/223 **p=0.027**[†]	(93, 99) 193/201 **p=0.0037**[†]	(12, 53) 6/21 **p<0.0001**[**] **p=0.0058**[†]	(71, 80) 251/334 **p<0.0001**[**] **p<0.0001**[†]	(3, 15) 6/89 **p<0.0001**[**] -	(91, 97) 251/266 p=0.41[**] p=0.86[†]

Within row p-values: [**]W4SS-positives vs -negatives.

Within column p-values: [*]Any vs. ≥2-week cough; [†]Cough (any) vs. CRP.

Abbreviations: CRP, C-reactive protein; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; W4SS, WHO-recommended four-symptom screen.

Appendix Table 4. Non-head-to-head diagnostic accuracy data of potential triage tests ("all" is the same as in **Appendix Table 2**). Diagnostic accuracy was like the head-to-head analyses done overall (**Table 2**) and in smear-negatives (**Appendix Table 2**). Data are %, 95% CI, and n/N.

	All				**CD4 count >350 cells/µl**				**CD4 count ≤350 cells/µl**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Irrespective of smear status												
Cough (any duration)	52 (43, 63) 56/107	73 (70, 76) 544/748	22 17, 28) 56/260	91 (89, 94) 544/595	39 (18, 65) 7/18	80 (75, 84) 246/309	10 (5, 20) 7/70	96 (93, 98) 246/257	53 (42, 65) 42/79 p=0.27**	68 (63, 73) 267/395 **p=0.0004****	25 (19, 32) 42/170 **p=0.010****	88 (84, 92) 267/304 **p=0.0010****
≥2 weeks	42 (33, 52) 45/107 p=0.13*	83 (81, 86) 621/748 **p<0.0001***	26 (20, 34) 45/172 p=0.27*	91 (89, 93) 621/683 p=0.75*	28 (10, 54) 5/18 p=0.48*	88 (84, 92) 271/309 **p<0.0065***	12 (4, 26) 5/43 p=0.789*	95 (93, 98) 271/284 p=0.87*	43 (32, 55) 34/79 p=0.23** p=0.20*	80 (76, 85) 317/395 **p=0.0082**** **p<0.0001***	30 (23, 40) 34/112 **p=0.016**** p=0.29*	88 (84, 91) 317/362 **p=0.0010**** p=0.92*
W4SS	76 (67, 84) 81/107	48 (45, 53) 362/748	17 (15, 22) 81/467	93 (91, 96) 362/388	50 (27, 74) 9/18	55 (50, 61) 171/309	6 (3, 12) 9/147	95 (91, 98) 171/180	80 (70, 88) 63/79 **p=0.0092****	44 (40, 50) 174/395 **p=0.0029****	22 (18, 28) 63/284 **p<0.0001****	92 (87, 96) 174/190 **p=0.19****
CRP ≥5mg/l	86 (78, 92) 89/104 p=0.070†	49 (45, 53) 339/696 p=0.91†	20 (17, 24) 89/446 p=0.31†	96 (94, 98) 339/354 p=0.14†	67 (41, 87) 12/18 p=0.31†	60 (55, 66) 176/294 p=0.26†	9 (5, 16) 12/130 p=0.33†	97 (93, 99) 176/182 p=0.42†	89 (81, 96) 68/76 **p=0.015**** p=0.094†	41 (36, 46) 148/365 **p<0.0001**** p=0.33†	24 (20, 30) 68/285 **p<0.0001**** p=0.65†	95 (91, 98) 148/156 **p=0.004**** p=0.23†
≥10mg/l	77 (68, 85) 80/104 p=0.84†	64 (61, 68) 445/696 **p<0.0001†**	24 (20, 30) 80/331 **p=0.018†**	95 (93, 97) 445/469 p=0.33†	39 (18, 65) 7/18 p=0.50†	72 (67, 77) 211/294 **p<0.0001†**	8 (4, 16) 7/90 p=0.62†	95 (92, 98) 211/222 p=0.98†	84 (75, 92) 64/76 **p<0.0001**** p=0.47†	59 (54, 64) 214/365 **p<0.0001**** **p<0.0001†**	30 (24, 37) 64/215 **p<0.0001**** p=0.054†	95 (91, 98) 214/226 p=0.87** p=0.21†
Smear-negatives												
Cough (any duration)	48 (38, 60) 44/91	73 (70, 77) 542/743	18 (14, 24) 44/245	92 (90, 95) 542/589	40 (17, 68) 6/15	80 (75, 84) 246/309	9 (4, 18) 6/69	96 (94, 99) 246/255	49 (90, 100) 34/69 p=0.52**	68 (64, 73) 265/390 **p=0.0010****	21 (16, 29) 34/159 **p=0.021****	88 (85, 92) 265/300 **p<0.0001****
≥2 weeks	37 (28, 49) 34/91 p<0.13*	83 (81, 86) 618/743 **p<0.0001***	21 (16, 29) 34/159 p=0.39*	92 (90, 94) 618/675 **p<0.0001***	27 (8, 56) 4/15 p=0.44*	88 (84, 92) 271/309 **p=0.0065***	10 (3, 23) 4/42 p=0.88*	96 (94, 99) 271/282 p=0.81*	39 (88, 100) 27/69 p=0.367** p=0.23*	81 (77, 85) 314/390 **p=0.011**** **p<0.0001***	26 (19, 36) 27/103 **p=0.026**** p=0.37*	88 (85, 92) 314/356 **p<0.0001**** p=0.96*
W4SS	75 (65, 84) 68/91	49 (46, 53) 362/743	15 (12, 19) 68/449	94 (92, 97) 362/385	47 (22, 74) 7/15	55 (50, 61) 171/309	5 (2, 10) 7/145	96 (92, 99) 171/179	80 (69, 89) 55/69 **p=0.0083****	45 (40, 50) 174/390 **p=0.0049****	20 (16, 26) 55/271 **p<0.0001****	93 (88, 96) 174/188 p=0.23**
CRP ≥5mg/l	84	49	17	96 (94, 98)	67	60	8	97	88	41	21	95

	(75, 92) 74/88 p=0.12†	(46, 53) 338/691 p=0.94†	(14, 22) 74/427 p=0.38†	338/352 p=0.27†	(39, 89) 10/15 p=0.27†	55, 66) 176/294 p=0.26†	(4, 14) 10/128 p=0.31†	(94, 100) 176/181 p=0.39†	(78, 95) 58/66 **p=0.043**** p=0.19†	(36, 47) 147/360 **p<0.0001**** p=0.29†	(17, 27) 58/271 **p=0.001**** p=0.75†	91, 98) 147/155 p=0.26** p=0.39†
≥10mg/l	74 (64, 83) 65/88 p=0.89†	64 (61, 68) 444/691 **p<0.0001**†	21 (17, 26) 65/312 **p=0.042**†	95 (93, 97) 444/467 p=0.50†	33 (12, 62) 5/15 p=0.46†	72 (67, 77) 211/294 **p<0.0001**†	6 (2, 13) 5/88 p=0.78†	95 (92, 98) 211/221 p=0.98†	82 (71, 91) 54/66 **p<0.0001**** p=0.76†	59 (54, 65) 213/360 **p<0.0001**** **p<0.0001**†	27 (21, 34) 54/201 **p<0.0001**** p=0.094†	95 (91, 98) 213/225 p=0.69** p=0.38†

Within row p-values: **CD4 count >350 vs. ≤350 cells/µl.
Within column p-values: *Any vs. ≥2-week cough, †W4SS vs. CRP.
Missing data or not done: no CD4 count within three months (n=76).
Abbreviations: CRP C-reactive protein; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis; W4SS, World Health Organization-recommended four-symptom screen.

Appendix Table 5. Head-to-head diagnostic accuracy of haemoglobin (<10 g/dl) stratified by CD4 count and W4SS statuses. Haemoglobin had increased sensitivity at lower CD4 cells counts but decreased specificity (same trend observed in W4SS-positives vs. negatives). Non-head-to-head data are not shown due to the many patients not receiving haemoglobin testing (it was done programmatically). Data are %, 95% CI, and n/N.

All participants (n=634)				**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**
31	88	27	90	8	91	5	95	37	86	35	86
(22, 43)	(86, 91)	(19, 38)	(87, 93)	(1, 37)	(87, 95)	(1, 24)	(91, 98)	(25, 50)	(81, 90)	(24, 48)	(82, 91)
25/81	487/553	25/91	487/543	1/13	210/230	1/21	210/222	23/63	254/297	23/66	254/294
								p=0.042[*]	**p=0.042**[*]	**p=0.0072**[*]	**p=0.0022**[*]
p<0.0001[†]	**p<0.0001**[†]	**p=0.041**[†]	p=0.062[†]	**p=0.013**[†]	**p<0.0001**	p=0.78[†]	p=0.97[†]	**p<0.0001**[†]	**p<0.0001**[†]	p=0.055[†]	p=0.066[†]
p<0.0001[‡]	**p<0.0001**[‡]	p=0.52[‡]	**p=0.002**[‡]	**p=0.050**[‡]	**p<0.0001**[‡]	p=0.63[‡]	p=0.83[‡]	**p<0.0001**[‡]	**p<0.0001**[‡]	p=0.44[‡]	**p=0.0022**[‡]
				W4SS-positive (n=354)				**W4SS-negative (n=280)**			
				39	86	37	86	0	91	0	93
				(28, 53)	(81, 90)	(26, 50)	(82, 91)	(0, 20)	(87, 95)	(0, 15)	(90, 97)
				25/64	248/290	25/67	248/287	0/17	239/263	0/24	239/256
								p<0.0001[*]	p=0.052[*]	**p<0.0001**[*]	**p=0.0079**[*]
				p<0.0001[‡‡]	**p<0.0001**[‡‡]	p=0.26[‡‡]	**p<0.0001**[‡‡]	**p=0.0080**[‡‡]	**p<0.0001**[‡‡]	p=0.13[‡‡]	p=0.76[‡‡]

Within row p-values: [*]CD4 count >350 vs. ≤350 cells/µl.

Within column p-values: [†]haemoglobin vs. W4SS in **Table 2**, [‡]haemoglobin vs. CRP (≥10mg/l) in **Table 2** (comparisons within a CD4 count stratum); [‡‡]haemoglobin vs. CRP (≥10mg/l) in **Appendix Table 1** (comparisons within a W4SS stratum).

Missing data or not done: CD4 count within three months (n=31); haemoglobin not programmatically done for n=190.

Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; W4SS; World Health Organization-recommended four-symptom screen.

Appendix Table 6. AUROCs, rule-in (~95% specificity), rule-out (~95% sensitivity), and Youden index (maximum sum of sensitivity and specificity) thresholds of W4SS, CRP, and haemoglobin, stratified by W4SS and CD4 count status. CRP had the highest AUROC. AUROCs were increased in W4SS-positives (vs. -negatives) and in people with CD4 counts >350 cells/µl (vs. ≤350). Haemoglobin followed similar trends. At rule-out thresholds, no tests met the WHO TPP minimal specificity point estimate of 70%, irrespective of patient group. Data are % (95% CI).

	W4SS			**CRP**			**Haemoglobin**		
	Overall								
	AUROC: 0.70 (0.64, 0.75)			AUROC: 0.78 (0.73, 0.84)			AUROC: 0.70 (0.64, 0.75)		
	Threshold (number)	Sensitivity	Specificity	Threshold (mg/l)	Sensitivity	Specificity	Threshold (g/dl)	Sensitivity	Specificity
Rule-in	≥3	32 (25, 40)	89 (87, 91)	>130	31 (20, 36)	95 (93, 96)	<9	13 (8, 21)	95 (93, 97)
Rule-out	≥0	100 (97, 100)	0 (0, 0)	>3	89 (83, 94)	39 (35, 42)	<14	95 (89, 98)	20 (17, 23)
Youden's index	≥2	62 (54, 70)	73 (70, 76)	>18	75 (66, 82)	75 (72, 77)	<13	82 (74, 89)	51 (47, 54)
				W4SS-positives					
				AUROC: 0.82 (0.77, 0.87)			AUROC: 0.72 (0.65, 0.78)		
Rule-in				>172	30 (22, 40)	95 (93, 97)	<9	23 (15, 33)	95 (92, 97)
Rule-out				>4	95 (89, 98)	38 (34, 43)	<14	95 (89, 99)	19 (15, 23)
Youden's index				>18	89 (81, 94)	65 (61, 69)	<12	71 (60, 80)	65 (60, 69)
				W4SS-negatives					
				AUROC: 0.58 (0.45, 0.70)			AUROC: 0.56 (0.45, 0.68)		
Rule-in				>72	4 (0, 18)	95 (92, 97)	<9	0 (0, 15)	95 (92, 97)
Rule-out				>3	67 (48, 82)	46 (41, 51)	<14	94 (76, 100)	27 (23, 32)
Youden's index				>5	58 (40, 75)	61 (56, 65)	Same as rule-out		
	CD4 count >350 cells/µl								
	AUROC: 0.58 (0.42, 0.73)			AUROC: 0.62 (0.47, 0.77)			AUROC: 0.59 (0.45, 0.74)		
Rule-in	≥3	93 (90, 95)	22 (8, 44)	>91	95 (92, 97)	17 (5, 38)	<10	0 (0, 21)	95 (92, 97)
Rule-out	≥0	100 (85, 100)	0 (0, 1)	>6	67 (45, 84)	64 (59, 68)	<14	92 (68, 100)	23 (18, 78)
Youden's index	≥2	22 (8, 44)	93 (90, 95)	Same as rule-out			<12	54 (29, 78)	68 (63, 73)
	CD4 count ≤350 cells/µl								
	AUROC: 0.69 (0.63, 0.76)			AUROC: 0.81 (0.75, 0.86)			AUROC: 0.70 (0.63, 0.76)		
Rule-in	≥3	32 (25, 40)	89 (87, 91)	>130	36 (26, 46)	95 (92, 96)	<9	12 (6, 21)	95 (93, 97)
Rule-out	≥0	100 (97, 100)	0 (0, 0)	>3	95 (88, 98)	30 (26, 35)	<14	95 (89, 99)	21 (18, 25)
Youden's index	≥2	62 (54, 70)	73 (70, 76)	>18	82 (73, 89)	70 (65, 74)	<13	83 (74, 90)	50 (45, 54)

Abbreviations: AUROC, area under the receiver operator characteristic curve; CRP, C-reactive protein; CI, confidence interval; W4SS, WHO-recommended four-symptom screen; WHO, World Health Organization; WHO TTP, World Health Organization target product profile.

Appendix Table 7. Head-to-head diagnostic accuracy of sputum Xpert and Ultra for TB stratified by CD4 cell count. Each test had higher sensitivities and similar specificities in people with lower vs. higher CD4 cell counts. Ultra had higher sensitivity than Xpert at CD4 counts ≤350 cells/µl. We did not sub-stratify CD4 cell counts categories further by smear status given low numbers of index test-detected cases. Data are %, 95% CI, and n/N.

	All (n=787)				**CD4 count >350 cells/µl (n=304)**				**CD4 count ≤350 cells/µl (n=432)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**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	18 (4, 44) 3/17	100 (99, 100) 287/287	100 (30, 100) 3/3	96 (93, 98) 287/301	13 (7, 23) 10/77 p=0.68*	99 (97, 100) 350/355 **p=0.044***	67 (39, 89) 10/15 p=0.24*	84 (81, 88) 350/417 **p<0.0001***
Xpert	56 (46, 66) 58/104 **p<0.0001**†	99 (98, 100) 676/683 p=0.56†	89 (79, 96) 58/65 p=0.13†	94 (92, 95) 676/722 **p=0.0010**†	30 (11, 56) 5/17 p=0.42†	99 (97, 100) 284/287 p=0.082†	63 (25, 92) 5/8 p=0.21†	96 (94, 98) 284/296 p=0.72†	60 (48, 71) 46/77 **p=0.023*** **p<0.0001**†	99 (98, 100) 351/355 p=0.92* p=0.74†	92 (81, 98) 46/50 **p=0.017*** **p=0.013**†	92 (89, 95) 351/382 **p=0.031*** **p<0.0001**†
Ultra-negative	10 (2, 27) 3/30	99 (98, 100) 663/668	38 (9, 76) 3/8	96 (94, 97) 663/690	0 (0, 31) 0/10	99 97, 100) 280/283	0 0, 71) 0/3	97 (94, 98) 280/290	17 4, 41) 3/18 p=0.17*	99 (98, 100) 342/344 p=0.50*	60 (15, 95) 3/5 p=0.090*	96 (93, 98) 342/357 p=0.62*
Ultra	71 (61, 80) 74/104 **p=0.014**‡	98 (96, 99) 668/683 p=0.086‡	83 (74, 90) 74/89 p=0.29‡	96 (94, 97) 668/698 p=0.083‡	42 (19, 68) 7/17 p=0.47‡	99 (97, 100) 283/287 p=0.70‡	64 (31, 90) 7/11 p=0.96‡	97 (94, 99) 283/293 p=0.68‡	77 (66, 86) 59/77 **p=0.0038*** **p=0.025**‡	97 (95, 99) 344/355 p=0.16* p=0.068‡	85 (74, 92) 59/70 p=0.10* p=0.21‡	96 (93, 98) 344/362 p=0.33* p=0.084‡
Xpert-negative	41 (27, 57) 19/46	98 (97, 99) 663/676	59 (41, 76) 19/32	96 (94, 97) 663/690	17 (3, 49) 2/12	99 (97, 100) 280/284	34 (5, 78) 2/6	97 (94, 99) 280/290	5 (34, 70) 16/31 **p=0.037***	98 (96, 99) 342/351 p=0.31*	64 (43, 83) 16/25 p=0.17*	96 (94, 98) 342/357 p=0.62*

Within row p-values: *CD4 count ≤350 vs. >350 cells/µl.
Within column p-values: †smear vs. Xpert; ‡Xpert vs. Ultra.
Missing data or not done: CD4 count (n=74).
Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Appendix Table 8. Diagnostic accuracy of Xpert and Ultra for TB stratified by W4SS status in smear-negative people. Trends were like those seen overall (**Table 3**). Data are %, 95% CI, and n/N.

	Head-to-head											
	All (n=787)			**W4SS-positive (n=357)**				**W4SS-negative (n=430)**				
	Sensitivity	**Specificity**	**PPV**	**Sensitivity**	**Specificity**	**PPV**	**Sensitivity**	**Specificity**	**PPV**	**Sensitivity**	**Specificity**	**PPV**
Xpert	50 (39, 61) 44/88	99 (98, 100) 671/678	86 (74, 94) 44/51	94 (92, 95) 671/715	22 (8, 44) 5/23	100 (98, 100) 328/331	63 (25, 92) 5/8	95 (92, 97) 328/346	60 (48, 72) 39/65 **p=0.0016**[*]	99 (98, 100) 343/347 p=0.75[*]	91 (78, 98) 39/43 **p=0.033**[*]	93 (90, 96) 343/369 p=0.31[*]
Ultra	67 (56, 77) 59/88 **p=0.022**[†]	98 (96, 99) 663/678 p=0.089[†]	80 (69, 88) 59/74 p=0.63[†]	96 (94, 97) 663/692 p=0.12[†]	44 (24, 66) 10/23 p=0.12[†]	99 (97, 100) 327/331 p=0.70[†]	72 (42, 92) 10/14 p=0.67[†]	97 (94, 98) 327/340 p=0.39[†]	76 (64, 86) 49/65 **p=0.0051**[*] p=0.061[†]	97 (95, 99) 336/347 p=0.083[*] p=0.068[†]	82 (70, 91) 49/60 p=0.39[*] p=0.20[†]	96 (93, 98) 336/352 p=0.64[*] p=0.15[†]
Non-head-to-head												
Xpert	50 (39, 61) 44/88	99 (98, 100) 672/679	86 (74, 94) 44/51	94 (92, 95) 672/716	22 (7, 44) 5/23	99 (97, 100) 328/331	63 (24, 91) 5/8	95 (92, 97) 328/346	60 (47, 72) 39/65 **p=0.0016**[*]	99 (97, 100) 344/348 p=0.75[*]	91 (78, 97) 39/43 **p=0.033**[*]	93 (90, 95) 344/370 p=0.31[*]
Ultra	68 (58, 78) 62/91 **p=0.014**[†]	98 (97, 99) 725/741 p=0.092[†]	79 (69, 88) 62/78 p=0.33[†]	96 (95, 97) 725/754 **p=0.043**[†]	76 (65, 86) 52/68 **p<0.0001**[†]	97 (95, 98) 367/379 **p=0.037**[†]	81 (70, 90) 52/64 p=0.22[†]	96 (93, 98) 367/383 p=0.51[†]	43 (23, 66) 10/23 **p=0.0033**[*] p=0.17[†]	99 (97, 100) 358/362 p=0.79[*] p=0.99[†]	71 (42, 92) 10/14 **p<0.0001**[*] p=0.071[†]	96 (94, 98) 358/371 p=0.63[*] **p=0.032**[†]

Within row p-values: [*]W4SS-positive vs. -negative.
Within column p-values: [†]Xpert vs. Ultra.
Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; Ultra, W4SS, WHO four-symptom screen; Xpert MTB/RIF; Xpert, Xpert MTB/RIF Ultra.

Appendix Table 9. Non-head-to-head diagnostic accuracy of Xpert and Ultra for TB detection stratified by CD4 count. Non-head-to-head diagnostic accuracy trends were like that in head-to-head comparisons (**Appendix Table 7**). Data are %, 95% CI, and n/N.

	All participants				**CD4 count >350 cells/µl**				**CD4 count ≤350 cells/µl**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Smear microscopy	15 (9, 24) 16/107	100 (99, 100) 743/748	77 (53, 92) 16/21	90 (87, 92) 743/834	17 (4, 42) 3/18	100 (99, 100) 309/309	10 (30, 100) 3/3	95 (93, 98) 309/324	13 (7, 23) 10/79 p=0.41*	99 (98, 100) 390/395 **p=0.044***	67 (39, 89) 10/15 p=0.19*	85 (82, 89) 390/459 **p<0.0001***
Xpert	56 (46, 66) 58/104 **p<0.0001†**	99 (98, 100) 677/684 p=0.46†	90 (80, 96) 58/65 p=0.13†	94 (92, 96) 677/723 **p=0.0016†**	29 (11, 56) 5/17 p=0.37†	99 (97, 100) 284/287 p=0.072†	63 (25, 92) 5/8 p=0.21†	96 (94, 98) 284/296 p=0.73†	60 (48, 71) 46/77 **p=0.023*** **p<0.0001†**	99 (98, 100) 352/356 p=0.924* p=0.86†	92 (81, 98) 46/50 **p=0.017*** **p=0.013†**	92 (89, 95) 352/383 **p=0.032*** **p=0.0019†**
Smear-negative	50 (40, 61) 44/88	99 (98, 100) 672/679	87 (74, 95) 44/51	94 (92, 96) 672/716	21 (5, 51) 3/14	99 (97, 100) 284/287	50 (12, 89) 3/6	96 (94, 99) 284/295	55 (43, 68) 37/67 **p=0.021***	99 (98, 100) 347/351 p=0.91*	90 (77, 98) 37/41 **p=0.010***	92 (89, 95) 347/377 **p=0.023***
Ultra-negative	10 (3, 27) 3/30	100 (99, 100) 663/668	38 (9, 76) 3/8	97 (95, 98) 663/690	0 (0, 31) 0/10	99 (97, 100) 280/283	0 (0, 71) 0/3	97 (94, 99) 280/290	17 (4, 42) 3/18 p=0.17*	99 (98, 100) 342/344 p=0.50*	60 (15, 95) 3/5 p=0.090*	96 (94, 98) 342/357 p=0.62*
Ultra	72 (63, 81) 77/107 **p=0.017‡**	97 (95, 98) 730/760 **p<0.0001‡**	83 (74, 90) 77/93 p=0.26‡	97 (95, 98) 730/760 **p=0.035‡**	44 (22, 70) 8/18 p=0.36‡	99 (97, 100) 305/309 p=0.78‡	67 (35, 91) 8/12 p=0.85‡	97 (95, 99) 305/315 p=0.56‡	77 (67, 86) 61/79 **p=0.0056*** **p=0.019‡**	97 (95, 99) 381/393 p=0.12* p=0.068‡	84 (74, 92) 61/73 p=0.17* p=0.17‡	95 (93, 98) 381/39 p=0.36* **p=0.039‡**
Smear-negative	69 (58, 78) 62/91 **p=0.010‡**	98 (97, 99) 725/741 p=0.092‡	80 (69, 88) 62/78 p=0.33‡	97 (95, 98) 725/754 **p=0.035‡**	40 (17, 68) 6/15 p=0.21‡	99 (97, 100) 305/309 p=0.78‡	60 (27, 88) 6/10 p=0.69‡	97 (95, 99) 305/314 p=0.55‡	74 (62, 84) 51/69 **p=0.011*** **p=0.023‡**	97 (95, 99) 376/388 p=0.12* p=0.068‡	81 (70, 90) 51/63 p=0.14* p=0.19‡	95 (93, 98) 376/394 p=0.24* p=0.052‡
Xpert-negative	42 (27, 57) 19/46 **p=0.0033‡**	99 (97, 99) 663/676 p=0.061‡	60 (41, 77) 19/32 p=0.27‡	97 (95, 98) 663/690 p>0.99‡	17 (3, 49) 2/12 p=0.18‡	99 (97, 100) 280/284 p=0.71‡	33 (5, 78) 2/6 p=0.26‡	97 (94, 99) 280/290 p>0.99‡	52 (34, 70) 16/31 **p=0.037*** **p=0.016‡**	97 (96, 99) 342/351 p=0.31* p=0.036‡	64 (43, 83) 16/25 p=0.17* p=0.87‡	96 (94, 98) 342/357 p=0.62* p>0.99‡

Within row p-values: *CD4 counts ≤350 vs. >350 cells/µl.
Within column p-values: †smear vs. Xpert and ‡Xpert vs. Ultra.
Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Appendix Table 10. Non-head-to-head diagnostic accuracy of Xpert and Ultra for TB overall and stratified by W4SS status. These findings are like those in the head-to-head analysis (**Table 3**). Data are %, 95% CI, and n/N.

	All				W4SS-negative				W4SS-positive			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Smear microscopy	15 (9, 24) 16/107	100 (99, 100) 743/748	77 (53, 92) 16/21	90 (87, 92) 743/834	12 (3, 31) 3/26	100 (99, 100) 362/362	100 (30, 100) 3/3	95 (92, 97) 362/385	17 (9, 26) 13/81 p=0.58*	99 (98, 100) 381/386 **p=0.029***	73 (47, 91) 13/18 p=0.29*	85 (82, 89) 381/449 **p<0.0001***
Xpert	56 (46, 66) 58/104 **p<0.0001**†	99 (98, 100) 677/684 p=0.46†	90 (80, 96) 58/65 p=0.13†	94 (92, 96) 677/723 **p=0.0016**†	27 (12, 48) 7/26 p=0.16†	100 (98, 100) 328/331 p=0.069†	70 (35, 94) 7/10 p=0.28†	95 (92, 97) 328/347 p=0.77†	66 (54, 76) 51/78 **p=0.0010*** **p<0.0001**†	99 (98, 100) 349/353 p=0.77* p=0.84†	93 (83, 98) 51/55 **p=0.033*** **p=0.022**†	93 (90, 96) 349/376 p=0.35* **p<0.0001**†
Ultra-negative	10 (3, 27) 3/30	100 (99, 100) 663/668	38 (9, 76) 3/8	97 (95, 98) 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=0.62*	60 (15, 95) 3/5 p=0.090*	97 (94, 99) 339/352 p=0.76*
Ultra	72 (63, 81) 77/107 **p=0.014**‡	97 (95, 98) 730/760 **p<0.0001**‡	83 (74, 90) 77/93 p=0.26‡	97 (95, 98) 730/760 **p=0.035**‡	47 (27, 67) 12/26 p=0.15‡	99 (98, 100) 358/362 p=0.79‡	75 (48, 93) 12/16 p=0.78‡	97 (94, 98) 358/372 p=0.27‡	81 (70, 89) 65/81 **p=0.001*** **p=0.035**‡	97 (95, 99) 372/384 p=0.057* p=0.064‡	85 (75, 92) 65/77 p=0.36* p=0.15‡	96 (94, 98) 372/388 p=0.79* p=0.61‡
Xpert-negative	42 (27, 57) 19/46 **p=0.0033**‡	99 (97, 99) 663/676 p=0.061‡	60 (41, 77) 19/32 p=0.27‡	97 (95, 98) 663/690 p>0.99‡	27 (10, 52) 5/19 **p=0.037**‡	99 (97, 100) 324/328 p=0.71‡	56 (22, 87) 5/9 p=0.091‡	96 (94, 98) 324/338 p>0.99‡	52 (32, 72) 14/27 p=0.083* **p=0.032**‡	98 (96, 99) 339/348 p=0.19* **p=0.036**‡	61 (39, 81) 14/23 p=0.78* p=0.97‡	97 (94, 99) 339/352 p=0.76* p>0.99‡

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

Appendix Table 11. Head-to-head diagnostic accuracy of Xpert and Ultra for TB detection by previous TB status. No specificity differences occurred when data stratified by previous TB status. Data are %, 95% CI, and n/N.

	All (n=787)				No previous TB (n=675)				Previous TB (n=112)			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert	56 (46, 66) 58/104	99 (98, 100) 676/683	89 (79, 96) 58/65	94 (92, 95) 676/722	58 (47, 69) 51/88	99 (98, 100) 581/587	90 (79, 97) 51/57	95 (92, 96) 581/618	44 (20, 70) 7/16 p=0.29*	99 (94, 100) 95/96 p=0.99*	88 (47, 100) 7/8 p=0.87*	91 (84, 96) 95/104 p=0.30*
Smear-negative	50 (39, 61) 44/88	99 (98, 100) 671/678	86 (74, 94) 44/51	94 (92, 95) 671/715	54 (42, 65) 40/75	99 (98, 100) 576/582	87 (74, 96) 40/46	95 (93, 96) 576/611	31 (9, 61) 4/13 p=0.13*	99 (94, 100) 95/96 p=0.99*	80 (28, 99) 4/5 p=0.67*	91 (84, 96) 95/104 p=0.25*
Ultra-negative	10 (2, 27) 3/30	99 (98, 100) 663/668	38 (9, 76) 3/8	96 (94, 97) 663/690	13 (3, 33) 3/24	100 (98, 100) 570/575	38 (9, 76) 3/8	97 (95, 98) 570/591	0 (0, 46) 0/6 p=0.36*	100 (96, 100) 93/93 p=0.37*	Non-calculable -	94 (87, 98) 93/99 p=0.23*
Ultra	71 (61, 80) 74/104 **p=0.021**‡	98 (96, 99) 668/683 p=0.086‡	83 (74, 90) 74/89 p=0.29‡	96 (94, 97) 668/698 p=0.083‡	73 (63, 82) 64/88 **p=0.039**‡	98 (97, 99) 575/587 p=0.15‡	85 (75, 92) 64/76 p=0.38‡	96 (95, 98) 575/599 p=0.11‡	63 (35, 85) 10/16 p=0.41* p=0.29‡	97 (91, 99) 93/96 p=0.50* p=0.31‡	77 (46, 95) 10/13 p=0.52* p=0.55‡	94 (87, 98) 93/99 p=0.35* p=0.48‡
Smear-negative	67 (56, 77) 59/88 **p=0.022**‡	98 (96, 99) 668/683 p=0.086‡	83 (74, 90) 74/89 p=0.63‡	96 (94, 97) 668/698 p=0.72‡	70 (58, 80) 52/75 **p=0.044**‡	98 (97, 99) 570/582 p=0.15‡	82 (70, 90) 52/64 p=0.43‡	97 (95, 98) 570/593 p=0.13‡	54 (25, 81) 7/13 p=0.27* p=0.23‡	97 (91, 99) 93/96 p=0.51* p=0.31‡	70 (35, 93) 7/10 p=0.41* p=0.68‡	94 (87, 98) 93/99 p=0.32* p=0.48‡
Xpert-negative	41 (27, 57) 19/46 **p=0.0033**‡	98 (97, 99) 663/676 p=0.061‡	59 (41, 76) 19/32 p=0.27‡	96 (94, 97) 663/690 p>0.99‡	44 (28, 61) 16/37 **p=0.011**‡	99 (97, 100) 570/581 p=0.14‡	60 (39, 78) 16/27 p=0.29‡	97 (95, 98) 570/591 p>0.99‡	33 (7, 70) 3/9 p=0.59* p=0.11‡	98 (93, 100) 93/95 p=0.89* p=0.16‡	60 (15, 95) 3/5 p=0.98* -	94 (87, 98) 93/99 p=0.23* p>0.99‡

Within row p-values: *no previous TB vs. previous TB.
Within column p-values: ‡Xpert vs. Ultra.
Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis; Ultra, Xpert MTB/RIF; Xpert, Xpert MTB/RIF.

Appendix Table 12. Effect of different trace re-categorisation strategies (reclassification or exclusion) on Ultra diagnostic accuracy, with stratification by previous TB status. Both the re-calculated estimate and change (Δ) are shown. In people without previous TB, re-categorising traces decreased sensitivity (reclassification only) and increases specificity (reclassification and exclusion strategies), while no differences were seen in people with previous TB. Data are %, 95% CI, and n/N.

	All (n=787)			**No previous TB (n= 675)**			**Previous TB (n=112)**		
	Overall	**Trace reclassified**	**Trace excluded**	**Overall**	**Trace reclassified**	**Trace excluded**	**Overall**	**Trace reclassified**	**Trace excluded**
Sensitivity	71 (61, 80) 74/104	66 (57, 76) 69/104 -5 (-10, 0) p=0.45†	70 (60, 79) 69/99 -1 (-11, 14) p=0.82† p=0.61‡	73 (63, 82) 64/88	67 (57, 77) 59/88 -6 (-12, 0) p=0.41†	71 (61, 81) 59/83 -2 (-15, 12) p=0.81† p=0.57‡	63 (35, 85) 10/16 p=0.41*	63 (35, 85) 10/16 0 (-6, 6) p=0.72* p>0.99†	63 (35, 85) 10/16 0 (-6, 6) p=0.49* p>0.99† p>0.99‡
Specificity	98 (96, 99) 668/683	99 (99, 100) 679/683 1 (1, 3) **p=0.011**†	99 (99, 100) 668/672 1 (0, 3) **p=0.012**† p=0.98‡	98 (97, 99) 575/587	99 (99, 100) 584/587 2 (0, 3) **p=0.019**†	99 (99, 100) 575/578 2 (0, 3) **p=0.021**† p=0.99‡	97 (91, 99) 93/96 p=0.50*	99 (94, 100) 95/96 2 (-2, 6) p=0.53* p=0.16†	99 (94, 100) 93/94 2 (-2, 6) p=0.52* p=0.32† p=0.99‡
PPV	83 (74, 90) 74/89	95 (87, 99) 69/73 11 (2, 21) **p=0.025**†	95 (87, 99) 69/73 11 (2, 21) **p=0.025**† p>0.99‡	85 (75, 92) 64/76	95 (87, 99) 59/62 11 (1, 21) **p=0.040**†	95 (87, 99) 59/62 11 (1, 21) **p=0.004**† p>0.99‡	77 (46, 95) 10/13 p=0.52*	91 (59, 100) 10/11 13 (-15, 43) p=0.57* p=0.36†	91 (59, 100) 10/11 13 (-15, 43) p=0.57* p=0.36† p>0.99‡
NPV	96 (94, 97) 668/698	95 (94, 97) 679/714 -1 (-3, 2) p=0.59†	96 (94, 98) 668/698 0 (-2, 2) p>0.99† p=0.59‡	96 (95, 98) 575/599	95 (94, 97) 584/613 -1 (-3, 2) p=0.54†	96 (95, 98) 575/599 p>0.99† p=0.54‡	94 (87, 98) 93/99 p=0.35*	94 (88, 98) 95/101 0 (-6, 6) p=0.60* p=0.97†	94 (87, 98) 93/99 0 (-7, 7) p=0.35* p>0.99† p=0.97‡

Within row p-values: *No previous TB vs. previous TB; †vs. overall in participants of the same previous TB status; ‡trace reclassified vs. excluded.
Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra.

Appendix Table 13. Diagnostic accuracy of urine tests (LF-LAM, concentrated Ultra) for TB overall and by W4SS status (head-to-head and non-head-to-head data). Urine Ultra and LF-LAM had similar sensitivity and specificity overall and by W4SS. A positive Ultra had higher PPV in W4SS-positives vs. -negatives, including in LF-LAM-negatives. Similarly, Ultra and LF-LAM had higher NPVs in W4SS-negatives vs. positives. Data are %, 95% CI, and n/N.

	Head-to-head											
	All participants (n=732)				**W4SS-negative (n=339)**				**W4SS-positive (n=393)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
LF-LAM	15 (9, 24) 14/97	99 (98, 100) 623/635	61 (39, 81) 14/23	89 (86, 91) 626/709	9 (2, 29) 2/23	100 (98, 100) 313/316	40 (6, 86) 2/5	94 (91, 97) 313/334	17 (9, 27) 12/74 p=0.37*	99 (96, 100) 313/319 p=0.32*	67 (41, 87) 12/18 p=0.28*	84 (80, 88) 313/375 **p<0.0001***
Conc. Ultra-negatives	9 (4, 18) 6/73	99 (98, 100) 619/628	40 (17, 68) 6/15	91 (88, 93) 619/686	5 (1,25) 1/20	100 (98, 100) 309/312	25 (1, 81) 1/4	95 (92, 97) 309/328	10 (4,21) 5/53 p=0.54*	99 (96, 100) 310/316 p=0.32*	46 (17, 77) 5/11 p=0.48*	87 (83,90) 310/358 **p<0.0001***
Conc. Ultra	25 (17, 35) 24/97 p=0.070**	99 (98, 100) 628/635 p=0.25**	78 (59, 91) 24/31 p=0.19**	90 (88, 92) 628/701 p=0.44**	14 (3, 34) 3/23 p=0.64**	99 (97, 100) 312/316 p=0.70**	43 (10, 82) 3/7 p=0.92**	94 (91, 97) 312/332 p=0.89**	29 (19,41) 21/74 p=0.14* p=0.076**	100 (98, 100) 316/319 p=0.69* p=0.31**	88 (68, 98) 21/24 **p=0.013*** p=0.10**	86 (82, 90) 316/369 **p<0.0001*** p=0.41**
LF-LAM-negatives	20 (12, 30) 16/83 **p=0.048**†**	99 (98, 100) 619/626 p=0.62†	70 (48, 87) 16/23 p=0.071†	91 (88, 93) 619/686 p>0.99†	10 (2, 31) 2/21 p=0.58†	99 (97, 100) 309/313 p=0.71†	34 (5, 78) 2/6 p=0.78†	95 (92, 97) 309/328 p>0.99†	23 (13, 35) 14/62 p=0.19* p=0.058†	100 (98, 100) 310/313 p=0.69* p=0.32†	83 (57, 97) 14/17 **p=0.025*** **p=0.041**†**	87 (83, 90) 310/358 **p<0.0001*** p>0.99†
	Non-head-to-head											
LF-LAM	16 (9, 4) 16/106	99 (98, 100) 732/745	56 (36, 74) 16/29	90 (87, 92) 732/822	8 (1, 26) 2/26	99 (97, 100) 356/362	25 (4, 66) 2/8	94 (91, 96) 356/380	18 (10, 28) 14/80 p=0.23*	99 (97, 100) 376/383 p=0.86*	67 (44, 86) 14/21 **p=0.044***	86 (82, 89) 376/442 **p<0.0001***
Conc. Ultra-negatives	9 (4, 18) 6/73	99 (98, 100) 732/745	34 (14, 60) 6/18	91 (88, 93) 616/683	5 (1, 25) 1/20	99 (97, 100) 356/362	15 (1, 58) 1/7	95 (92, 97) 306/325	10 (4, 21) 5/53 p=0.54*	99 (97, 100) 376/383 p=0.86*	46 (17, 77) 5/11 p=0.17*	87 (83, 90) 310/358 **p=0.001***
Conc. Ultra	25 (17, 35) 24/97 p=0.084**	99 (98, 100) 629/636 p=0.32**	78 (59, 91) 24/31 p=0.068**	90 (88, 92) 629/702 p=0.73**	14 (3, 34) 3/23 p=0.54**	99 (97, 100) 312/316 p=0.67**	43 (10, 382) 3/7 p=0.46**	94 (91,9 7) 312/332 p=0.87**	29 (19, 41) 21/74 p=0.137* p=0.11**	100 (98, 100) 317/320 p=0.692* p=0.32**	88 (68, 98) 21/24 **p=0.013*** p=0.094**	86 (82, 90) 317/370 **p<0.0001*** p=0.81**
LF-LAM-negatives	20 (12, 30) 16/83	99 (98, 100) 616/623	70 (48, 87) 16/23	91 (88, 93) 616/683	10 (2, 31) 2/21	99 (97, 100) 306/310	34 (5,78) 2/6	95 (92, 97) 306/325	23 (13, 35) 14/62 p=0.19*	100 (98, 100) 310/313 p=0.69*	83 (57, 97) 14/17 **p=0.025***	87 (83, 90) 310/358 **p<0.0001***

	p=0.048[†]	p=0.34[†]	**p=0.021**[†]	p>0.99[†]	p=0.58[†]	p=0.69[†]	p=0.42[†]	p>0.99[†]	p=0.058[†]	p=0.338[†]	**p=0.041**[†]	p>0.99[†]

Within row p-values: [*]W4SS-positive vs. -negatives.
Within column p-values: [**]LF-LAM vs. conc. Ultra; [†]LF-LAM in Ultra-negatives vs. Ultra in LF-LAM-negatives.
Abbreviations: CI, confidence interval; Conc., concentrated; LF-LAM, Determine TB LAM Ag test (LF-LAM), NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis, Ultra, Xpert MTB/RIF Ultra; W4SS, WHO-recommended four-symptom screen.

Appendix Table 14. Number of people who could expectorate different numbers of sputum specimens (≥1 ml each). Whether people could naturally expectorate sputum before induction was offered was only successfully recorded for the last 163 patients (**Methods**) and, of these, five did not have any culture attempted due to human error and are omitted (the number of sputa induction successfully produced are specified in the footnote). The proportion of people with a culture-positive result is shown overall and stratified by W4SS status. All people who received induction could produce at least one induced sputum. Data are % (n/N).

	No. expectorated sputa			
	None[†] (i.e., induction attempted for all sputa)	≥1	≥2	≥3 (i.e., no induction required)
Overall	31 (49/158)	69 (109/158)	64 (101/158)	62 (98/158)
Culture-positive[*]	20 (10/49)	16 (17/105)	15 (15/97)	14 (14/94)
W4SS-negative	20 (11/55)	80 (44/55)	73 (40/55)	71 (39/55)
Culture-positive	0 (0/11)	7 (3/44)	5 (2/40)	5 (2/39)
W4SS-positive	37 (38/103)	63 (65/103)	59 (61/103)	57 (59/103)
Culture-positive	26 (10/38)	23 (14/61)	23 (13/57)	22 (12/55)

[*]Four patients with only contaminated cultures excluded from culture-positivity rate calculation.
[†]Number of sputum produced as a result of induction in people who could not expectorate any: 7/49 (14%), 5/49 (10%) and 37/49 (76%) people produced 1, 2, or 3 induced sputa, respectively.
Abbreviations: TB, tuberculosis; W4SS, WHO-recommended four-symptom screen.

Appendix Table 15. Yield (proportion of people with at least one positive confirmatory test result detected by a test) and sensitivity and specificity of individual tests according to whether people could expectorate. For each test, yield, sensitivity, and specificity did not differ based on whether a person was able to expectorate sputum, however, this conclusion is limited by the relatively small number of people with a positive test result in the subset of people with sputum expectoration status information. Data are %, 95% CI, and n/N.

	All participants		**Expectorator**		**Required induction**	
Yield						
Sputum						
Xpert	59 (39, 76) 17/29		55 (32, 77) 11/20		67 (30, 93) 6/9	
Ultra	66 (46, 82) 19/29		65 (41, 85) 13/20		67 (30, 93) 6/9	
Culture	83 (64, 94) 24/29		80 (56, 94) 16/20		89 (52, 100) 8/9	
Urine						
Ultra*	34 (18, 54) 10/29		25 (9, 49) 5/20		56 (21, 86) 5/9	
LF-LAM	24 (10, 44) 7/29		15 (3, 38) 3/20		44 (14, 79) 4/9	
Sensitivity and specificity						
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Sputum						
Xpert	67 (45, 84) 16/24	99 (94, 100) 96/97	63 (35, 85) 10/16	98 (92, 100) 63/64	75 (35, 97) 6/8	100 (89, 100) 33/33
Ultra	75 (53, 90) 18/24	99 (94, 100) 97/98	75 (48, 93) 12/16	98 (92, 100) 63/64	75 (35, 97) 6/8	100 (90, 100) 34/34
Urine						
Conc. Ultra	32 (14, 55) 7/22	100 (96, 100) 84/84	21 (5, 51) 3/14	100 (94, 100) 57/57	50 (16,84) 4/8	100 (87, 100) 27/27
LF-LAM	25 (10, 47) 6/24	99 (94, 100) 97/98	13 (2, 38) 2/16	98 (92, 100) 63/64	50 (16,84) 4/8	100 (90, 100) 34/34

*Conc. and unconc. positive results included.

Abbreviations: CI, confidence interval; Conc., concentrated; LF-LAM, Determine TB LAM Ag test, NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis, Ultra, Xpert MTB/RIF Ultra; Unconc., unconcentrated; W4SS, WHO-recommended four-symptom screen.

References

1. Boditech. iChromax CRP (Package insert). 12 May 2021 2021. http://web.labindustrias.com/wp-content/uploads/2022/02/ichroma-CRP-Rev.24.pdf (accessed 04 April 2022).
2. Cepheid. Xpert MTB/RIF Ultra: Raising the Standard for Tuberculosis Diagnosis (Package insert). . 2022. on:https://p.widencdn.net/sihiwi/Cepheid-Xpert-MTB-RIF-Ultra-Brochure-CE-IVD-3098-English (accessed 04 April 2022).
3. Xpert MTB/RIF (Package insert) Cepheid S, CA 300-6081, Rev, AA Available from https://www.cepheid.com/administrator/components/com_productcatalog/library-files/d09039779cfd2ef61d7ec992cbca395b-2aa30e2b9837e71bec72ec6bf53197b6-Xpert-MTBRIF-US-0421-04.pdf (Accessed January 2021). Cepheid; 2018.
4. Xpert MTB/RIF Ultra (Package insert) Cepheid, S, CA. 300-6081, Rev. AA. Available on:https://p.widencdn.net/sihiwi/Cepheid-Xpert-MTB-RIF-Ultra-Brochure-CE-IVD-3098-English. accessed on 20 July 2020. Cepheid; May 2018.
5. Alere DETERMINE™ TB LAM AG (Package insert). Abbott, USA. Available on: https://www.alere.com/en/home/product-details/determine-tb-lam.html . Accessed on 20 July 2020. Abbott; July 2020.
6. Drain PK, Gardiner J, Hannah H, et al. Guidance for studies evaluating the accuracy of biomarker-based nonsputum tests to diagnose tuberculosis. 2019; **220** (Supplement_3): S108-S15.
7. Schumacher SG, Wells WA, Nicol MP, et al. Guidance for studies evaluating the accuracy of sputum-based tests to diagnose tuberculosis. 2019; **220** (Supplement_3): S99-S107.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	3-4 (abstract); 12 (results)
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	3-4
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	9 (methods)
	4	Study objectives and hypotheses	7-8 (introduction)
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Prospective; 9 (methods)
Participants	**6**	Eligibility criteria	9
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	9
	8	Where and when potentially eligible participants were identified (setting, location and dates)	9
	9	Whether participants formed a consecutive, random or convenience series	9-11
Test methods	**10a**	Index test, in sufficient detail to allow replication	9-11
	10b	Reference standard, in sufficient detail to allow replication	9-10
	11	Rationale for choosing the reference standard (if alternatives exist)	9
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	9; 2, Appendix
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	2, Appendix
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	10
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	10
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	10
	15	How indeterminate index test or reference standard results were handled	Appendix 2
	16	How missing data on the index test and reference standard were handled	14
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	14-16
	18	Intended sample size and how it was determined	4, Appendix
RESULTS			
Participants	**19**	Flow of participants, using a diagram	23-24 Figure 1, 2
	20	Baseline demographic and clinical characteristics of participants	28, Table 1
	21a	Distribution of severity of disease in those with the target condition	28, Table 1
	21b	Distribution of alternative diagnoses in those without the target condition	N/A
	22	Time interval and any clinical interventions between index test and reference standard	10
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	33-35, Table 2-3; 12-19, Appendix Tables 1-5; 21-28, Appendix Tables 7-13
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	12-15
	25	Any adverse events from performing the index test or the reference standard	10
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	19-20
	27	Implications for practice, including the intended use and clinical role of the index test	18

OTHER INFORMATION			
	28	Registration number and name of registry	NCT03187964
	29	Where the full study protocol can be accessed	9-11
	30	Sources of funding and other support; role of funders	11; 21-22

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.