Clinical Infectious Diseases

MAJOR ARTICLE

Oral Washes and Tongue Swabs for Xpert MTB/RIF Ultra–Based Tuberculosis Diagnosis in People With and Without the Ability to Make Sputum

Loren Rockman,[1] Shima M. Abdulgader,[1] Stephanie Minnies,[1] Zaida Palmer,[1] Charissa C. Naidoo,[1] Rouxjeane Venter,[1] Daphne Naidoo,[1] Gcobisa Ndlangalavu,[1] Byron W. P. Reeve,[1,a] Andrew Medina-Marino,[2] Tim J. Bull,[3] Alaina M. Olson,[4] Rachel Wood,[4] Gerard A. Cangelosi,[4] Robin M. Warren,[1] and Grant Theron[1]

[1]Department of Science and Innovation-National Research Foundation 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]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [3]St George's School of Health and Medical Sciences, City and St George's University of London, London, United Kingdom; and [4]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington, USA

Background. Oral samples show promise for tuberculosis (TB) diagnosis, but data from different sample types and sputum-scarce individuals remain limited.

Methods. We evaluated Xpert MTB/RIF Ultra (Ultra) in symptomatic clinic attendees (cohort A, n = 891) and people initiating antiretroviral therapy without symptom screening (cohort B, n = 258). In cohort A, we collected oral washes (OWs) and, separately, tongue swabs (flocked or foam with heat). In cohort B, we collected OWs, 3 flocked tongue swabs (1 heated, 2 pooled), and, separately, buccal swabs and periodontal brushes. Sputum induction was offered, and different culture methods were applied to a subset of cohort B tongue swabs.

Results. In cohort A, Ultra sensitivity was 80% (95% confidence interval [CI], 56%–94%) for OWs, 59% (95% CI, 53%–65%) for flocked swabs, and 65% (95% CI, 58%–72%) for foam swabs, with high specificity. Foam swabs detected more people with Ultra sputum semi-quantitation categories of low or less than flocked swabs (53% [95% CI, 41%–64%] vs 37% [95% CI, 29%–46%]). In cohort B, OWs and single heated swabs had sensitivities of 71% (95% CI, 42%–92%) and 64% (95% CI, 35%–87%), respectively. Pooled tongue swabs, buccal swabs, and brushes had lower sensitivity. MGIT960 showed the highest sensitivity (64% [95% CI, 35%–87%]) among culture methods. Oral sampling identified TB in sputum-scarce people: 25% (7/28) positive by flocked or foam swabs (cohort A); 18% (10/56) were OW and 23% (13/56) single swab positive (cohort B). In cohort B, this could double Ultra positivity if induction were unavailable.

Conclusions. Ultra on OWs or foam swabs offers higher sensitivity than other oral methods and effectively detects TB in sputum-scarce individuals.

Keywords. Xpert MTB/RIF Ultra; tuberculosis; oral washes; tongue swabs.

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* complex (*Mtb*), is a global pandemic [1]. Molecular tests like Xpert MTB/RIF Ultra (Ultra; Cepheid, Sunnyvale, California, USA) and Truenat MTB Plus (Molbio Diagnostics, Goa, India) were designed for sputum. Non-sputum tests are urgently needed to reach the millions of people with TB who go undiagnosed each year [2–4].

Tongue swabs, which collect biofilm from the dorsum tongue, have data to support use with the Ultra and Truenat MTB Ultima (Ultima) tests: Ultra has sensitivities from 72% to 75% with high specificity, with similar performance described for Ultima [5]. Furthermore, discrete choice studies show that tongue swabs are more acceptable than sputum to some people despite potential sensitivity trade-offs [6]. Importantly, a test with diminished sensitivity can increase diagnostic yield if the number of individuals tested is higher [7, 8]; however, there are few empiric data in the context of oral samples [9].

Key gaps remain regarding the diagnostic potential of oral samples, including samples other than tongue swabs. Early proof-of-concept studies used in-house rather than commercially available tests or, as tongue swabs are likely most

Received 13 March 2025; editorial decision 10 July 2025; published online 18 July 2025

[a]Deceased.

Correspondence: G. Theron, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Room 2035, 2nd Floor BMRI Building, Francie Van Zijl Drive, Tygerberg 7505, South Africa (gtheron@sun.ac.za).

Clinical Infectious Diseases®

https://doi.org/10.1093/cid/ciaf397

beneficial in people unable to expectorate, paradoxically preselected people based on their ability to naturally produce sputum [10–13]. Furthermore, the type of tongue swab (flocked, foam) and processing method may influence performance [14]. Equally important, data from people with risk factors for TB not preselected based on symptoms, who are less able to make sputum and increasingly targeted as part of facility active case finding strategies, are scarce. More data are urgently needed to inform global policy.

Data are also scarce regarding *Mtb* culturability from tongue swabs. One study found 44%–58% of swabs to be MGIT960 culture positive in people with TB [15]. Culturable *Mtb* from the swab could reinforce microbiological reference standards for diagnostic accuracy evaluations and provide material for drug susceptibility testing. Last, culture methods that do not involve harsh N-Acetyl-L-Cysteine-Sodium Hydroxide (NaOH) decontamination and are designed for paucibacillary samples may have added utility.

To address these knowledge gaps, we evaluated different types of oral samples, swabs, and processing methods for TB diagnosis in a human immunodeficiency virus (HIV) high-burden setting where people were offered sputum induction. We hypothesized that Ultra would detect *Mtb* in oral specimens with high concordance to sputum.

MATERIALS AND METHODS

Ethics Statement

This work was approved by the Health Research Ethics Committee of Stellenbosch University (N14/10/136, M20/06/017, M20/06/018) and City of Cape Town (10570). Written informed consent was obtained.

Study Cohorts

In cohort A, adults (≥18 years) self-reporting with presumptive pulmonary TB symptoms meeting World Health Organization (WHO) criteria (at least 1 symptom for people with HIV [PWH], ≥2 for those without HIV) were recruited at clinics in Cape Town, South Africa. The standard WHO-defined symptoms for TB were used (cough, weight loss, fever, drenching night sweats), with 2 required if HIV negative and 1 if HIV positive (for people without HIV, cough needed to have persisted for at least 2 weeks; for PWH, any cough duration qualified). For cohort B, antiretroviral therapy (ART) initiators (≥18 years), regardless of symptoms, were recruited in Cape Town as previously described [16, 17].

Definitions

In cohort A, people were classified as having TB using an extended microbiological reference standard (eMRS) if sputum was Ultra positive or had *Mtb*-positive culture growth. Those categorized as not having TB had no positive sputum Ultra or at least 1 negative culture result. Those missing a culture or Ultra result were, if other results were negative, classified as not having TB.

In cohort B, the eMRS consisted of 2 sputum cultures. If at least 1 culture was *Mtb* positive, participants were classified as having TB. Participants with negative culture(s) did not have TB.

For both cohorts, treatment decisions were made outside of the study without knowledge of oral test results.

Specimen Collection

All oral samples were collected prior to sputum collection after ≥30 minutes had passed since food or fluid was ingested and teeth were brushed. Figure 1 summarizes specimen collection and testing.

Oral Washes

Each person (cohort A, n = 40; cohort B, n = 156) received 2 vials of 20 mL sterile water. One vial was used to rinse the mouth and discarded. The second vial's contents were swirled in the mouth for a few seconds and spat into the vial. All samples were stored at −20°C until processed. In cohort A, oral washes (OWs) were collected from separate people than tongue swabs. In cohort B, OWs were collected in everyone who gave tongue swabs and collection was done immediately after swab collection.

Tongue Swabs

In cohort A, 1 dry flocked swab (FLOQSwabs, code 520C; Copan Italia S.p.A., Brescia, Italy) followed by a dry foam swab (Medline Industries, Northfield, Illinois, USA) were consecutively collected from 851 and 550 people, respectively, by scraping the tongue dorsum for 10–15 seconds and placing the swab in a dry tube (550 people had both swab types). As a control for each person, an air swab was collected in the same space as the participant by waving a foam swab (flocked prior to the start of foam swab sampling) in the air for 10–15 seconds immediately before sampling.

In cohort B, 1 flocked swab was collected in 800 µL Tris–ethylenediaminetetraacetic acid (TE) buffer from 156 people. The last 122 were asked to provide an additional 2 flocked swabs subsequently pooled into a single tube with 800 µL TE buffer. A flocked air swab was done as for cohort A. In both cohorts, samples were stored (−20°C) until processed.

Buccal Swabs and Periodontal Brushes

In cohort B, paired buccal swabs and periodontal brushes (n = 102) were collected (Supplementary Text p. 2) and stored (−20°C) until processed. Collection occurred prior to the start of tongue swab and OW collection.

Sputum

Cohort A participants were each asked to provide 2 sputum samples, while cohort B participants were each asked to provide 3 sputum samples. Sputa were used for Ultra and MGIT960 culture (with 1% NALC-NaOH decontamination and *Mtb*

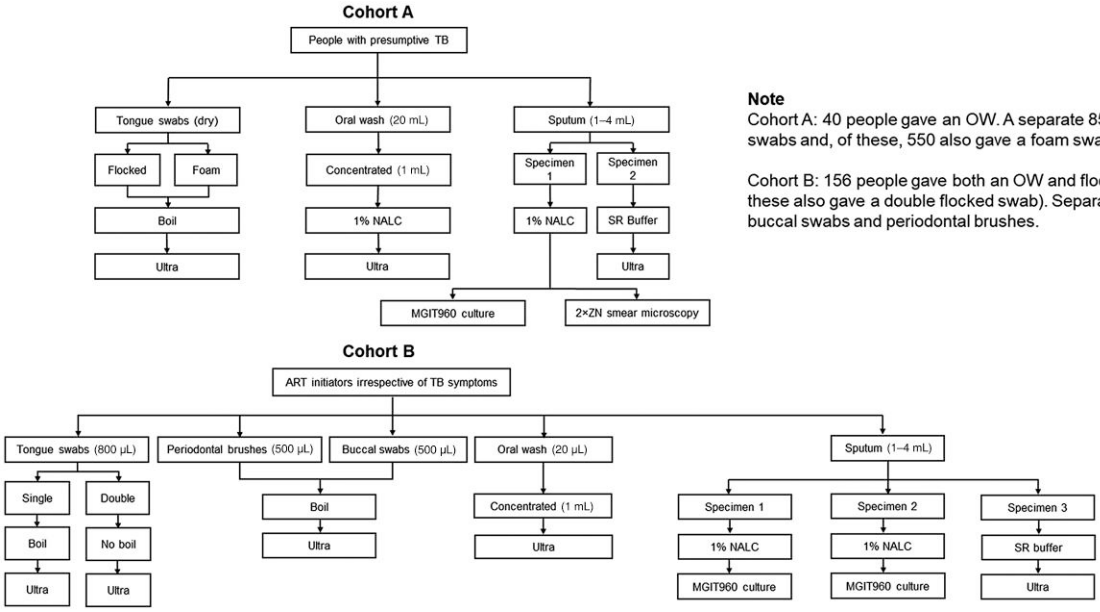

Figure 1. Flowchart showing enrollment, specimen collection and processing, and tests in cohorts A and B. Abbreviations: ART, antiretroviral therapy; NALC, N-Acetyl-L-Cysteine; OW, oral wash; SR, sample reagent; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; ZN, Ziehl-Neelsen.

confirmation using MTBDR*plus*, Bruker-Hain Diagnostics, Nehren, Germany; 1 culture in cohort A, 2 in cohort B). Sputum induction was done [16, 18]; however, it was only recorded whether people definitively required induction to make at least one ≥1 mL sputum sample in 584 cohort A people and the 156 cohort B people that gave tongue swabs. Figure 1 summarizes specimens collected by cohort.

Specimen Processing and Testing

All processing was performed in a Biosafety Level 3 laboratory.

Oral Wash Testing Using Ultra (Cohorts A and B)

In cohort A, OWs were concentrated, decontaminated, and processed with sample reagent (SR; Cepheid) per Supplementary Text p. 2. Decontamination was not done in cohort B.

Tongue Swab Testing Using Ultra (Cohorts A and B)

In cohort A, swabs were removed from storage and placed into a heating block (100°C, 10 minutes), after which TE buffer was added to 2 mL and the whole volume was tested with Ultra [19] with no SR.

In cohort B, single flocked swabs in TE buffer were removed from storage and immediately boiled (100°C, 10 minutes), followed by SR addition (1.6 mL to 800 µL sample) and Ultra [19]. For the 122 people who also gave a double swab, no heating was done, 2:1 SR was added, and Ultra was done (Supplementary Figure 1).

Air Swab Controls in People Who had Tongue Swabs (Cohorts A and B)

Air swabs were processed and tested with Ultra using the same procedure as tongue swabs in every tenth patient. If a participant had a positive tongue swab, their air swab was tested.

Buccal Swabs and Periodontal Brushes (Cohort B)

Stored samples were removed, processed, and tested with Ultra using the same procedure as tongue swabs in cohort B.

Tongue Swab Culture (Cohort B)

We did different types of culture (MGIT960, TiKa [20], early bactericidal activity [21]; see methodology in Supplementary Text p. 2) on single flocked tongue swabs (without heat inactivation). Speciation on positive growth was done using Ultra on a concentrated MGIT960 tube [22].

Analysis

Methods and reporting are per Standard for Reporting Diagnostic Accuracy (STARD) guidelines [23]. Diagnostic accuracy metrics were calculated using Excel software (Microsoft, Redmond, Washington, USA) and compared using prtest [24] and Fisher exact test [25] in Stata software version 16.0 (StataCorp, College Station, Texas, USA). The prtest or Fisher exact test (where $n < 5$) Stata functions were used to calculate *P* values for comparisons between proportions. The effect on sensitivity and specificity of Ultra trace results reclassified to Ultra negative in 2×2 tables was evaluated, as was Ultra trace exclusion with these people removed from 2×2 tables. In cohort B, head-to-head comparisons involved

Table 1. Demographic and Clinical Characteristics

Characteristic	Cohort A			Cohort B		
	Overall (n = 891)	Definite TB (n = 315)	Non-TB (n = 576)	Overall (n = 258)	Definite TB (n = 28)	Non-TB (n = 230)
Demographic characteristics						
Age, y	38 (30–48)	38 (30–47)	38 (30–48) *P* = .364*	37 (31–44) *P* = .956**	36 (31–44) *P* = .378**	39 (31 -43) *P* = .352* **P = .002****
Female sex	445/891 (50)	131/315 (42)	314/576 (55) **P = .002***	156/258 (60) **P = .003****	11/28 (39) *P* = .813**	145/230 (63) **P = .015*** **P = .027****
Clinical characteristics						
HIV-positive	355/891 (40)	136/315 (44)	219/576 (38) *P* = .133*	258/258 (100) **P < .0001****	28/28 (100) **P < .0001****	230/230 (100) **P < .0001****
CD4 count, cells/μL	305 (269–903)	582 (250–945)	305 (290–605) *P* = .587*	293 (143–434) **P < .0001****	59 (22–135) **P < .0001****	299 (171–438) **P = .013*** **P < .0001****
TBScoreII	3 (2–3)	3 (3–4)	3 (2–3) **P < .0001***	2 (0–2) **P < .0001****	3 (2–4) **P = .001****	1 (0–2) **P = .0006*** **P < .0001****
Previous TB	398/891 (45)	114/315 (36)	284/576 (49) **P = .0002***	26/258 (10) **P < .0001****	2/28 (7) **P = .002****	24/230 (10) *P* = .585* **P < .0001****

P values in bold are significant. Data are presented as no./No. (%) or median (interquartile range). The breakdown of individual culture and Ultra results used to classify people is shown in the Supplementary Results.

Abbreviations: HIV, human immunodeficiency virus; TB, tuberculosis.

*Within-row *P* values: definite TB vs non-TB within the same cohort.

**Within-row *P* values: across cohorts for people of the same TB status.

the same people who received both sampling approaches. Non–head-to-head comparisons compared people within the same cohort who were not necessarily the same individuals (in other words, not everyone received the exact same sampling approaches). Continuous data were compared with GraphPad Prism version 8.0.1 (GraphPad Software, San Diego, California, USA), also used for linear regression and correlation analysis. Diagnostic yield (DYT, diagnostic yield in those tested; DYD, diagnostic yield in those diagnosed) was calculated as described previously [8] and defined in the Supplementary Text p. 3. Morbidity score information (TBScoreII) was calculated from 7 clinical variables (cough, hemoptysis, chest pain, night sweats, pale conjunctivae, body mass index, and mid-upper arm circumference). Higher scores indicate higher morbidity [26]. Venn diagrams were made using InteractiVenn [27]. In cohort B, people were designated asymptomatic based on the WHO 4-symptom screen [28]. *P* values ≤.05 were significant. Unsuccessful results are those not positive or negative by any test. We compared sample processing control (SPC) cycle threshold (C_T) values from Ultra to measure inhibition (lower SPC C_T means less inhibition) [29].

RESULTS

Participant Demographics

People in cohort A were, compared to cohort B, less likely to be female, have higher morbidity, and more likely to have previous TB (Table 1). In both cohorts, people with TB were more likely to be male and have higher morbidity.

Diagnostic Accuracy of Ultra

Data are summarized in Figures 2 and 3. All air swabs were negative.

Cohort A

For OWs, no unsuccessful results occurred. Sensitivity and specificity were both 80% (95% confidence interval [CI], 56%–94%). Sensitivity was higher among those without HIV compared to PWH (94% [95% CI, 70%–100%] vs 25% [95% CI, 1%–81%]; *P* = .002) (Table 2). Four false-positive results occurred (all trace semi-quantitation, 2 previous TB).

In tongue swabs, 3% (25/851) of flocked swabs and 2% (13/550; *P* = .611) of foam swabs had unsuccessful results (3 both, 22 flocked only, 10 foam only; mostly overpressure errors). Flocked swabs had lower sensitivity than foam swabs (59% [95% CI, 53%–65%] vs 65% [95% CI, 58%–72%]; *P* = .176) with high specificities (94% [95% CI, 91%–96%] vs 92% [95% CI, 89%–95%]; *P* = .292) (Table 2). Among false-positive swabs, 26% (9/34) of flocked swabs and 39% (11/28) of foam swabs were from people programmatically empirically treated (no positive bacteriology at treatment start). Different trace recategorization strategies resulted in small sensitivity decreases and large specificity increases (Supplementary Table 1).

Cohort B

In OWs, no unsuccessful results occurred. Sensitivity was 71% (95% CI, 42%–92%) and specificity 92% (95% CI, 86%–96%). Of the 12 false positives, all were Ultra semi-quantitation category trace. All tested air swab Ultra results were negative. Seventeen percent (2/12) were Ultra flocked tongue swab positive.

In buccal swabs and periodontal brushes, no unsuccessful results occurred. Sensitivity was 7% (95% CI, 0%–34%) and 14% (95% CI, 2%–43%), respectively, with 98% (95% CI, 92%–100%) specificity for both (Supplementary Table 2).

In tongue swabs, 1% (2/156) of single swabs and 1% (1/122; $P = .711$) of double swabs generated unsuccessful results. In head-to-head analyses, single versus double swab sensitivity was 67% (95% CI, 30%–93%) and 22% (95% CI, 3%–60%; $P = .068$) and specificity was 82% (95% CI, 74%–89%) versus 96% (95% CI, 91%–99%; $P = .009$) (Table 3). Alternative evidence of TB occurred in 23% (6/26) of false-positive single flocked swabs (2 sputum Ultra-positive, 1 Alere Determine TB LAM Ag [Abbott, Lake Forest, Illinois, USA] positive, 3 positive MGIT960 tongue swab cultures) and 25% (1/4) of false-positive double swabs (sputum Ultra positive). Different trace recategorization strategies resulted in small sensitivity decreases and large specificity increases (Supplementary Table 3).

Fifty-eight percent (150/258) of people were asymptomatic and, of these, 6% (9/150) had TB. Thirty-three percent (3/9), 22% (2/9), and 0% (0/9) were positive using OW, single flocked swab, and double flocked swabs, respectively.

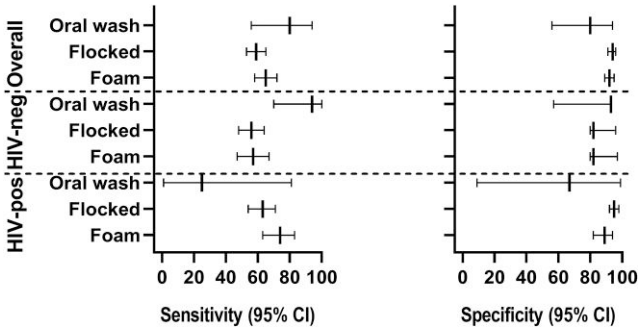

Figure 2. Forest plot of Xpert MTB/RIF Ultra sensitivity and specificity on different oral samples in cohort A, overall and stratified by human immunodeficiency virus status. Oral washes, foam swabs, and flocked swabs had the highest sensitivity. Abbreviations: CI, confidence interval; HIV-neg, human immunodeficiency virus negative; HIV-pos, human immunodeficiency virus positive.

Ultra Oral Sample Performance by Sputum Bacillary Load

In both cohorts, tongue swab sensitivity was highest in people whose Ultra sputum semi-quantitation categories were highest (Table 4). In cohort A, the proportion of people with a high, medium, low, very low, or trace sputum semi-quantitation category detected as positive using flocked or foam swabs was 68% (95% CI, 57%–77%) versus 73% (95% CI, 61%–83%; $P = .475$), 63% (95% CI, 47%–78%) versus 100% (95% CI, 77%–100%; $P = .011$), 56% (95% CI, 42%–58%) versus 77% (95% CI, 60%–89%; $P = .05$), 23% (95% CI, 12%–41%) versus 41% (95% CI, 23%–62%; $P = .167$), and 7% (95% CI, 1%–26%) versus 15% (95% CI, 4%–42%; $P = .390$) (Figure 4, Supplementary Tables 6–8). In sputum Ultra-positive people with a sputum semi-quantitation of "low" or less, 37% (95% CI, 29%–46%) of people were Ultra flocked swab positive versus 53% (95% CI, 41%–64%; $P = .04$) for foam swabs. For OWs in both cohorts and the other sampling approaches in cohort B, proportion Ultra oral sample positive by sputum semi-quantitation grade is in Supplementary Tables 9–11.

Diagnostic Accuracy of Tongue Swab Culture in Cohort B

MGIT960 and TiKa had 64% (95% CI, 35%–87%) versus 36% (95% CI, 13%–65%; $P = .131$) sensitivity, while specificity was 88% (95% CI, 82%–93%) versus 94% (95% CI, 89%–98%; $P = .060$) (Supplementary Table 4). For comparison, sputum TiKa culture sensitivity and specificity were 79% (95% CI,

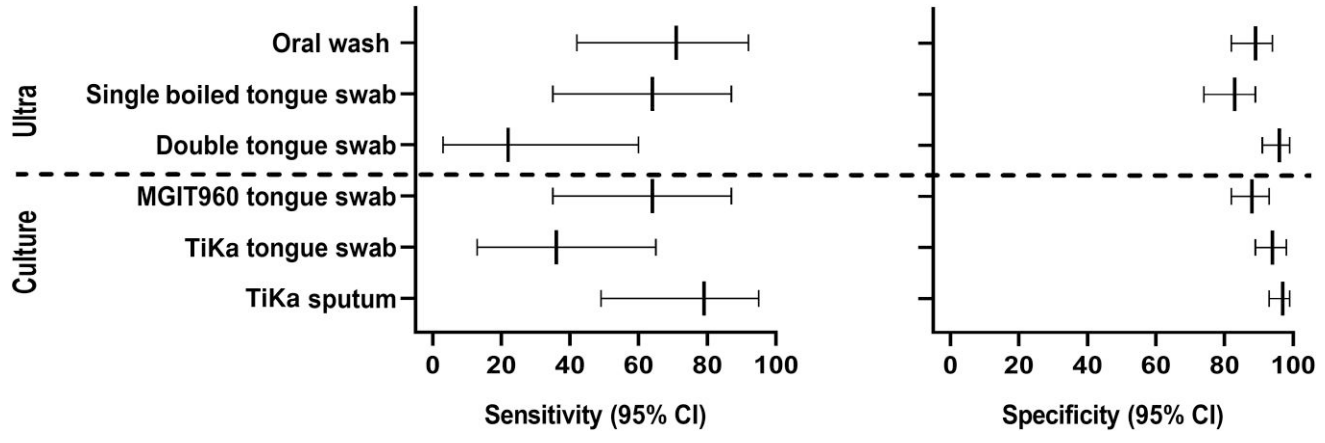

Figure 3. Forest plot of Xpert MTB/RIF Ultra and culture sensitivity and specificity on oral samples in cohort B. Oral washes had the highest sensitivity of oral samples. Abbreviations: CI, confidence interval.

Table 2. Diagnostic Accuracy of Xpert MTB/RIF Ultra on Oral Wash or Tongue Swabs Compared to an Extended Microbiological Reference Standard for the Detection of Tuberculosis Stratified by Human Immunodeficiency Virus Status in Cohort A

	Overall[a] (n = 891)				HIV Negative (n = 526)				HIV Positive (n = 355)			
Method	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
Oral wash	80 (56–94) 16/20	80 (56–94) 16/20	80 (56–94) 16/20	80 (56–94) 16/20	94 (70–100) 15/16	93 (57–96) 14/17	83 (59–96) 15/18	93 (68–100) 14/15	25 (1–81) 1/4 ***P* = .013***	67 (9–99) 2/3 *P* ≤ .999*	50 (1–99) 1/2 *P* ≤ .999*	40 (5–85) 2/5 ***P* = .048***
Flocked swab	59 (53–65) 164/277	94 (91–96) 515/549	83 (77–88) 164/198	82 (79–85) 515/628	56 (48–64) 84/150	82 (90–96) 307/329	79 (70–87) 84/106	82 (78–86) 307/373	63 (54–71) 78/124 *P* = .247*	95 (92–98) 204/214 *P* = .330*	89 (80–94) 78/88 *P* = .079*	82 (76–86) 204/250 *P* = .822*
Foam swab	65 (58–72) 123/188 *P* = .176**	92 (89–95) 321/349 *P* = .292**	81 (74–87) 123/151 *P* = .740**	83 (79–87) 321/386 *P* = .639**	57 (47–67) 58/101 *P* = .823**	82 (90–97) 197/209 *P* = .661**	83 (72–91) 58/70 *P* = .553**	82 (77–87) 197/240 *P* = .944**	74 (63–83) 62/84 ***P* = .020*** *P* = .100**	89 (82–94) 119/134 *P* = .067* ***P* = .022****	81 (70–89) 62/77 *P* = .715* *P* = .147**	84 (77–90) 119/141 *P* = .562* *P* = .484**

P values in bold are significant. Data are presented as % (95% confidence interval) and no./No.

Abbreviations: HIV, human immunodeficiency virus; NPV, negative predictive value; PPV, positive predictive value.

[a]Ten people had unknown HIV status.

*Within-row *P* values: HIV negative vs HIV positive.

**Within-column *P* values: foam vs flocked swab.

Table 3. Diagnostic Accuracy of Xpert MTB/RIF Ultra on Oral Wash or Tongue Swabs Compared to a Double Sputum Culture as an Extended Microbiological Reference Standard for Tuberculosis Detection in Cohort B

	Head-to-Head (n = 122)				Non–Head-to-Head (n = 156)			
Method	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
Oral wash	56 (21–86) 5/9	89 (82–94) 101/113	29 (10–56) 5/17	96 (91–99) 101/105	71 (42–92) 10/14 *P* = .435*	92 (86–96) 130/142 *P* = .556*	45 (24–68) 10/22 *P* = .307*	97 (93–99) 130/134 *P* = .725*
Single boiled tongue swab	67 (30–93) 6/9 *P* = .065**	82 (74–89) 91/110 *P* = .151**	23 (9–44) 6/26 *P* = .695**	97 (91–99) 91/94 *P* = .864**	64 (35–87) 9/14 *P* = .907* *P* = .705**	81 (74–87) 114/140 *P* = .79* ***P* = .018****	26 (12–43) 9/35 *P* = .813* *P* = .143**	96 (90–99) 114/119 *P* = .700* *P* = .609**
Double tongue swab	22 (3–60) 2/9 *P* = .160** *P* = .068***	96 (91–99) 108/112 ***P* = .042**** ***P* = .009*****	33 (4–78) 2/6 *P* = .858** *P* = .639***	94 (88–98) 108/115 *P* = .480** *P* = .395***	22 (3–60) 2/9 ***P* = .022**** ***P* = .048*****	96 (91–99) 108/112 *P* = .117** ***P* = .0004*****	33 (4–78) 2/6 *P* = .595** *P* = .729***	94 (88–98) 108/115 *P* = .247** *P* = .526***

P values in bold are significant. Data are presented as % (95% confidence interval) and no./No.

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

*Within-row *P* values: head-to-head vs non–head-to-head.

**Within-column *P* values: vs oral wash.

***Within-column *P* values: vs single boiled swab.

Table 4. Sensitivity of Tongue Swab Xpert MTB/TIF Ultra Relative to Ultra Sputum Semi-quantitative Results

Sputum Semi-quantitative Results	Cohort A			
	Sensitivity, Flocked		Sensitivity, Foam	
	No.	% (95% CI)	No.	% (95% CI)
High	54/72	75 (63–84)	41/56	73 (60–84)
Medium	21/31	68 (49–83)	13/13	100 (75–100)
Low	31/50	62 (47–75)	24/31	77 (59–90)
Very low	10/34	29 (15–47)	9/22	41 (21–64)
Trace	1/17	6 (0–29)	2/13	15 (2–45)
Total	160/266	60 (54–66)	120/180	67 (59–74)
	Cohort B			
	Sensitivity, Single Heated Flocked		Sensitivity, Double Flocked	
	No.	% (95% CI)	No.	% (95% CI)
High	…	…	…	…
Medium	3/3	100 (29–100)	0/1	0 (0–97)
Low	1/3	33 (1–91)	1/2	50 (1–99)
Very low	3/4	75 (19–99)	1/4	25 (1–81)
Trace	3/10	30 (7–65)	1/8	13 (0–53)
Total	10/20	50 (27–73)	3/15	20 (4–48)

As sputum bacillary load decreased, tongue swab Ultra sensitivity decreased.

Abbreviations: CI, confidence interval.

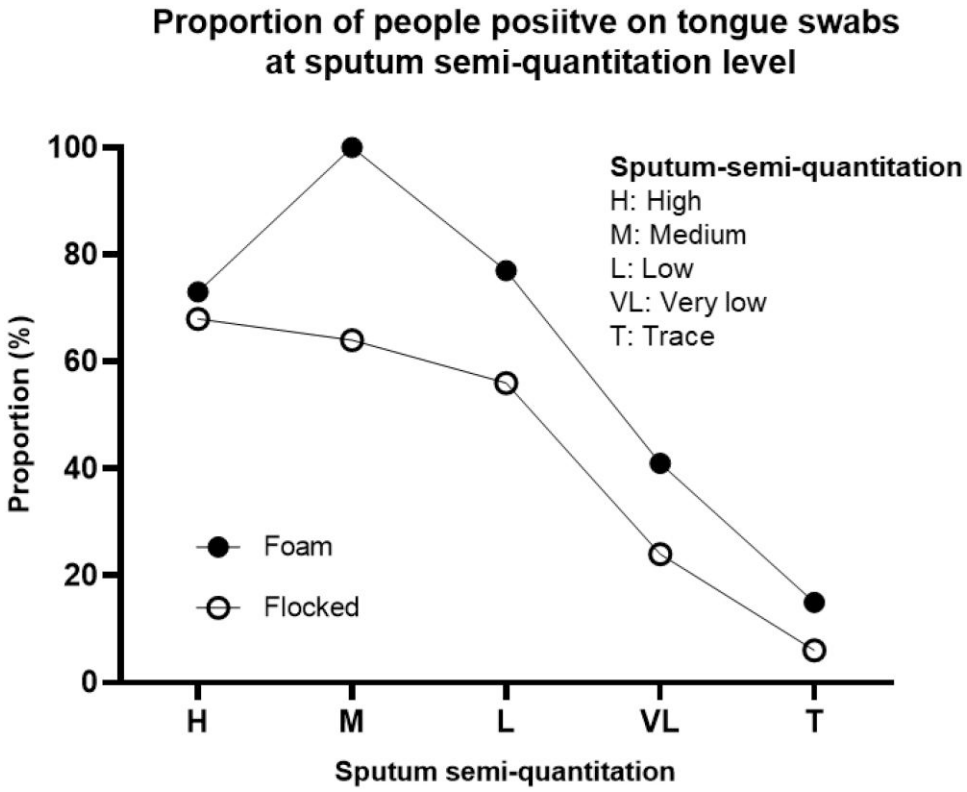

Figure 4. The proportion of people positively detected by Xpert MTB/RIF Ultra on tongue swabs at specific sputum semi-quantitation levels. Foam swabs were more adept at detecting people with lower sputum bacillary load.

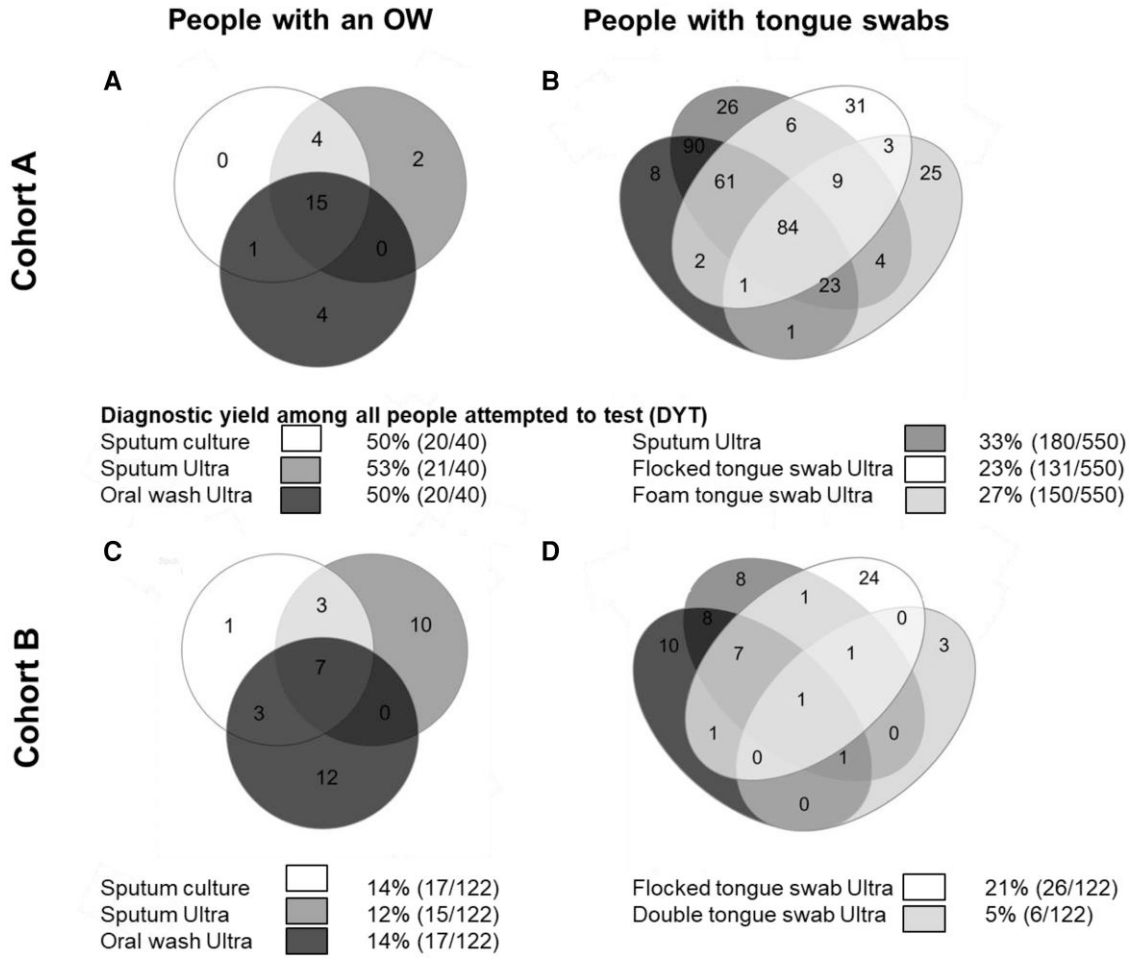

Figure 5. Number of people with a positive result in cohorts *A* (part figure A and B) and *B* (part figure C and D). The first column includes people with an oral wash (OW), whereas the second includes people with tongue swabs (in cohort *B*, OWs and tongue swabs were done in the same people but are disaggregated here for clarity). In both cohorts, more people were detected exclusively using OWs than any other approach. In comparisons of different tongue swab approaches, foam tongue swabs detected the most people. Abbreviations: DYT, diagnostic yield in those tested; OW, oral wash.

49%–95%; $P = .022$ vs TiKa on tongue swabs) and 97% (95% CI, 93%–99%; $P = .238$), respectively. In the subset who underwent tongue swab early bactericidal activity culture, sensitivity and specificity were 33% (95% CI, 4%–78%) and 95% (95% CI, 85%–99%), respectively.

Yield of Ultra

Figure 5 shows people who tested positive by Ultra on different samples, as well as sputum culture. Ultra yield metrics are compared in Supplementary Table 5. Similar patterns for DYD occurred as those described for DYT. We also calculate yields from tongue swab culture (Supplementary Text p. 4), which were low.

Cohort A

Overall, OW DYT (50% [20/50]) was like that for sputum. Foam tongue swabs had a higher DYT (27% [150/550]) point estimate than flocked tongue swabs (23% [131/150]), but this did not reach significance. Sputum DYT was significantly higher (33% [180/550]; $P < .0001$ and $P = .048$) than flocked and foam swabs, respectively.

No one with an OW had sputum induction information. In people with both tongue swab types, 5% (28/550) could not expectorate sputum. Twenty-five percent (7/28) and 25% (7/28) were foam or flocked tongue swab positive, respectively. If tongue swabs were done in people who could not expectorate sputum, people bacteriologically diagnosed rapidly would increase from 175 (Ultra positive on expectorated sputum) to 182 for flocked and foam swabs (4% [2%–6%] increase).

Cohort B

Overall, among people with all 3 sample types, DYT point estimates were highest for single flocked tongue swab compared to OWs and double tongue swabs (21% [26/122] vs 14% [17/122] and 5% [6/122]). Sputum DYT was 12% (5/122).

In people who had all 3 sample types (OWs, single and double swabs), 35% (43/122) could not expectorate sputum, of

which 16% (7/43), 26% (11/43), and 5% (2/43) were positive on each oral sample type. If oral sampling was done in people who could not expectorate sputum, bacteriological diagnoses would change from 8 (Ultra-positive on expectorated sputum) to 15, 19, and 10 for OWs, single swabs, and double swabs, respectively (+88% [82%–93%], +138% [130%–145%], and 25% [17%–33%], respectively). If induction were unavailable, sputum DYT decreased to 7% (8/122), lower than OWs (14% [17/122]; $P = .057$) and single flocked swabs (21% [26/122]; $P = .0009$).

Ultra Inhibition

In cohort A, oral washes had less polymerase chain reaction (PCR) inhibition than sputum; however, this did not occur in cohort B. No other differences of note occurred across sample types (Supplementary Text p. 4).

DISCUSSION

Our key findings are that (1) OW Ultra had the highest sensitivity among the methods tested (71%–80%); (2) Ultra on foam swabs had higher sensitivity than flocked swabs (65% vs 59% in cohort A) and was more adept at detecting people with lower sputum bacillary load; (3) Ultra on oral samples swabs diagnosed TB in many people who could not naturally expectorate, permitting yield to exceed that of sputum-based testing in cohort B; and (4) other approaches (Ultra on double tongue swabs, buccal swabs, and periodontal brushes) were suboptimal. These data suggest that oral sampling, especially oral washes and foam swabs, can improve TB diagnosis, especially when sputum scarcity is accounted for.

OW Ultra had a higher sensitivity point estimate than other approaches. To our knowledge, we report the first study using OW Ultra. Previous studies [30, 31] used in-house PCR methods applied to OWs and reported sensitivities of 77% and 88%, broadly like our findings. We suggest that OWs be included as a comparator in research on tongue swabs for TB diagnosis going forward as, unlike swabs, they require less processing.

We observed a small sensitivity increase with foam swabs compared to flocked swabs, likely attributable to the larger amount of biomass bound to the foam swab [15, 32]. This added benefit of foam swabs was most pronounced in people with lower sputum load. Foam swabs have an added advantage in that they are relatively cheap (0.14 US dollar per swab) compared to flocked swabs. On a practical note, however, participants report that foam swabs make their tongues feel "dry." Furthermore, the lack of a breakpoint means swab heads need to be cut off; however, this is addressable. The magnitude of the sensitivity improvement from foam swabs may, versus flocked swabs, be even greater in people with earlier-stage disease.

We showed that Ultra on swabs can detect TB in people who cannot make sputum. As TB testing expands to people with early-stage subclinical TB, the proportion of people with sputum-scarce TB who require testing will increase. Thus, our study addresses a key gap: Most studies on oral samples for TB diagnosis have either not recruited people with sputum-scarce TB or offered induction and not disaggregated performance in people who, without induction, cannot expectorate. This is important because, even if tongue swabs perform well in sputum expectorators, it is hard to justify not testing sputum if available. Our data are thus novel in that they give performance data in the type of person most likely to benefit from nonsputum tests. For example, in cohort B, the number of people with a positive Ultra result would at least double with the use of oral sampling.

We compared a single flocked swab with heat lysis, which had high sensitivity in earlier work using contrived samples [33], to a double flocked swab without heat lysis, which had lower sensitivity. This comparison involves 2 different sampling approaches that include differences other than just an extra swab (heat lysis) and suggests that heating is critical to release DNA. It remains to be seen if heating with a double swab improves sensitivity further; however, given our results from foam swabs, it is likely that steps that increase input material would be beneficial. In addition to the double swab method, we evaluated other specimen types (buccal swabs, periodontal brushes) and different tongue swab culture methods; however, neither was promising.

Although eMRS was used in each cohort and air controls (all of which were negative), we found specificities slightly less than those previously reported for other tongue swab–based approaches. Furthermore, OW specificity was relatively low, although some OW false-positives were independently treated empirically. While some of this may be due to small sample size (and hence requires prospective validation in larger studies), this could be due to laboratory errors, reference standard limitations, *Mtb* colonization, or paucibacillary disease. Such future work should include longitudinal follow-up of people who test "false positive."

Our study has strengths and limitations. Recruitment was programmatic in that it included people with symptoms and non-symptom-based risk factors for which South African guidelines [34, 35] require molecular testing. More data are needed from asymptomatic people [7]. Participants did not all receive the same oral sampling methods due to evolving protocols, logistical challenges, coronavirus disease 2019–related delays, and early discontinuation of less promising methods like buccal and periodontal swabs. Although this variation is a limitation, it is partly mitigated through head-to-head comparisons within cohorts. In cohort B, performance differences between tongue swab methods may stem from both processing steps and not just swab number. The timing of sample collection (eg, OWs vs tongue swabs) may affect results and should be explored further. Future studies should use larger sample sizes, particularly for OWs, and apply the most promising sampling approaches in a staggered, head-to-head design. Different processing methods, such as those which maximize time to

recover DNA (buccal swabs and periodontal brushes were in buffer for short periods) should be explored. Last, the inclusion of people who cannot make sputum naturally and the availability of induction information are strengths of the study, as well as the double culture reference standard in cohort B.

In conclusion, Ultra on oral samples—especially foam tongue swabs and OW—is sensitive and highly specific and can significantly increase the overall number of people with a rapid positive bacteriological result when applied to people who cannot naturally make sputum.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. The authors thank the National Health Laboratory Services, Cape Town, South Africa; members of the Clinical Mycobacteriology and Epidemiology (CLIME) research group; and the study participants.

Disclaimer. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the South African Medical Research Council (SAMRC). Cepheid donated cartridges used in the study but did not have a role in study design or result interpretation.

Financial support. This work was supported by the South African Medical Research Council, National Research Foundation, through its Division of Research Capacity Development under the SAMRC Internship Scholarship Programme. The postgraduate bursary department of Stellenbosch University is hereby also acknowledged for funding this project. L. R. acknowledges funding from the SAMRC. G. T. reports funding from the EDCTP2 program supported by the European Union (grant numbers RIA2018D-2509, PreFIT; RIA2018D-2493, SeroSelectTB; RIA2020I-3305, CAGE-TB) and the National Institutes of Health (grant numbers D43TW010350, U01AI152087, U54EB027049, and R01AI136894).

Potential conflicts of interest. The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. World Health Organization (WHO). Global tuberculosis report 2023. Geneva, Switzerland: WHO, **2023**.
2. World Health Organization (WHO). Target product profile for tuberculosis diagnosis and detection of drug resistance. Geneva, Switzerland: WHO, **2024**.
3. 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**; 78:103939.
4. Abdulgader SM, Okunola AO, Ndlangalavu G, et al. Diagnosing tuberculosis: what do new technologies allow us to (not) do? Respiration **2022**; 101:797–813.
5. Wood RC, Luabeya AK, Dragovich RB, et al. Diagnostic accuracy of tongue swab testing on two automated tuberculosis diagnostic platforms, Cepheid Xpert MTB/RIF Ultra and Molbio Truenat MTB Ultima. J Clin Microbiol **2024**; 62:e0001924.
6. Shah K, Nalugwa T, Marcelo D, et al. A multi-country evaluation of patient preferences for tuberculosis diagnostics: a discrete choice experiment to inform WHO target product profiles. medRxiv [Preprint]. Posted online 22 June 2024. doi:10.1101/2024.06.19.24309124
7. de Nooy A, Ockhuisen T, Korobitsyn A, et al. Trade-offs between clinical performance and test accessibility in tuberculosis diagnosis: a multi-country modelling approach for target product profile development. Lancet Glob Health **2024**; 12:e1139–48.
8. Broger T, Marx FM, Theron G, et al. Diagnostic yield as an important metric for the evaluation of novel tuberculosis tests: rationale and guidance for future research. Lancet Glob Health **2024**; 12:e1184–91.
9. Bezuidenhout C, Long L, Nichols B, et al. Using sputum and tongue swab specimens for in-home point-of-care targeted universal testing for tb of household contacts: an acceptability and feasibility analysis. medRxiv [Preprint]. Posted online 2 November 2024. doi:10.1101/2024.11.01.24316570
10. Valinetz ED, Cangelosi GA. A look inside: oral sampling for detection of non-oral infectious diseases. J Clin Microbiol **2021**;59:e0236020.
11. Luabeya AK, Wood RC, Shenje J, et al. Noninvasive detection of tuberculosis by oral swab analysis. J Clin Microbiol **2019**; 57:e01847-18.
12. Nicol MP, Wood RC, Workman L, et al. Microbiological diagnosis of pulmonary tuberculosis in children by oral swab polymerase chain reaction. Sci Rep **2019**; 9:10789.
13. Wood RC, Luabeya AK, Weigel KM, et al. Detection of *Mycobacterium tuberculosis* DNA on the oral mucosa of tuberculosis patients. Sci Rep **2015**; 5:8668.
14. Church EC, Steingart KR, Cangelosi GA, Ruhwald M, Kohli M, Shapiro AE. Oral swabs with a rapid molecular diagnostic test for pulmonary tuberculosis in adults and children: a systematic review. Lancet Glob Health **2024**; 12:e45–54.
15. Wood RC, Andama A, Hermansky G, et al. Characterization of oral swab samples for diagnosis of pulmonary tuberculosis. PLoS One **2021**; 16:e0251422.
16. Reeve BW, 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**; 12:e793–803.
17. Mann T, Gupta RK, Reeve BW, et al. Blood RNA biomarkers for tuberculosis screening in people living with HIV before antiretroviral therapy initiation: a diagnostic accuracy study. Lancet Glob Health **2024**; 12:e783–92.
18. 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. Lancet Respir Med **2020**; 8:368–82.
19. Cepheid Inc. Xpert MTB/RIF Ultra product brochure. Sunnyvale, CA: Cepheid Inc, **2017**.
20. Goosen WJ, Kleynhans L, Kerr TJ, et al. Improved detection of *Mycobacterium tuberculosis* and *M. bovis* in African wildlife samples using cationic peptide decontamination and mycobacterial culture supplementation. J Vet Diagn Invest **2022**; 34:61–7.
21. Diacon AH, Donald PR. The early bactericidal activity of antituberculosis drugs. Expert Rev Anti Infect Ther **2014**; 12:223–37.
22. Ghebrekristos YT, Beylis N, Centner CM, et al. Xpert MTB/RIF Ultra on contaminated liquid cultures for tuberculosis and rifampicin-resistance detection: a diagnostic accuracy evaluation. Lancet Microbe **2023**; 4:e822–9.
23. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open **2016**; 6:e012799.
24. StataCorp LLC. prtest—tests of proportions. Available at: https://www.stata.com/manuals/rprtest.pdf. Accessed 22 October 2023.
25. McDonald JH. Handbook of biological statistics. 3rd edn. Baltimore, MD: Sparky House Publishing, **2014**.
26. Rudolf F, Lemvik G, Abate E, , et al. TBScore II: refining and validating a simple clinical score for treatment monitoring of patients with pulmonary tuberculosis. Scand J Infect Dis **2013**; 45:825–36.
27. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics **2015**; 16:169.
28. World Health Organization (WHO). WHO consolidated guidelines on tuberculosis. Module 2: screening—systematic screening for tuberculosis disease. Geneva, Switzerland: WHO, **2021**.
29. Blakemore R, Nabeta P, Davidow AL, et al. A multisite assessment of the quantitative capabilities of the Xpert MTB/RIF assay. Am J Respir Crit Care Med **2011**; 184:1076–84.
30. Davis JL, Huang L, Kovacs JA, et al. Polymerase chain reaction of secA1 on sputum or oral wash samples for the diagnosis of pulmonary tuberculosis. Clin Infect Dis **2009**; 48:725–32.

31. Evans D, Goyal M, Taylor IK, Shaw RJ. Identification of *M. tuberculosis* ribosomal RNA in mouthwash samples from patients with tuberculosis. Respir Med **1994**; 88:687–91.
32. Olson AM, Wood RC, Weigel KM, et al. High-sensitivity detection of *Mycobacterium tuberculosis* DNA in tongue swab samples. J Clin Microbiol **2025**; 63:e0114024.
33. Andama A, Reza T, Jaganath D, et al. Accuracy of tongue swab testing using Xpert MTB-RIF Ultra for tuberculosis diagnosis. J Clin Microbiol **2022**; 60:e00421–22.
34. Martinson NA, Nonyane BA, Genade LP, et al. Evaluating systematic targeted universal testing for tuberculosis in primary care clinics of South Africa: a cluster-randomized trial (The TUTT Trial). PLoS Med **2023**; 20:e1004237.
35. Department of Health, Republic of South Africa. TB Screening and Testing Standard Operating Procedure. Pretoria: National Department of Health. 2022. Available from: https://sahivsoc.org/Files/NDoH_TB SCREENING AND TESTING SOP 2022.pdf. Accessed 30 July 2025.