Performance of Xpert MTB/RIF Ultra for tuberculosis diagnosis in the context of passive and active case finding

Belén Saavedra[1,2], Edson Mambuque[1], Dinis Nguenha[1], Neide Gomes[1], Shilzia Munguambe[1], Juan Ignacio García[3], Santiago Izco[1,2], Sozinho Acacio[1], Adrià Murias-Closas[2], Marta Cossa[1], Irene Losada[2], Hadrián Pernas-Pardavila[4], Laura Oliveras[2,5,6], Grant Theron[7] and Alberto L. García-Basteiro[1,2]

[1]Centro de Investigação em Saúde de Manhiça (CISM), Maputo, Mozambique. [2]ISGlobal, Hospital Clínic – Universitat de Barcelona, Barcelona, Spain. [3]TB Group, Population Health Programme, Texas Biomedical Research Institute, San Antonio, TX, USA. [4]Infectious Diseases, Complexo Hospitalario Universitario de Santiago, Santiago de Compostela, Spain. [5]Agència de Salut Pública de Barcelona, Barcelona, Spain. [6]Institut d'Investigació Biomèdica Sant Pau (IIB Sant Pau), Barcelona, Spain. [7]DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research and SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa.

Corresponding author: Alberto L. García-Basteiro (alberto.garcia-basteiro@manhica.net)

Shareable abstract (@ERSpublications)
Ultra has higher sensitivity and lower specificity than Xpert for passive case finding. Lower specificity estimates do not seem to be driven by false-positive results among previously treated participants. https://bit.ly/3wcsg62

Cite this article as: Saavedra B, Mambuque E, Nguenha D, *et al*. Performance of Xpert MTB/RIF Ultra for tuberculosis diagnosis in the context of passive and active case finding. *Eur Respir J* 2021; 58: 2100257 [DOI: 10.1183/13993003.00257-2021].

This article has supplementary material available from erj.ersjournals.com

Received: 27 Jan 2021
Accepted: 11 May 2021

Abstract

Aims We present a field evaluation of the diagnostic accuracy of Xpert MTB/RIF ("Xpert") and Xpert MTB/RIF Ultra ("Ultra") using two cohorts in a high tuberculosis/HIV burden setting in Southern Mozambique.

Methods Single respiratory specimens from symptomatic adults accessing healthcare services (passive case finding (PCF) cohort) and from household and community close contacts (active case finding (ACF) cohort) were tested by smear microscopy, culture, Xpert and Ultra. Liquid and solid culture served as a composite reference standard. We explored the impact of trace results on specificity *via* their recategorisation to negative (in all and just among those previously treated individuals).

Results 1419 and 252 participants were enrolled in the PCF and ACF cohorts, respectively. For the PCF cohort, Ultra showed higher sensitivity than Xpert overall (0.95 (95% CI 0.90–0.98) *versus* 0.88 (96% CI 0.82–0.93); p<0.001) and among smear-negative patients (0.84 (96% CI 0.71–0.93) *versus* 0.63 (96% CI 0.48–0.76)). Ultra's specificity was lower than Xpert's (0.96 (96% CI 0.95–0.97) *versus* 0.98 (96% CI 0.97–0.99); p=0.008). For ACF, sensitivities were the same (0.67 (95% CI 0.22–0.96) for both tests), although Ultra detected a higher number of microbiologically confirmed samples than Xpert (4.7% (12 out of 252) *versus* 2.7% (seven out of 252)). Conditional recategorisation of trace results among previously treated participants maintained differences in specificity in the PCF cohort.

Conclusion These results add evidence on the improved sensitivity of Ultra and support its use in different case finding scenarios.

Introduction

Tuberculosis (TB) is the leading cause of death by a single pathogen worldwide. In 2019, 1.2 million people died from TB among HIV-negative patients, 208 000 among HIV-positive patients and 10 million were estimated to have fallen sick, although only 70% were diagnosed [1]. The gap between TB notifications and estimated existing cases is still unacceptable. The low TB case detection rate in many high-burden settings urgently demands effective strategies and tools to identify TB patients and ensure their linkage to healthcare [2].

Most national plans for TB control are built based on passive case finding (PCF) strategies (health facility-based diagnosis) which mostly detect the most severe TB cases, missing many others in the community [3]. The End TB Strategy goals will not be achieved without the implementation of active case finding (ACF) strategies [4]. Systematic screening for those at higher risk of exposure (close contacts) or those who do not experience typical symptoms (*e.g.* people living with HIV) has been widely recommended [1, 3, 5]. Broad access to highly sensitive TB assays may minimise delays in diagnosis by identifying patients in early stages of the disease [6, 7]. Nevertheless, evaluations on the added yield of applying molecular tests in ACF approaches are limited [8].

In 2010, Xpert MTB/RIF ("Xpert"; Cepheid, Sunnyvale, CA, USA) arose as a promising molecular test with substantial improved sensitivity compared with traditional smear microscopy. Although it rapidly became the front-line test for TB investigation in most high-burden settings [9, 10], Xpert showed suboptimal sensitivity in paucibacillary cases and concerns about its ability to detect certain rifampicin silent mutations [11]. To overcome those limitations, in 2017, Xpert Ultra MTB/RIF ("Ultra"; Cepheid) appeared as a next-generation assay. It uses larger reaction volumes and the multicopy specific genes (IS6110 and IS1081) of *Mycobacterium tuberculosis* as new targets to increase DNA detection (limit of detection 15.6 $CFU \cdot mL^{-1}$), adding a new category for the lowest burden of DNA, "trace call". Results on its analytical evaluation are currently on the rise [12–15]. Comparative studies showed better diagnostic sensitivity of Ultra compared with Xpert, lower specificity, and similar figures for sensitivity and specificity in rifampicin resistance detection [16]. However, there are few studies comparing both tests, and they show high heterogeneity in study design, setting characteristics, samples and reference standard, as well as uncertainty about the interpretation of trace results [17–19]. For this reason, the recent World Health Organization (WHO) consolidated guidelines for TB diagnosis recommend both Xpert and Ultra as front-line diagnostic tools for ruling out TB [5].

We conducted a field study to assess the diagnostic accuracy of Ultra and Xpert in a high HIV/TB burden setting using the same single specimen. We evaluated the performance of Ultra in two separate populations: a PCF (healthcare centre-based) cohort and an ACF (field-based) cohort. We hypothesised that Ultra would have different diagnostic performance depending on the programmatic approach used for TB diagnosis.

Methods

Study design

This is a prospective cross-sectional diagnostic accuracy evaluation of Xpert and Ultra in a high TB/HIV burden area in Southern Mozambique [20, 21]. The study was conducted in the District of Manhiça, Maputo Province, a semirural area 80 km away from the capital of Maputo, with a population of approximately 201 845 inhabitants living in 46 441 households [22].

From November 2017 to March 2019 the whole district participated in the Xpatial-TB study (TB REACH, wave 5, funded by the Stop TB Partnership). This project introduced an innovative ACF initiative which stratified contact tracing activities depending on spatial and microbiological parameters of index TB cases. It was implemented by the Centro de Investigação em Saúde de Manhiça (CISM; Maputo, Mozambique) in coordination with the Ministry of Health (supplementary material). Nested in the Xpatial-TB study, the field evaluation of Xpert and Ultra was conducted, where samples from both the ACF intervention and the routine PCF diagnostic pathway of the National Tuberculosis Programme were tested from November 2017 to June 2018.

Sample collection

PCF cohort

All samples from presumptive TB adults (those presenting suggestive TB symptoms, as per international definitions [23]), who attended any of the health units or hospitals in the district, were consecutively recruited. As part of the national TB diagnostic work-up, participants were instructed to deliver one early morning sputum sample to the reference health centre (based on proximity) on the following day. Samples were sent to the CISM TB BSL3 laboratory (ISO 15789 certified). A map of the District of Manhiça and participating health facilities, as well as the laboratory flowchart, is provided in the supplementary material.

ACF cohort

Sample collection was based on the procedures of the Xpatial-TB study. Briefly, throughout the Health and Demographic Surveillance System of Manhiça, every index case who started treatment in the district and lived in the study area generated a list of contacts for TB investigation [24]. After informed consent, contacts filled out a study questionnaire and following the testing algorithm (supplementary material), spot

sputum samples were collected in the field. Sputum induction was used at participants' households for those symptomatic or HIV-positive, but unable to provide spontaneous sputum, by administrating one inhalation of salbutamol (100 µg) followed by 15 min of nebulisation with hypertonic saline solution (5%) with portable nebulisers (MicroAIR U22; OMRON, Kyoto, Japan).

STARD (Standards for Reporting of Diagnostic Accuracy Studies) guidelines were used as reference for the study design and analysis.

Laboratory procedures

All participants were requested to provide one sputum specimen with a minimum volume of 2 mL.

For smear microscopy, sputum specimens were stained by the Ziehl–Neelsen method and results were reported as negative or on a scale of positive grades according to international standards [25].

For Xpert and Ultra, raw samples were tested according to the manufacturer's instructions. Invalid results were excluded. The semiquantitative results for Ultra fell under the following categories: trace, very low, low, medium or high. For Xpert the categories were: very low, low, medium or high.

For solid and liquid culture, remnant raw samples were decontaminated by the modified Kubica method [26]. Afterwards, 500 µL was inoculated into Mycobacteria Growth Indicator Tube (MGIT) liquid medium and incubated in a BACTEC MGIT 960 mycobacterial detection instrument (BD, Franklin Lakes, NJ, USA). Additionally, 200 µL was cultured in BD Löwenstein–Jensen (LJ) solid medium. After 42 days (for liquid culture) or 8 weeks (for solid culture) without growth, samples were classified as negative. Incubation times for those with trace call results were doubled. If MGIT culture was contaminated, the decontamination procedure was repeated over the liquid medium. Final contaminated cultures were excluded from the analysis. Nontuberculous mycobacteria growths were considered negative for *M. tuberculosis*.

Phenotypic drug susceptibility tests were conducted on all positive cultures using the BACTEC MGIT 960 system and samples of isolates of positive cultures were stored at −80°C.

Statistical analysis

Statistical analysis was performed using R version 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria).

In order to detect a difference of 5% in sensitivity and 3% in specificity of Ultra compared with Xpert [27], the sample size needed for the PCF cohort was 1450 participants. More details on sample size calculations are provided in the supplementary material. Logical relations of positivity among tests were analysed in Venn diagrams. Results on rifampicin resistance were described and cross-tabulated. The sensitivity, specificity, positive predictive value and negative predictive value for the detection of *M. tuberculosis* complex were calculated on the basis of aggregated culture results (liquid and solid) as reference standard. We used the two-sided McNemar's test statistic for paired data, and its exact version for small groups, to test whether there was a systematic difference between Ultra and Xpert test results. A p-value of 0.05 was considered as the threshold for statistical significance. Test values were calculated overall and stratified by Ziehl–Neelsen smear microscopy, and reported with 95% confidence intervals. Discordant results were defined as samples with a positive molecular test and negative MGIT and LJ culture.

Trace call management and analysis

Trace results are obtained when only the multicopy genes IS6110 and IS1081 are detected by the PCR. Patients who received a trace result were managed following WHO recommendations (clinical evaluation and context assessment), and they were followed-up for between 3 weeks and 6 months after enrolment. To explore the significance of this category, patients with a trace call were theoretically recategorised using two strategies: recategorisation to negative and conditional reclassification for those previously treated individuals, regardless of the lapse of time since last previous treatment.

Bacillary burden analysis

The Ultra cycle threshold (C_t) for the IS1081/IS6110 target and the minimum Xpert C_t (excluding zero when markers of resistance were present) were used as a measure of mycobacterial concentration for correlation assessment. We applied the Kruskal–Wallis test to compare bacillary burden among smear

microscopy and molecular tests, and Spearman's correlation coefficient to assess the correlation of C_t values between Xpert and Ultra.

Ethical considerations

The protocol was approved by CISM's Internal Scientific Committee, CISM's Internal Bioethics Committee and the National Bioethics Committee at the Ministry of Health (369/CNBS/17). All individuals provided written informed consent to participate. The study methods were carried out in accordance with the guidelines and regulations established by the National Bioethics Committee.

Results

Participants

Between November 2017 and June 2018, we initially enrolled a total of 1739 participants (1441 and 298 in the PCF and ACF cohorts, respectively) who met all inclusion criteria for this study (figure 1). After exclusions, 1419 and 252 participants were included in the analyses of the PCF and ACF cohorts, respectively. Overall, participants were more commonly women (ACF 64.7% *versus* PCF 51.4%), the median (interquartile range) age was 39.0 (28.7–52.0) years and the proportion of retreatments was similar for both groups (p>0.05). Data on HIV status was fully available only among those reached through the ACF study and 31% of them were HIV-positive. Table 1 shows the sociodemographic characteristics of participants.

Diagnostic accuracy evaluation and predictive values

Passive case finding

11.1% of suspected patients (157 out of 1419) were positive by culture (solid and/or liquid), 7.9% of samples (112 out of 1419) by smear microscopy, 11.7% (166 out of 1419) by Xpert and 14.2% (201 out of 1419) by Ultra. Results on test parameters are displayed in table 2. From those positive by the reference standard (n=157), Ultra correctly identified 149 (sensitivity 0.95 (95% CI 0.90–0.98)) and Xpert correctly identified 138 (sensitivity 0.88 (95% CI 0.82–0.93); p<0.001). When stratified by smear results, Ultra also showed greater sensitivity among smear-negative patients. Conversely, specificity of Ultra was lower than Xpert in both cases (0.96 (95% CI 0.95–0.97) *versus* 0.98 (95% CI 0.97–0.99); p=0.05) (table 2). Ultra detected 24 more positive cases than Xpert and culture together (11.5% of positive samples), and 35

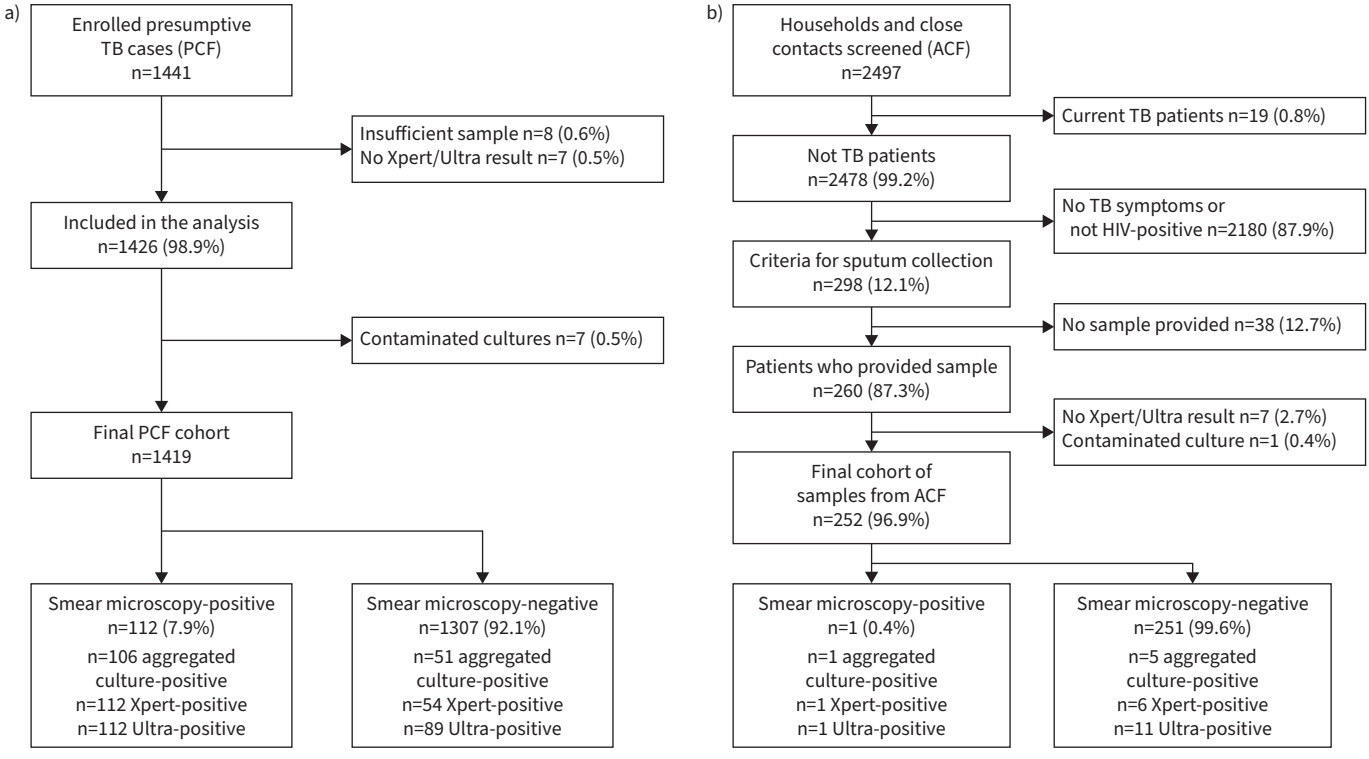

FIGURE 1 Study flowchart for the a) passive case finding (PCF) and b) active case finding (ACF) cohorts. TB: tuberculosis.

TABLE 1 Sociodemographic characteristics of participants: overall data and stratification by cohort

	Passive case finding cohort	Active case finding cohort	Overall
Age, years	39.0 (30.0–53.0)	36.1 (18.5–50.3)	39.0 (28.7–52.0)
Women	51.4 (729/1419)	64.7 (163/252)	53.4 (892/1671)
Retreatment	2.1 (30/1415)[#]	4.0 (10/250)[#]	2.4 (40/1665)[#]
Symptoms	NA	38.0 (99/260)	NA
HIV-positive status		31.0 (78/252)[¶]	

Data are presented as mean (interquartile range) or % (n/N). NA: not applicable. [#]: some participants had missing data for this variable; [¶]: samples were collected in the field from symptomatic contacts or asymptomatic HIV-positive households.

(16.7% of positive samples) more than Xpert alone with 54.3% (19 out of 35) of them being trace results (figure 2). The number of patients needed to test (NNT) *via* PCF to diagnose one bacteriologically confirmed case of TB was 7 (1419/201) using Ultra and 8.5 (1419/166) using Xpert.

Active case finding

2.4% of samples (six out of 252) were positive by culture, 2.8% (seven out of 252) by Xpert and 4.8% (12 out of 252) by Ultra. Only one patient (0.4%) was smear microscopy-positive. Both assays achieved equal overall sensitivity (0.67 (95% CI 0.22–0.96)) and similar values for specificity: Ultra 0.97 (95% CI 0.94–0.99) *versus* Xpert 0.99 (95% CI 0.96–1.00); p=0.06. Full diagnostic parameters can be found in table 2.

TABLE 2 Comparison of smear microscopy, Xpert and Ultra accuracy by cohort and stratified by smear microscopy and trace recategorisation: contingency table showing the distribution of results by tests

	Passive case finding cohort (n=1419)				Active case finding cohort (n=252)			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Smear microscopy	0.68 (0.60–0.75) 106/157	1.00 (0.99–1.00) 1256/1262	0.95 (0.89–0.98) 106/112	0.96 (0.95–0.97) 1256/1307	0.10 (0.0–0.64) 1/6	1.00 (0.99–1.00) 246/246	1.00 (0.02–1.00) 1/1	0.98 (0.95–0.99)
Xpert	0.88 (0.82–0.93) 138/157	0.98 (0.97–0.99) 1234/1262	0.83 (0.77–0.88) 138/166	0.98 (0.98–0.99) 1234/1253	0.67 (0.22–0.96) 4/6	0.99 (0.96–1.00) 243/246	0.57 (0.18–0.90) 4/7	0.99 (0.97–1.00)
Ultra	0.95 (0.90–0.98) 149/157	0.96 (0.95–0.97) 1210/1262	0.74 (0.67–0.80) 149/201	0.99 (0.99–1.00) 1210/1218	0.67 (0.22–0.96) 4/6	0.97 (0.94–0.99) 238/246	0.33 (0.10–0.65)	0.99 (0.97–1.00)
Statistical test[#]	p<0.001	p=0.008			NA	p=0.06		
	Smear-negative (n=1307)				**Smear-negative (n=251)**			
Xpert	0.63 (0.48–0.76) 32/51	0.98 (0.97–0.99) 1234/1256	0.59 (0.45–0.72) 32/54	0.98 (0.98–0.99) 1234/1253	0.60 (0.15–0.95) 3/5	0.99 (0.96–1.00) 243/246	0.50 (0.12–0.88) 3/6	0.99 (0.97–1.00) 243/245
Ultra	0.84 (0.71–0.93) 43/51	0.96 (0.95–0.97) 1210/1256	0.48 (0.38–0.59) 43/89	0.99 (0.99–1.00) 1210/1218	0.60 (0.15–0.95) 3/5	0.97 (0.94–0.99) 238/246	0.27 (0.06–0.61) 3/11	0.99 (0.97–1.00) 238/240
Statistical test[#]	p<0.001	p=0.05			NA	p=0.10		
	Trace recategorisation as negative							
Ultra	0.90 (0.85–0.95) 143/157	0.97 (0.96–0.98) 1228/1262	0.81 (0.74–0.86) 143/177	0.99 (0.98–0.99) 1228/1242	0.67 (0.22–0.96) 4/6	0.99 (0.96–1.00) 243/246	0.57 (0.18–0.90) 4/7	0.99 (0.97–1.00) 243/245
Statistical test[#]	p=0.02	p=0.5			NA	p=1		
	Trace conditional recategorisation[¶]							
Ultra	0.94 (0.89–0.97) 148/157	0.96 (0.95–0.97) 1215/1262	0.76 (0.69–0.82) 148/195	0.99 (0.99–81.00) 1215/1224	0.67 (0.22–0.96) 4/6	0.97 (0.94–0.99) 239/246	0.30 (0.07–0.65) 3/10	0.99 (0.97–1.00) 239/242
Statistical test[#]	p<0.0001	p=0.03			NA	p=0.10		

Data are presented with 95% CI or as n (comparator test)/N (gold standard), unless otherwise stated; absolute numbers for these calculations are given in the supplementary material. The average monthly Mycobacteria Growth Indicator Tube contamination rate during the study period was 4.0%. PPV: positive predictive value; NPV: negative predictive value; NA: not applicable. [#]: McNemar's test for evaluation of differences in test parameters among Xpert and Ultra; [¶]: recategorisation of trace results as negative, if patients had been previously treated.

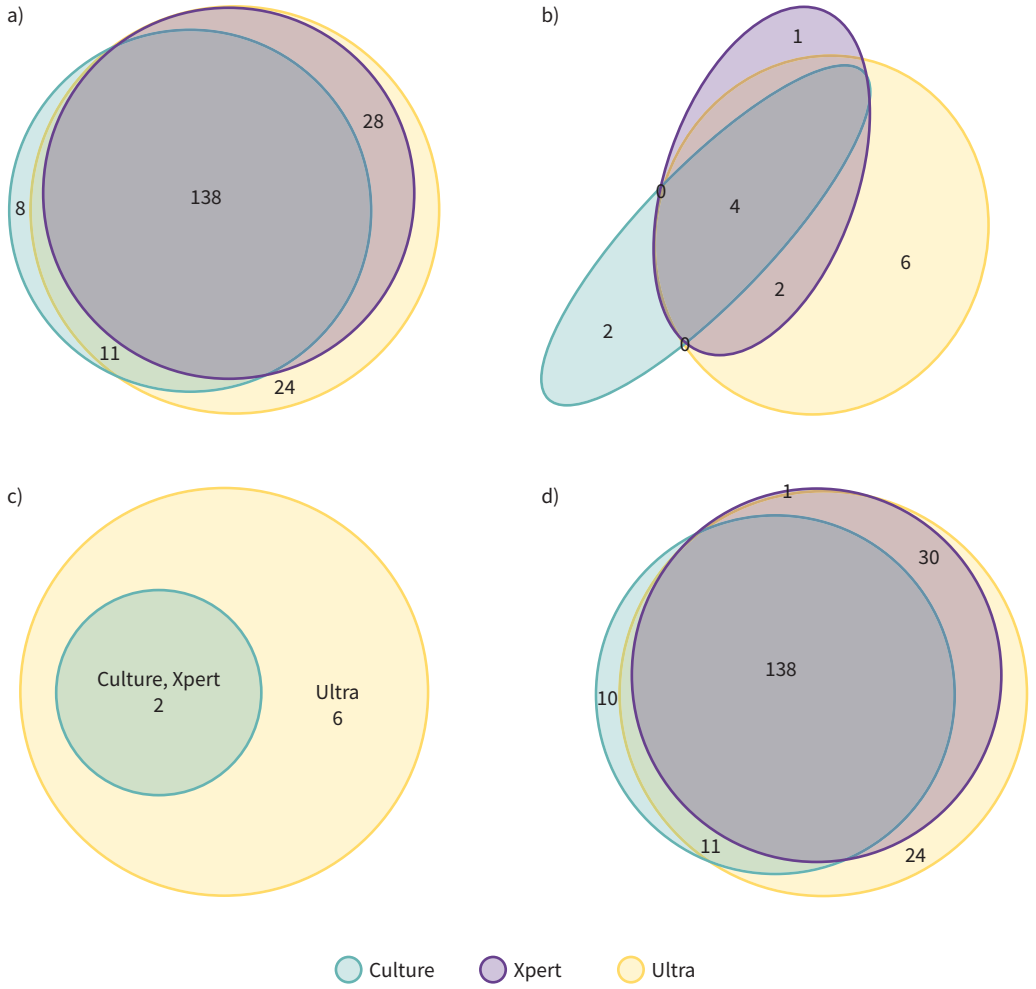

FIGURE 2 Venn diagrams showing logical relations of positive results among culture, Xpert and Ultra for a) passive case finding, b) active case finding, c) previously treated patients and d) new patients.

Ultra detected six additional cases compared with culture (culture detected 50% of those positive by Ultra) (figure 2). Of those additional cases, four (66%) were "trace". The NNT *via* ACF to diagnose one bacteriologically confirmed case was 21 (252/12) using Ultra and 36 (252/7) using Xpert.

Logical relations of positive results

Logical relations of positive results among culture, Xpert and Ultra are shown in figure 2a–d for both cohorts (PCF and ACF), as well as for previously treated and new patients.

Trace reclassification

29 patients obtained trace call results. Only one patient (3.4%) was lost to follow-up and six patients died (20.7%) (four before starting treatment (13.4%) and two during treatment (6.9%)). 22 patients could be re-assessed (75.9%), of whom 13 (59.1%) were HIV-positive and six (27.3%) had been previously treated for TB. All these 22 suspected patients started treatment because they fulfilled the criteria for clinically diagnosed TB.

PCF cohort

Of those samples positive by Ultra, 11.9% (24 out of 201) had a trace result. Reclassifying all traces as negative decreased the sensitivity of Ultra, although it still had superior point sensitivity than Xpert (0.91 (95% CI 0.85–0.95) *versus* 0.88 (95% CI 0.82–0.93); p=0.02); Ultra's specificity was improved (0.97) substantially close to Xpert's estimates (0.98). Six of those 24 samples (25%) with "trace call" belonged to patients who had been previously treated; thus, under the second recategorisation approach (considering

those results as negative), Ultra maintained superior sensitivity. Conversely, specificity remained lower than that of Xpert.

ACF cohort

41.7% (five out of 12) of Ultra-positive tests had trace results. None of them was culture-positive, thus recategorisations maintained Ultra's sensitivity. The conditional recategorisation resulted in the reclassification of two trace results from previously treated patients.

Combined cohort

The analysis was also repeated for both cohorts combined. Overall, test values did not differ from those achieved in the PCF group. Detailed information can be found in the supplementary material. Figure 3 summarises test values in comparison with published data.

Drug resistance detection

Ultra detected rifampicin resistance in 8% of cases (17 out of 213) and Xpert in 8.7% of cases (15 out of 173). There were two cases in which Ultra tagged resistance markers that were not detected by Xpert. One of them also showed phenotypic resistance by culture; however, the other was not identified by the conventional drug susceptibility tests. Conversely, four drug susceptible results detected by Xpert were identified as trace by Ultra, missing the information about drug susceptibility (supplementary material).

Correlation of measures of bacillary burden

Ultra IS1081/IS6110 C_t and the minimum *rpoB* C_t for Xpert directly correlated to smear grade, and variability in C_t values was higher for Xpert than Ultra (supplementary material). Additionally, we plotted C_t values to explore associations between both molecular assays. Patterns of correlation are shown in a smoothed curve (figure 4a). The Spearman rank correlation coefficient revealed a good monotonic positive correlation among both molecular tests (r=0.79, p<0.001) excluding negative results. From the curve, two different trends of association can be observed. Lastly, exploring data stratified by whether Ultra was positive or not, we can observe that most cases with negative Xpert and positive Ultra results corresponded to a low bacillary burden specimen (high C_t values) (figure 4b).

Discussion

This is the first field diagnostic evaluation of Ultra and Xpert under different programmatic approaches, and one of the few using the same single specimen under field conditions [28]. Test parameters varied by

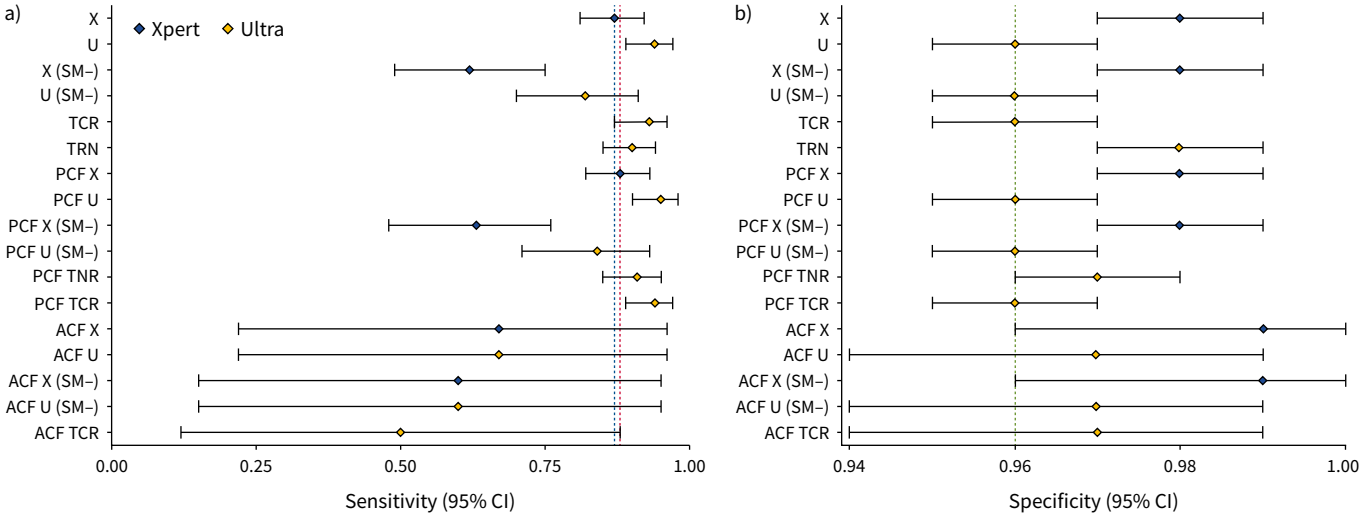

FIGURE 3 Sensitivity and specificity values of the different diagnostic methods, overall and for each cohort, and comparison with previously published data. a) Comparison of sensitivity values of Xpert and Ultra. Dotted lines represent overall sensitivity estimated for the two systematic reviews published so far: blue line, data from Zhang *et al.* [16] (sensitivity 0.87); red line, data from the World Health Organization [17, 18] (sensitivity 0.88). b) Comparison of specificity values of Xpert and Ultra. The dotted green line represents overall specificity estimated for the previously cited articles (specificity 0.96). X: Xpert; U: Ultra; SM: smear microscopy; TCR: trace conditional recategorisation; TRN: trace recategorisation as negative; PCF: passive case finding; ACF: active case finding.

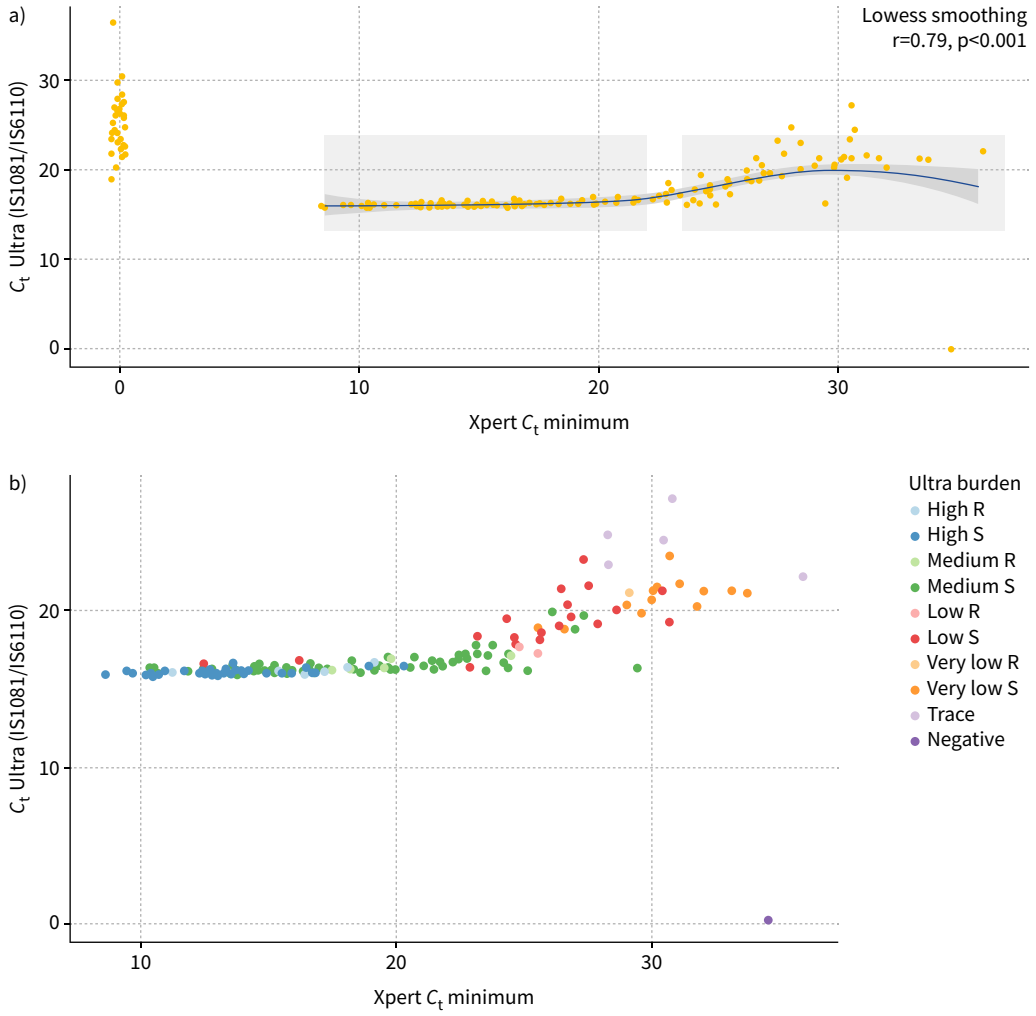

FIGURE 4 Correlation of cycle threshold (C_t) values. The *x*-axis represents the minimum (non-zero) *rpoB* C_t for Xpert. The *y*-axis represents the Ultra IS1081/IS6110 C_t value. a) Overall, two trends of correlation can be observed (shaded boxes): a first fitted curve with a slope close to zero and a second ascendent fitted curve. b) Correlation of C_t values, dissagregated by categories of Ultra results. R: rifampicin resistance detected; S: rifampicin resistance not detected.

strategy used to diagnose patients. Under PCF conditions, Ultra achieved higher sensitivity and lower specificity compared with Xpert. 12% of participants with any positive result were detected only by Ultra. In the ACF cohort, sensitivity of Xpert and Ultra was similar and lower (0.67) than for the PCF cohort. Specificity parameters remained high for both tests. A minority of trace results belonged to participants reporting a previous history of TB treatment.

In our PCF cohort, Xpert and Ultra tests reached slightly improved values for sensitivity than those estimated in the two recently published systematic reviews (figure 3) [16, 19]. This improvement in sensitivity was obtained at the expense of slightly diminished specificity. Conversely, for the ACF group, exploration of test values was limited by the number of positive cases. Only 2.4% (six out of 260) of those who provided a sample were positive by the composite reference standard. However, Ultra doubled the yield of positive results (4.8% *versus* 2.4% of cultured-confirmed and 2.8% of Xpert-confirmed). This higher positivity rate of Ultra could be due to the paucibacillary nature of TB disease in cases found in the community, where only a low proportion of participants are symptomatic (38% (99 out of 260)).

In diagnostic accuracy studies, it is generally assumed that gold standards are hypothetical error-free tests [29]. However, culture for *M. tuberculosis* also shows limitations to identify paucibacillary specimens and

nonviable DNA. Among others, narrow margins of decontamination challenge whether a positive molecular test (culture-negative) is a real false-positive result. Culture contamination rates were validated as per the international laboratory quality standards and protocols were adapted fitting established contamination ranges (5–10% for liquid culture) [30, 31]. Indeed, an excess of decontamination might have played a role in Ultra's sensitivity findings, given the high yield of molecular tests compared with culture in both cohorts. We speculate that the use of the modified Kubica method for decontamination purposes and the high proportion of nontuberculous mycobacteria isolates (aligned with some previous studies [32]) could affect the growth of specimens with low bacillary burden. Thus, misclassification of apparent false-negative samples would overestimate the true sensitivity and underestimate the specificity of the assay under evaluation.

Previous studies also showed lower specificity of Ultra compared with Xpert [16, 17, 19] and some associate this to false-positive results among previously treated patients [15, 27]. Since the compromised specificity of Ultra is largely based on trace results, their interpretation and how to translate the test into daily clinical practice remains controversial [15, 17, 18]. Their reclassification as negative maintained the superiority of Ultra's sensitivity and equated its specificity compared with Xpert. When trace recategorisation was conditioned by previous TB treatment, specificity differences between both assays remained meaningful. Notably, by extending incubation and using combined medium we achieved 20.7% of grown cultures. In addition, most patients with trace results were put on treatment and mortality within this group was very high (21%). These findings may shed light on the fact that the impaired specificity of Ultra might not only depend on the detection of genetic material stemming from previous disease [33] but perhaps on stochastic variation (sample used, sample processing variability). Thus, clinical implications of the interpretation of Ultra's trace calls should consider other information (such as presence of TB-compatible symptoms, epidemiological links or compatible radiography findings) [34]. In addition, we explored C_t values as a proxy of bacillary burden. The correlation between Ultra and Xpert C_t is in agreement with the initial technical evaluation of Ultra [35]. As the new target of the Ultra PCR is a multicopy region of the bacteria, Ultra improved the identification of samples with low DNA concentration and thus a considerable number of additional cases were detected.

Ultra was endorsed by the WHO for TB diagnosis, being widely adopted in most national diagnostic guidelines (PCF-based) [5]. However, recommendations on its use in screening activities (such as ACF) remain unclear due to its apparent lower specificity (attributed, to a great extent, to the interpretation of trace results as positive cases). False-positive results have implications for patient clinical management and related costs [36, 37]. In the ACF cohort, 40% (six out of 15) of participants with a positive test were identified just by Ultra; four out of those six had a trace result and only one participant had been previously treated. Operationally, the prevalence of TB in contact tracing studies is generally low [38], thus a decrease in Ultra's specificity implies a reduction in the positive predictive value. However, apparent false-positive results could indeed be true-positive results, thus Ultra's trace result in the context of ACF activities might allow the identification of early stages of the disease, which might be paucibacillary and/or subclinical, and in consequence foster more effective community-based activities [7].

Our study has limitations. First, the risk of active TB after exposure is highest during the first year [6], but the number of cases confirmed by culture through ACF was low. This could be explained by the early screening of household contacts and the small sample size, which affected the precision of Ultra performance estimates. Second, the low rate of resistance softens our ability to make strong conclusions with regard to accuracy for rifampicin resistance. Third, as mentioned, some negative culture but positive molecular results (Xpert and/or Ultra) may be due to a higher concentration of sodium hydroxide used for decontamination. Lastly, in the PCF cohort, many participants were missing HIV status as this is often not completed under routine conditions.

Conclusions

Ultra showed lower specificity but superior sensitivity than Xpert in the evaluation of samples in the PCF cohort. Test performance partly depended on the interpretation of trace results. Under the ACF strategy, sensitivity and specificity for both tests were similar, although Ultra detected a higher number of microbiologically confirmed samples. Our results add evidence for the use and interpretation of the Ultra assay as the front-line tool for TB diagnosis in different case finding scenarios

Acknowledgements: We thank all participants and their families for participating in the research project. We are also thankful for all the support received by the Mozambican's National TB Control Programme and the Stop TB Partnership TB REACH initiative.

Data availability: The datasets generated during the current study are kept at CISM's data centre. Individual de-identified participant data and data dictionaries can be shared with interested investigators upon a formal request to CISM's Internal Scientific Committee (cci@manhica.net) and the corresponding author. Data will be available immediately following publication with no end date.

Conflict of interest: None declared.

Support statement: This work was supported by the Stop TB Partnership (W5_MOZ_CDT1_CISM-FM). We acknowledge support from the Spanish Ministry of Science, Innovation and Universities through the "Centro de Excelencia Severo Ochoa 2019–2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program. B. Saavedra receives a pre-doctoral fellowship from the Secretariat of Universities and Research, Ministry of Enterprise and Knowledge of the Government of Catalonia and co-funded by European Social Fund (AGAUR). This project has received donations of consumables from Cepheid. Cepheid was not involved in the data analysis or decision to submit this manuscript for publication. G. Theron is supported by the EDCTP2 program supported by the European Union (SF1401, OPTIMAL DIAGNOSIS) and The Center for Innovation in Point-of-Care Technologies for HIV/AIDS at Northwestern University (National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health) (U54EB027049). Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 World Health Organization. Global Tuberculosis Report 2020. Geneva, WHO, 2020.
2 Albert H, Nathavitharana RR, Isaacs C, *et al.* Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J* 2016; 48: 516–525.
3 World Health Organization. Systematic Screening for Active Tuberculosis. Principles and Recommendations. Geneva, WHO, 2013.
4 World Health Organization. WHO Guidelines on Tuberculosis Infection Prevention and Control, 2019 Update. Geneva, WHO, 2019.
5 World Health Organization. WHO Consolidated Guidelines on Tuberculosis. Module 3: Diagnosis. Geneva, WHO, 2020.
6 Fox GJ, Barry SE, Britton WJ, *et al.* Contact investigation for tuberculosis: a systematic review and meta-analysis. *Eur Respir J* 2013; 41: 140–156.
7 Lönnroth K, Corbett E, Golub J, *et al.* Systematic screening for active tuberculosis: rationale, definitions and key considerations. *Int J Tuberc Lung Dis* 2013; 17: 289–298.
8 Kempker RR, Chkhartishvili N, Kinkladze I, *et al.* High yield of active tuberculosis case finding among HIV-infected patients using Xpert MTB/RIF testing. *Open Forum Infect Dis* 2019; 6: ofz233.
9 Drain PK, Hyle EP, Noubary F, *et al.* Diagnostic point-of-care tests in resource-limited settings. *Lancet Infect Dis* 2014; 14: 239–249.
10 World Health Organization. Unitaid Tuberculosis Diagnostics Technology Landscape. 5th Edn. Geneva, WHO, 2017.
11 World Health Organization. Expert Group Meeting Report: Using the Xpert MTB/RIF Assay to Detect Pulmonary Tuberculosis and Rifampicin Resistance in Adults and Children. Geneva, WHO, 2013.
12 Piersimoni C, Gherardi G, Gracciotti N, *et al.* Comparative evaluation of Xpert MTB/RIF and the new Xpert MTB/RIF Ultra with respiratory and extra-pulmonary specimens for tuberculosis case detection in a low incidence setting. *J Clin Tuberc Other Mycobact Dis* 2019; 15: 100094.
13 Donovan J, Thu DDA, Phu NH, *et al.* Xpert MTB/RIF Ultra *versus* Xpert MTB/RIF for the diagnosis of tuberculous meningitis: a prospective, randomised, diagnostic accuracy study. *Lancet Infect Dis* 2020; 20: 299–307.
14 Beutler M, Plesnik S, Mihalic M, *et al.* A pre-clinical validation plan to evaluate analytical sensitivities of molecular diagnostics such as BD MAX MDR-TB, Xpert MTB/Rif Ultra and FluoroType MTB. *PLoS One* 2020; 15: e0227215.
15 Mishra H, Reeve BWP, Palmer Z, *et al.* Xpert MTB/RIF Ultra and Xpert MTB/RIF for diagnosis of tuberculosis in an HIV-endemic setting with a high burden of previous tuberculosis: a two-cohort diagnostic accuracy study. *Lancet Respir Med* 2020; 8: 368–382.
16 Zhang M, Xue M, He J-Q. Diagnostic accuracy of the new Xpert MTB/RIF Ultra for tuberculosis disease: a preliminary systematic review and meta-analysis. *Int J Infect Dis* 2020; 90: 35–45.
17 FIND. A Multicentre Non-Inferiority Diagnostic Accuracy Study of the Ultra Assay Compared to the Xpert MTB/RIF Assay. Geneva, FIND, 2017.
18 World Health Organization. WHO Meeting Report of a Technical Expert Consultation: Non-inferiority Analysis of Xpert MTB/RIF Ultra Compared to Xpert MTB/RIF. Geneva, WHO, 2017.
19 Horne DJ, Kohli M, Zifodya JS, *et al.* Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2019; 6: CD009593.

20 Garcia-Basteiro AL, Hurtado JC, Castillo P, *et al.* Unmasking the hidden tuberculosis mortality burden in a large *post mortem* study in Maputo Central Hospital, Mozambique. *Eur Respir J* 2019; 54: 1900312.
21 García-Basteiro AL, López-Varela E, Respeito D, *et al.* High tuberculosis burden among people living with HIV in southern Mozambique. *Eur Respir J* 2015; 45: 547–549.
22 Nhacolo A, Jamisse E, Augusto O, *et al.* Cohort profile update: Manhiça Health and Demographic Surveillance System (HDSS) of the Manhiça Health Research Centre (CISM). *Int J Epidemiol* 2021; 50: 395.
23 Stop TB Partnership. Framework for the Evaluation of New Tests for Tuberculosis Infection. Geneva, Stop TB Partnership, 2020.
24 Sacoor C, Nhacolo A, Nhalungo D, *et al.* Profile: Manhica Health Research Centre (Manhica HDSS). *Int J Epidemiol* 2013; 42: 1309–1318.
25 Lumb R, Van Deun A, Bastian I, *et al.* Laboratory Diagnosis of Tuberculosis by Sputum Microscopy. The Handbook. Adelaide, SA Pathology, 2013.
26 Cercenado E, Cantón R, Alcaide Fernández De Vega F, *et al.* Procedimientos en Microbiología Clínica: Recomendaciones de la Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica. [Procedures in Clinical Microbiology: Recommendations of the Spanish Society of Infectious Diseases and Clinical Microbiology.] Madrid, EIMC, 2005.
27 Dorman SE, Schumacher SG, Alland D, *et al.* Xpert MTB/RIF Ultra for detection of *Mycobacterium tuberculosis* and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *Lancet Infect Dis* 2018; 18: 76–84.
28 Pereira GR, Barbosa MS, Dias NJD, *et al.* Evaluation of Xpert MTB/RIF Ultra performance for pulmonary tuberculosis (TB) diagnosis in a city with high TB incidence in Brazil. *Respir Med* 2020; 162: 105876.
29 Tang S, Hemyari P, Canchola JA, *et al.* Dual composite reference standards (dCRS) in molecular diagnostic research: a new approach to reduce bias in the presence of Imperfect reference. *J Biopharmac Stat* 2018; 28: 951–965.
30 European Center for Diseases Prevention and Control. Handbook on Tuberculosis Laboratory Diagnostic Methods in the European Union. Stockholm, ECDC, 2018.
31 Global Laboratory Initiative. Mycobacteriology Laboratory Manual. Geneva, GLI, 2014.
32 López-Varela E, García-Basteiro AL, Augusto OJ, *et al.* High rates of non-tuberculous mycobacteria isolation in Mozambican children with presumptive tuberculosis. *PLoS One* 2017; 12: e0175613.
33 Theron G, Venter R, Calligaro G, *et al.* Xpert MTB/RIF results in patients with previous tuberculosis: can we distinguish true from false positive results? *Clin Infect Dis* 2016; 62: 995–1001.
34 Global Laboratory Initiative. Planning for Country Transition to Xpert MTB/RIF Ultra Cartridges. Geneva, GLI, 2017.
35 Chakravorty S, Simmons AM, Rowneki M, *et al.* The new Xpert MTB/RIF Ultra: improving detection of *Mycobacterium tuberculosis* and resistance to rifampin in an assay suitable for point-of-care testing. *mBio* 2017; 8: e00812-17.
36 World Health Organization. Systematic Screening for Active Tuberculosis: An Operational Guide. Geneva, WHO, 2015.
37 Chadha VK, Praseeja P. Active tuberculosis case finding in India – the way forward. *Indian J Tuberc* 2019; 66: 170–177.
38 Cudahy PGT, Andrews JR, Bilinski A, *et al.* Spatially targeted screening to reduce tuberculosis transmission in high-incidence settings. *Lancet Infect Dis* 2019; 19: e89–e95.