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

Hridesh Mishra, Byron W P Reeve*, Zaida Palmer, Judy Caldwell, Tania Dolby, Charissa C Naidoo, Jennifer G Jackson, Samuel G Schumacher, Claudia M Denkinger, Andreas H Diacon, Paul D van Helden, Florian M Marx, Robin M Warren, Grant Theron*

Summary

Background Xpert MTB/RIF Ultra (Ultra) is a new test for tuberculosis undergoing global roll-out. We assessed the performance of Ultra compared with Xpert MTB/RIF (Xpert) in an HIV-endemic setting where previous tuberculosis is frequent and current test performance is suboptimal.

Methods In this two-cohort diagnostic accuracy study, we used sputum samples from patients in South Africa to evaluate the accuracy of Ultra and Xpert against a single culture reference standard. For the first cohort (cohort A), we recruited adults (aged ≥18 years) with symptoms of presumptive tuberculosis at Scottsdene clinic in Cape Town, South Africa. We collected three sputum samples from each patient in cohort A, two at the first visit of which one was tested using Xpert and the other was tested using culture, and one sample the next morning which was tested using Ultra. In a separate cohort of patients with presumptive tuberculosis and recent previous tuberculosis (≤2 years) who had submitted sputum samples to the National Health Laboratory Services (cohort B), decontaminated sediments were, after processing, randomly allocated (1:1) for testing with Ultra or Xpert. For both cohorts we calculated the sensitivity and specificity of Ultra and Xpert and evaluated the effects of different methods of interpreting Ultra trace results.

Findings Between Feb 6, 2016, and Feb 2, 2018, we recruited 302 people into cohort A, all of whom provided sputum samples and 239 were included in the head-to-head analyses of Ultra and Xpert. For cohort B, we collected sputum samples from eligible patients who had submitted samples between Dec 6, 2016, and Dec 21, 2017, to give a cohort of 831 samples, of which 352 were eligible for inclusion in analyses and randomly assigned to Ultra (n=173) or Xpert (n=179). In cohort A, Ultra gave more non-actionable results (not positive or negative) than did Xpert (28 [10%] 275 *vs* 14 [5%] 301; p=0·011). In the head-to-head analysis, in smear-negative patients, sensitivity of Ultra was 80% (95% CI 64–90) and of Xpert was 73% (57–85; p=0·45). Overall, specificity of Ultra was lower than that of Xpert (90% [84–94] *vs* 99% [95–100]; p=0·001). In cohort B, overall sensitivity was 92% (81–98) for Xpert versus 86% (73–95; p=0·36) for Ultra and overall specificity was 69% (60–77) for Ultra versus 84% (78–91; p=0·005) for Xpert. Ultra specificity estimates improved after reclassification of results with the lowest Ultra-positive semiquantitation category (trace) to negative (15% [8–22]). In cohort A, the positive predictive value (PPV) for Ultra was 78% (67–87) and for Xpert was 96% (87–99; p=0·004); in cohort B, the PPV for Ultra was 50% (43–57) and for Xpert was 70% (61–78; p=0·014). Ultra PPV estimates in previously treated patients were low: at 15% tuberculosis prevalence, half of Ultra-positive patients with presumptive tuberculosis would be culture negative, increasing to approximately 70% in patients with recent previous tuberculosis. In cohort B, 21 (28%) of 76 samples that were Ultra positive were rifampicin indeterminate (all trace) and, like cohort A, most were culture negative (19 [90%] of 21).

Interpretation In a setting with a high burden of previous tuberculosis, Ultra generated more non-actionable results and had diminished specificity compared with Xpert. In patients with recent previous tuberculosis, a quarter of Ultra-positive samples were indeterminate for rifampicin resistance and culture negative, suggesting that additional drug-resistance testing will probably be unsuccessful. Our data have implications for the handling of Ultra-positive results in patients with previous tuberculosis in high burden settings.

Funding South African Medical Research Council, the EDCTP2 program, and the Faculty of Medicine and Health Sciences, Stellenbosch University.

Lancet Respir Med 2020

Published Online
February 14, 2020
https://doi.org/10.1016/S2213-2600(19)30370-4

See Online/Comment
https://doi.org/10.1016/S2213-2600(20)30050-3

*Joint first authors

NRF-DST Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics (H Mishra PhD, B W P Reeve PhD, Z Palmer MSc, C C Naidoo PhD, J G Jackson PhD, Prof P D van Helden PhD, Prof R M Warren PhD, G Theron PhD), **and Desmond Tutu TB Centre, Department of Paediatrics and Child Health** (F M Marx PhD), **Faculty of Medicine and Health Sciences, Division of Medical Physiology** (Prof A H Diacon MD), **and DST-NRF South African Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA)** (F M Marx), **Stellenbosch University, Cape Town, South Africa; City of Cape Town, Cape Town, South Africa** (J Caldwell MPH); **National Health Laboratory Services, Cape Town, South Africa** (T Dolby DMM); **FIND, Geneva, Switzerland** (S G Schumacher PhD, C M Denkinger MD); **and University of Heidelberg, Division of Tropical Medicine, Center of Infectious Diseases, Heidelberg, Germany** (C M Denkinger)

Correspondence to:
Dr Grant Theron, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town 7550, South Africa
gtheron@sun.ac.za

Introduction

Xpert MTB/RIF (Xpert) has been scaled-up for the diagnosis of tuberculosis and rifampicin resistance; however, Xpert performs suboptimally, especially in smear-negative sputum samples, which are frequently obtained from patients who are HIV positive.[1–3]

Research in context

Evidence before this study

Xpert MTB/RIF Ultra (Ultra) is a new test for tuberculosis and rifampicin susceptibility endorsed by WHO. In January, 2016, we searched PubMed for studies published in 2015 or earlier using the search term "Xpert MTB/RIF Ultra", with no language restrictions. A multicentre study described increased sensitivity for smear-negative pulmonary tuberculosis. However, overall, little data are available from HIV-endemic settings and especially from patients who have been previously treated for tuberculosis. Such patients, especially those with recent previous tuberculosis, often re-present with symptoms but have old *Mycobacterium tuberculosis* complex DNA in their sputum, which diminishes specificity and increases the proportion of results that are false positive. Ultra positive results with the lowest test semiquantitation value (trace), which are frequent in patients with paucibacillary sputum and do not have useful results on rifampicin resistance, also require clarification.

Added value of this study

In this large study, among patients with presumptive tuberculosis in a high HIV and previous tuberculosis prevalence setting in Cape Town, South Africa, two cohorts were studied. In cohort A, sputum samples from all patients were tested using Ultra, its predecessor Xpert MTB/RIF (Xpert), and culture. 25% of patients who were Ultra positive were culture negative, meaning they were erroneously classified as positive for tuberculosis by Ultra against the conventionally used reference standard. This diminished specificity was associated with previous tuberculosis. In cohort B, in which samples were used from patients who had recent previous tuberculosis treatment (within the past 2 years), half of all samples that were Ultra positive were culture negative; indicating that a positive Ultra result has little tuberculosis diagnostic use in this epidemiologically important population. When estimated across a range of different tuberculosis prevalences, the positive predictive value of Ultra remained suboptimal. Importantly, trace reclassification recovered some specificity and positive predictive value. Cohort B had a high proportion of patients who were trace positive, most of whom were culture negative, indicating that culture is probably not a useful follow-on test for drug susceptibility despite the higher risk of drug resistance in patients who have been treated previously.

Implications of all the available evidence

All available evidence shows that a positive Ultra result (even after trace exclusion) has little diagnostic use in patients who have been previously treated for tuberculosis. Furthermore, the diagnosis of rifampicin resistance in these patients, who have an increased baseline risk of resistance, will prove difficult in those who are trace positive. These findings, combined with high rates of non-actionable results (ie, not positive or negative), add complexity to clinical and programmatic decision making and require careful consideration against the possible gains offered by Ultra compared with Xpert.

This diminished sensitivity prompted the development of the next generation Xpert MTB/RIF Ultra (Ultra) test, which was endorsed by WHO in 2018.[4]

Ultra can detect lower concentrations of *Mycobacterium tuberculosis* complex than Xpert (16 *vs* 113 colony forming units per mL).[5] In a multicentre study that used sputum samples from patients with presumptive tuberculosis,[6] sensitivity of Ultra was superior to that of Xpert in patients who were smear negative for *M tuberculosis* (63% for Ultra *vs* 46% for Xpert; difference of 17% [95% CI 10–24]). However, sensitivity data are lacking from HIV-endemic settings.[3,7,8]

In the same multicentre study, results that were Ultra positive and culture negative (ie, false positives) were more frequent in patients who had previously had treatment for tuberculosis than those who had not (88% *vs* 83%; change in specificity of –5% [95% CI –9 to –3]). The recency of previous treatment for tuberculosis is a known risk factor for diminished specificity of Xpert[9,10] and recency of previous treatment is also a risk factor for recurrence, with the risk being highest within 2 years of treatment.[11,12] Thus, symptomatic patients with recent previous treatment for tuberculosis are a diagnostic challenge. Despite being a potentially important target for tuberculosis control interventions,[13] the overall number of patients who have been previously treated in whom Ultra has been evaluated (n=249 with any history, n=55 treated within the past 2 years)[6] is inadequate to inform algorithms.

One of several improvements offered by Ultra over Xpert is the addition of a new lowest semiquantitation category (trace) that reflects the positive detection of *M tuberculosis* complex DNA.[5] Trace is potentially a major source of incremental sensitivity (trace-positive samples have amplification of the multicopy IS*6110* and IS*1081* regions but not the single copy *rpoB* region).[5] Different strategies for interpreting results that are Ultra positive and categorised as trace have been explored to improve specificity.[14] However, the optimal strategy, including stratification by tuberculosis treatment history, remains unclear.[3,15] If, as considered by some tuberculosis programmes, traces are not reported as positive, the benefits offered by Ultra over Xpert will be reduced (Xpert relies only on *rpoB* amplification). Furthermore, Ultra does not produce an actionable (positive or negative) result on rifampicin resistance in patients who are trace positive because no *rpoB* amplification occurs. Therefore, programmes that use Ultra might be confronted with increased numbers of patients with unknown rifampicin resistance profiles, which might be especially problematic in patients who have been previously treated who have, on average, a five-times increased risk of drug-resistant tuberculosis.[16]

Therefore, to inform Ultra's uptake, we did a head-to-head evaluation of the diagnostic accuracy of Ultra and Xpert in a setting with high prevalence of HIV (approximately 20%) and previous tuberculosis treatment (approximately 26% of tuberculosis cases in Cape Town, South Africa, were treated recently within the past 2 years).[17] To further assess accuracy in patients who are the most at risk of false positive results, we leveraged an opportunity to test routinely collected sputum sediment remnants from symptomatic patients with recent previous tuberculosis. We estimated how different semiquantitation recategorisation strategies changed sensitivity and specificity, the effect of a simulated four-culture reference standard in a scenario maximally generous to Ultra, and Ultra and Xpert predictive values across different tuberculosis prevalences and populations with recent previous tuberculosis treatment.

Methods

Study design and population

In this diagnostic study, we used sputum samples from two cohorts of patients with presumptive tuberculosis to determine the diagnostic accuracy, sensitivity, and specificity of Ultra and Xpert compared with culture. For the first cohort (cohort A), adults (aged ≥18 years) with presumptive tuberculosis symptoms[18] were consecutively recruited at Scottsdene clinic in Cape Town, South Africa. A patient with presumptive tuberculosis was defined as presenting with symptoms or signs suggestive of tuberculosis as per standard criteria.[19] Demographic, clinical (including tuberculosis symptom score [TBscore II][20]), and microbiological data were captured on REDCap.[21] For the second cohort (cohort B), we collected sputum samples that had been submitted to the National Health Laboratory Services (NHLS) under programmatic conditions from health facilities across Cape Town by individuals with presumptive tuberculosis who self-reported having recently had previous tuberculosis treatment within 2 years (exact period since treatment not recorded). For both cohorts, we excluded patients who, at the time of submission of samples, had treatment for tuberculosis within 2 months, an unknown treatment status, or were missing a paired positive or negative culture result.

This study was approved by the Stellenbosch University Faculty of Health Sciences Research Ethics Committee (cohort A, N14/10/136; cohort B, N09/11/296) and the City of Cape Town (cohort A, 10483; cohort B, 10570). Informed consent was obtained from patients in cohort A. Because de-identified, routinely collected sample remnants were used for cohort B, informed consent was waived.

Procedures

Patients in cohort A provided two sputum samples at the first visit. Of the two samples, the more viscous[22] sample was decontaminated with Mycoprep (BD, Johannesburg, South Africa) and used for double Ziehl-Neelsen smear microscopy and a Mycobacteria Growth Indicator Tube (MGIT) 960 liquid culture. The other sputum sample was used for testing with Xpert. The next day, patients provided a third sputum sample in the early morning, which was used for testing with Ultra. Sputum samples were typically induced with a nebuliser (Ultrasonic Hospital Grade WH-802, Hitech Therapy, Johannesburg, South Africa) with 5% sodium chloride solution (Ysterplaat Medical Supplies, Cape Town, South Africa) for 7–10 min.[1] If a test result was non-actionable (the results were not useful for clinical decision making—ie, not positive or negative), the specimen sample reagent mix was re-tested and if an insufficient volume remained for such re-testing, the initial result was used.

In cohort B, each patient provided two sputum samples, both of which were decontaminated with N-acetyl-L-cysteine sodium hydroxide solution and used for Ziehl-Neelsen smear microscopy (50 µL each). One decontaminated sediment then had bleach added and the other was resuspended in phosphate-buffered saline (PBS; Sigma Aldrich, Modderfontein, South Africa). Approximately 500 µL of the remaining non-bleach-treated resuspended sediment was used for a MGIT960 culture. The remaining resuspended sediment was collected, concentrated into pellets and resuspended in PBS, and randomly assigned (1:1) to testing with Xpert or Ultra (more details are in the appendix [p 3]). We analysed samples to check eligibility and removed any duplicates.

See Online for appendix

In both cohorts, the MGIT960 culture was used as a reference standard and MTBDR*plus* (Hain LifeSciences, Nehren, Germany) was done on culture-positive isolates for the detection of *M tuberculosis* complex and rifampicin and isoniazid resistance.

Patients in cohort A who were identified as positive for tuberculosis by Ultra or Xpert testing, or both, and were found to be culture negative, were followed up for typically at least 1 year to ascertain if they started treatment programmatically in the interim. If not on treatment after 1 year, sputum induction was done and patients were re-tested using Ultra, Xpert, and culture.

Predictive value estimates across different settings

We estimated predictive values of the Xpert and Ultra tests using 2×2 tables and Ultra and Xpert sensitivity and specificity estimates from cohorts A (stratified by treatment history) and B at different pre-test probabilities (ie, prevalence of single MGIT960 culture positivity). We also did analyses using sensitivity and specificity estimates from a four-culture reference standard scenario (comprising two MGIT960 and two Löwenstein–Jensen solid cultures), trace recategorisation scenarios, and sensitivity estimates from a multicentre evaluation of Ultra and Xpert.[6] Furthermore, we estimated predictive values in patients who had previous tuberculosis treatment based on the proportion of patients with presumptive tuberculosis with recent previous tuberculosis (appendix p 4).

Ultra trace results can be recategorised using the following strategies. Trace reclassification involves the recategorisation of trace results that were Ultra positive to Ultra negative, whereas trace exclusion excludes trace results from 2×2 tables. The minimum *rpoB* cycle threshold value for bacillary load (C_{Tmin}) is the lowest non-zero C_T value for an Ultra-positive result that is not trace.[23]

We did subanalyses by previous tuberculosis and HIV status.

Outcomes

In cohort A, we compared the sensitivity, specificity, and predictive values of Ultra and Xpert with those of culture, overall and stratified by previous tuberculosis status. In cohort B, we compared the sensitivity, specificity, and predictive values of Ultra and Xpert with those of culture, in recent previous tuberculosis only. In both cohorts, we investigated the effect of trace recategorisation on diagnostic accuracy and estimated predictive values of Ultra and Xpert in different settings.

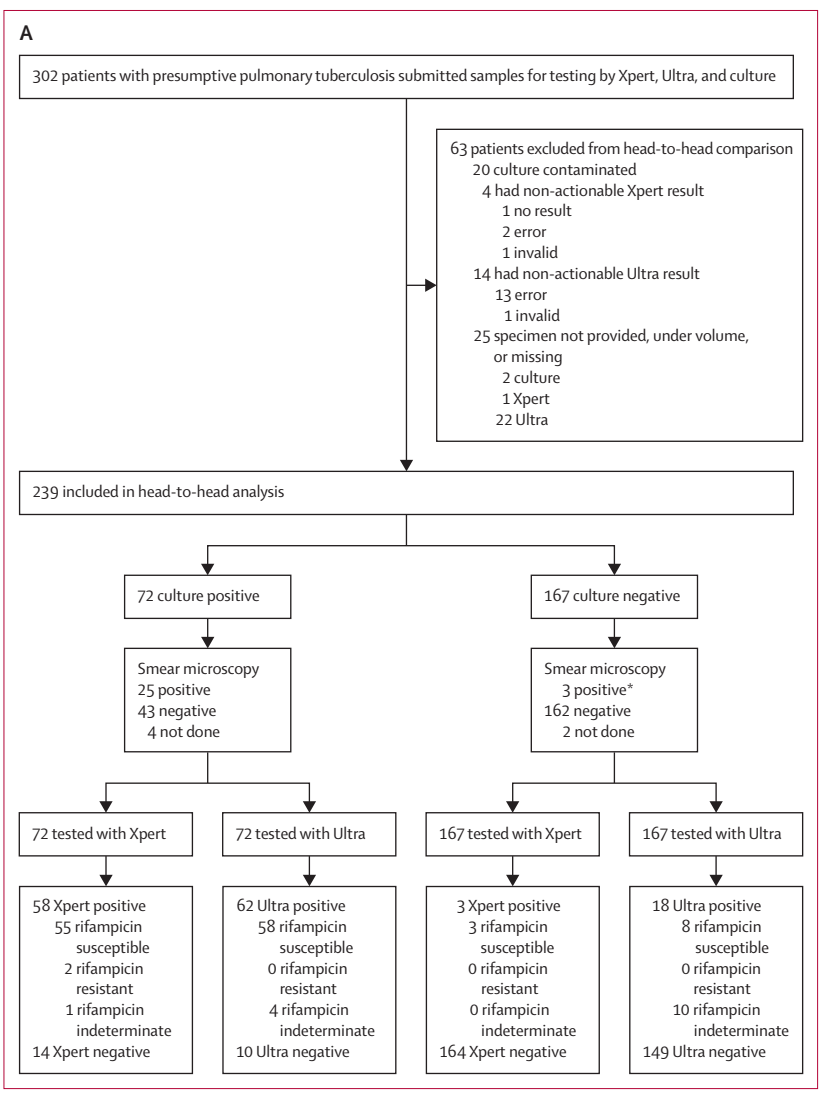


Figure 1 continues on next page

Statistical analysis

We included patients in the head-to-head analysis if they had an actionable Xpert, Ultra, and culture result. We included patients in the non-head-to-head analysis if they had a non-actionable Xpert but actionable Ultra and culture or non-actionable Ultra but actionable Xpert and culture.

We followed the STARD guidelines for this study design and analysis (appendix pp 22–23).[24] We analysed data using the χ^2 test (including McNemar's test) to calculate differences in diagnostic accuracy metrics; the Mann-Whitney *U* test to calculate differences in non-parametric data; the two-sample proportion test to compare, for example, specificity across two groups; and the Kruskal-Wallis and Spearman's ρ tests for bacterial load (time to positivity or smear grade or Ultra C_T). In cohort A, we did not record respiratory rate and we calculated TBscore II out of a total of seven points. In both cohorts, we calculated the potential effect of a four-culture reference standard on specificity in a scenario maximally generous to Ultra in both cohorts (full details are in the appendix [p 4]).

We projected positive predictive values (PPVs) using the common formula PPV = (sensitivity × prevalence) / (sensitivity × prevalence + [1–specificity] × [1–prevalence]). PPVs were first calculated separately for presumptive patients with recent previous treatment and those with non-recent previous treatment using different estimates of test specificity obtained via the study because test specificity is expected to be significantly lower among individuals recently treated for tuberculosis to due to residual *M tuberculosis* complex DNA from the previous tuberculosis episode. PPV estimates were then combined for all previously treated patients, weighted for the percentage of patients with recent tuberculosis treatment. We obtained PPV projections for varying (assumed) levels of tuberculosis prevalence (pre-test probability; range 0–30%) and of the proportion of individuals with recent previous treatment (0–2 years ago; range 0–35%). Projections in the three-dimensional figure refer to 1% increases in each pre-test probability and percentage recent tuberculosis treatment, as shown through the small gridlines in the graphs. The smooth surfaces in the graphs result from the fact that higher PPVs are strictly positively correlated with higher tuberculosis prevalence (pre-test probability) and lower proportions of individuals are correlated with recent tuberculosis treatment (and hence higher test specificity; see formula).

We used Stata version 15 and GraphPad Prism version 7.0 for all analyses.

Role of the funding sources

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or

writing of the report. The corresponding author had full access to all data and had final responsibility for the decision to submit for publication.

Results

Between Feb 6, 2016, and Feb 2, 2018, we recruited 302 people with presumptive tuberculosis into cohort A, all of whom were enrolled and provided sputum samples. the cohort comprised 144 (48%) women and 158 (52%) men, with an overall median age of 36 years (IQR 26–49). In cohort A, we included samples from 239 (79%) of 302 patients in the head-to-head analyses (figure 1). 72 (30%) of 239 were culture positive for *M tuberculosis*, of whom 43 (63%) of 68 who had available data were smear negative for *M tuberculosis* (table 1). The proportion of patients who were culture negative for *M tuberculosis* who had previously had tuberculosis treatment was high (69 [41%] of 167; and 15 [22%] of 69 who had recent previous tuberculosis treatment). A higher proportion of patients who were culture positive than who were culture negative were of black ethnicity (14 [19%] of 72 *vs* 16 [10%] of 167; p=0·035) and HIV positive (21 [29%] of 72 *vs* 27 [16%] of 166; p=0·023). Patients who were culture positive had greater morbidity than those who were culture negative (median TBscore II score of 3 [IQR 2–4] *vs* 2 [2–3]; p<0·0001) and lower haemoglobin concentrations (13 g/dL [11–14] *vs* 15 g/dL [14–16]; p<0·0001).

For cohort B, we collected sputum samples from eligible patients who had submitted samples between Dec 6, 2016, and Dec 21, 2017, to give us a cohort of 831 samples. 352 patient samples were eligible (characteristics overall and by allocated testing method are in the appendix [pp 10–11]). 95 (27%) of 346 samples were culture positive for *M tuberculosis*, of these, 36 (43%) of 84 with available data were smear negative. More patients who gave samples for cohort B than patients in cohort A were male (212 [60%] of 351 *vs* 123 [51%] of 239; p=0·032) and HIV positive (124 [44%] of 283 with known HIV status *vs* 48 [20%] of 238; p<0·0001).

Overall in cohort A, sensitivity of Ultra for detection of *M tuberculosis* was 87% (95% CI 76–94) and of Xpert was 81% (70–89; p=0·37; table 2). Ultra had reduced specificity compared with Xpert (90% [84–94] *vs* 99% [95–100]; p=0·001) and PPV (78% [67–87] *vs* 96% [87–99]; p=0·004). NPVs did not differ (p=0·57). By HIV status, sensitivity of Ultra for detection of *M tuberculosis* was similar in those who were HIV positive and those who were HIV negative (p=0·42). Ultra had reduced NPV in those who were HIV positive than in those who were HIV negative (p=0·030). Similar trends were observed for Xpert (table 2). In patients who were smear negative for acid fast bacilli, the sensitivity of Ultra and Xpert were similar (p=0·45). Testing with Xpert missed 14 patients who were culture positive, of whom seven (50%) were detected with Ultra (the remaining seven were also Ultra negative), and all three patients who were Ultra negative and culture positive were Xpert positive. Hence, Ultra resulted in an 8% (four of 52) increase in case detection. By HIV status, specificity of Ultra was reduced in those who were HIV positive versus those who were HIV negative (p=0·038; table 2). This reduction in specificity was higher in patients with previous tuberculosis who were HIV positive and culture negative than in those who were HIV negative and culture negative (16 [59%] of

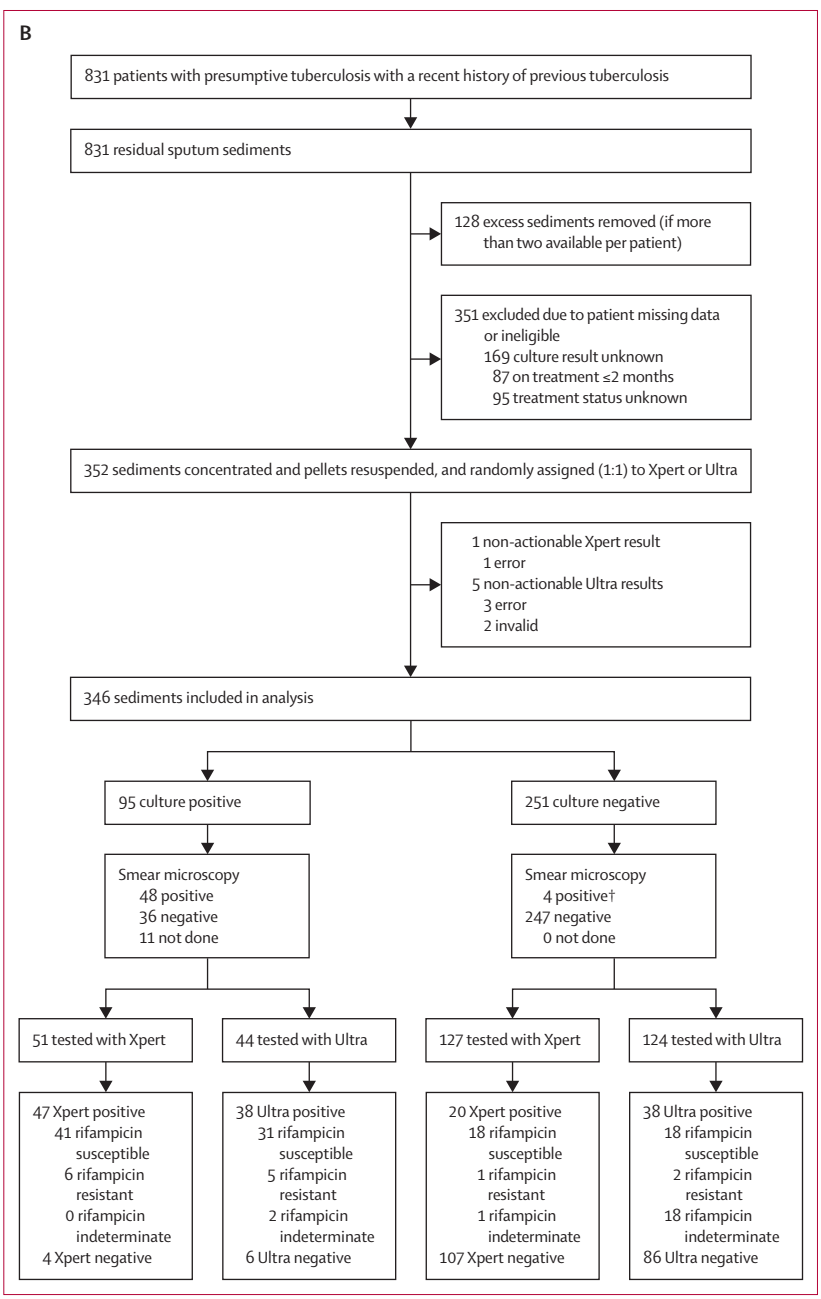


Figure 1: **Study profile for cohorts A and B**
Xpert=Xpert MTB/RIF. Ultra=Xpert MTB/RIF Ultra. *Three smear-positive, culture-negative samples were scanty (all were both Xpert negative and Ultra negative). †Three samples that were smear positive and culture negative were scanty and one was +1 grade (all scanty samples were Xpert negative or Ultra negative, and the +1 grade sample was Xpert positive).

	Overall (n=239)	Culture-positive patients			Culture-negative patients (n=167)
		All culture-positive patients (n=72)	Smear-positive patients (n=25)*	Smear-negative patients (n=43)*	
Demographics					
Age, years	37 (27–50)	36 (28–45)	33 (24–46)	36 (28–45)	39 (28–51)
Sex					
Female	116/239 (49%)	35/72 (49%)	13/25 (52%)	21/43 (49%)	81/167 (49%)
Male	123/239 (51%)	37/72 (51%)	12/25 (48%)	22/43 (51%)	86/167 (51%)
Ethnicity					
Mixed ancestry	209/239 (87%)	58/72 (81%)	22/25 (88%)	32/43 (74%)	151/167 (90%)
Black	30/239 (13%)	14/72 (19%)	3/25 (12%)	11/43 (26%)	16/167 (10%)
Tobacco smoker (past or current)	187/239 (78%)	54/72 (75%)	19/25 (76%)	33/43 (77%)	133/167 (80%)
Clinical					
HIV status†					
Positive	48/238 (20%)	21/72 (29%)	2/25 (8%)	17/43 (40%)	27/166 (16%)
Negative	190/238 (80%)	51/72 (71%)	23/25 (92%)	26/43 (60%)	139/167 (84%)
CD4, cells per μL‡	316 (178–503)	251 (137–397)	Insufficient data	293 (137–432)	323 (227–558)
TBscore II§	2 (2–3)	3 (2–4)	3 (2–4)	3 (3–5)	2 (2–3)
Haemoglobin, g/dL	14 (13–15)	13 (11–14)	13 (12–14)	13 (11–14)	15 (14–16)
Previous tuberculosis	94/239 (39%)	25/72 (35%)	6/25 (24%)	16/43 (37%)	69/167 (41%)

Data are median (IQR) or n/N (%). TBscore II=tuberculosis symptom score II. Xpert=Xpert MTB/RIF. Ultra=Xpert MTB/RIF Ultra. *Four patients had data missing for smear results. †One patient had data missing for HIV status. ‡Two patients had data missing for CD4 cell count. §13 patients had data missing for TBscore II.

Table 1: **Demographic and clinical characteristics of patients in cohort A, by culture and smear status**

27 *vs* 53 [38%] of 139; p=0·042). Specificity of Ultra was reduced in patients with previous tuberculosis treatment compared with those with no previous treatment (83% [95% CI 72–91] *vs* 94% [88–98]; p=0·022), whereas Xpert was unaffected (100% [95–100] *vs* 97% [92–100]; p=0·16; appendix pp 14–15). This effect persisted independently of HIV status: when analyses were restricted to only patients who were HIV positive, specificity of Ultra was higher in those without previous tuberculosis treatment than in those with previous tuberculosis treatment (100% [72–100] *vs* 63% [36–85]; p=0·021). Similar results were obtained in the non-head-to-head comparison of Ultra and Xpert (appendix p 13). We followed up patients who were positive via Ultra and culture negative (n=18; median of 435 days [IQR 347–533] until follow-up visit) and re-tested patients who were culture negative and initially positive by Xpert and Ultra (n=3) or only Ultra (n=15; none were positive by Xpert alone; appendix p 16). Only two (11%) of 18 patients who were followed up were treated programmatically in the interim, and of the remaining 16, five (31%) were unavailable (four lost to follow-up, one declined to participate further), four (36%) of 11 remained Ultra positive at re-testing (all four had previous tuberculosis), of whom three were now culture positive and one was still culture negative. The remaining seven (64%) of 11 patients were negative by Xpert, Ultra, and culture at follow-up. When testing the Ultra and sputum bacillary load, both Ultra IS*6110–1081* C_T and Ultra *rpoB* C_{Tmin} were proportional to traditional measures of sputum mycobacterial load (smear microscopy grade and culture time to positivity; figure 2).

After excluding non-actionable results in cohort B, from 346 eligible samples, 168 sediments were allocated to Ultra and 178 to Xpert. Overall sensitivity of Ultra for detection of *M tuberculosis* was 86% (95% CI 73–95) and of Xpert was 92% (81–98; p=0·36) and in samples that were smear negative sensitivity of Ultra was 76% (50–92) and of Xpert was 79% (54–93; p=0·86), with no differences by HIV status (table 3). Overall specificity for Ultra was 69% (60–77) and for Xpert was 84% (78–91; p=0·005). The PPV for Ultra was 50% (43–57) and for Xpert was 70% (61–78; p=0·014). The NPV for Ultra was 93% (87–97) and for Xpert was 96% (91–99; p=0·34). We estimated that in a four-culture scenario maximally generous to Ultra, only negligible improvement in specificity would be seen (appendix p 17).

In cohort A, of the 80 patients who were Ultra positive in the head-to-head comparison, 13 (16%) were trace positive, of whom six (46%) had previous tuberculosis and seven (54%) did not have previous tuberculosis. Similarly, in cohort B, of 76 patients who were Ultra positive, 21 (28%) were trace positive, all of whom had recent previous tuberculosis. In cohort A, for positive trace results that were re-categorised to negative, specificity of Ultra increased by 5% (95% CI 1 to 9; p=0·003) overall and by 7% (0 to 15; p=0·025) in patients who had been previously treated (table 4). Overall sensitivity of Ultra decreased (–6% [–14 to 1]; p=0·046]. Although trace re-categorisation resulted in both the

	All patients (n=239)				HIV-negative patients (n=190)				HIV-positive patients (n=48)			
	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
Smear microscopy	25/68 (37%, 26–50)	162/165 (99%, 95–100)	25/28 (90%, 72–98)	162/205 (80%, 73–85)	23/49 (47%, 33–62)	135/138 (98%, 94–100)	23/26 (89%, 70–98)	135/161 (84%, 78–90)	2/19 (11%, 2–34)	26/26 (100%, 87–100)	2/2 (100%, 16–100)	26/43 (61%, 45–76)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·005	p=0·45	p=0·61	p=0·001
Smear *vs* Xpert	p<0·0001	p=0·99	p=0·31	p<0·0001	p<0·0001	p=0·99	p=0·44	p=0·002	p<0·0001	p>0·99	p>0·99	p=0·08
Xpert												
All samples	58/72 (81%, 70–89)	164/167 (99%, 95–100)	58/61 (96%, 87–99)	164/178 (93%, 88–96)	44/51 (87%, 74–95)	136/139 (98%, 94–100)	44/47 (94%, 83–99)	136/143 (96%, 91–99)	14/21 (67%, 44–86)	27/27 (100%, 88–100)	14/14 (100%, 77–100)	27/34 (80%, 63–92)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·06	p=0·44	p=0·33	p=0·002
Smear negative patients	31/43 (73%; 57–85)	159/162 (99%; 95–100)	31/34 (92%; 77–99)	159/171 (93%; 89–97)	20/26 (77%; 57–92)	132/135 (98%; 94–100)	20/23 (87%; 67–98)	132/138 (96%; 91–99)	11/17 (65%; 39–86)	26/26 (100%;87–100)	11/11 (100%;72–100)	26/32 (82%; 64–93)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·38	p=0·44	p=0·21	p=0·004
Ultra												
All samples	62/72 (87%; 76–94)	149/167 (90%; 84–94)	62/80 (78%; 67–87)	149/159 (94%; 89–97)	45/51 (89%; 77–96)	127/139 (92%; 86–96)	45/57 (79%; 67–89)	127/133 (96%; 91–99)	17/21 (81%; 59–95)	21/27 (78%; 58–92)	17/23 (74%; 52–90)	21/25 (84%; 64–96)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·42	p=0·038	p=0·626	p=0·030
Ultra *vs* Xpert	p=0·37	p=0·001	p=0·004	p=0·57	p=0·77	p=0·017	p=0·034	p=0·88	p=0·29	p=0·009	p=0·037	p=0·66
Smear negative patients	34/43 (80%; 64–90)	144/162 (89%; 84–94)	34/52 (66%; 51–79)	144/153 (95%; 90–98)	21/26 (81%; 61–94)	123/135 (92%; 85–96)	21/33 (64%; 46–80)	123/128 (97%; 92–99)	13/17 (77%; 51–94)	20/26 (77%; 57–92)	13/19 (69%; 44–88)	20/24 (84%; 63–96)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·74	p=0·036	p=0·73	p=0·015
Smear-negative Ultra *vs* smear-negative Xpert	p=0·45	p=0·001	p=0·007	p=0·68	p=0·73	p=0·017	p=0·053	p=0·86	p=0·45	p=0·009	p=0·037	p=0·84
Xpert negative patients	7/14 (50%; 24–77)	149/164 (91%; 86–95)	7/22 (32%; 14–55)	149/156 (96%; 91–99)	2/7 (29%; 4–71)	127/136 (94%; 88–97)	2/11 (19%; 3–52)	127/132 (97%; 92–99)	5/7 (72%; 30–97)	21/27 (78%; 58–92)	5/11 (46%; 17–77)	21/23 (92%; 72–99)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·11	p=0·010	p=0·17	p=0·29

Data are n/N and % and 95% CI in parentheses, with p values where appropriate. Data are missing for one patient for HIV status and from six patients for smear results. Likelihood ratios are in the appendix (p 12). NPV=negative predictive value. PPV=positive predictive value. Xpert=Xpert MTB/RIF. Ultra=Xpert MTB/RIF Ultra.

Table 2: **Head-to-head comparison of Xpert and Ultra accuracy in cohort A, by smear and HIV status**

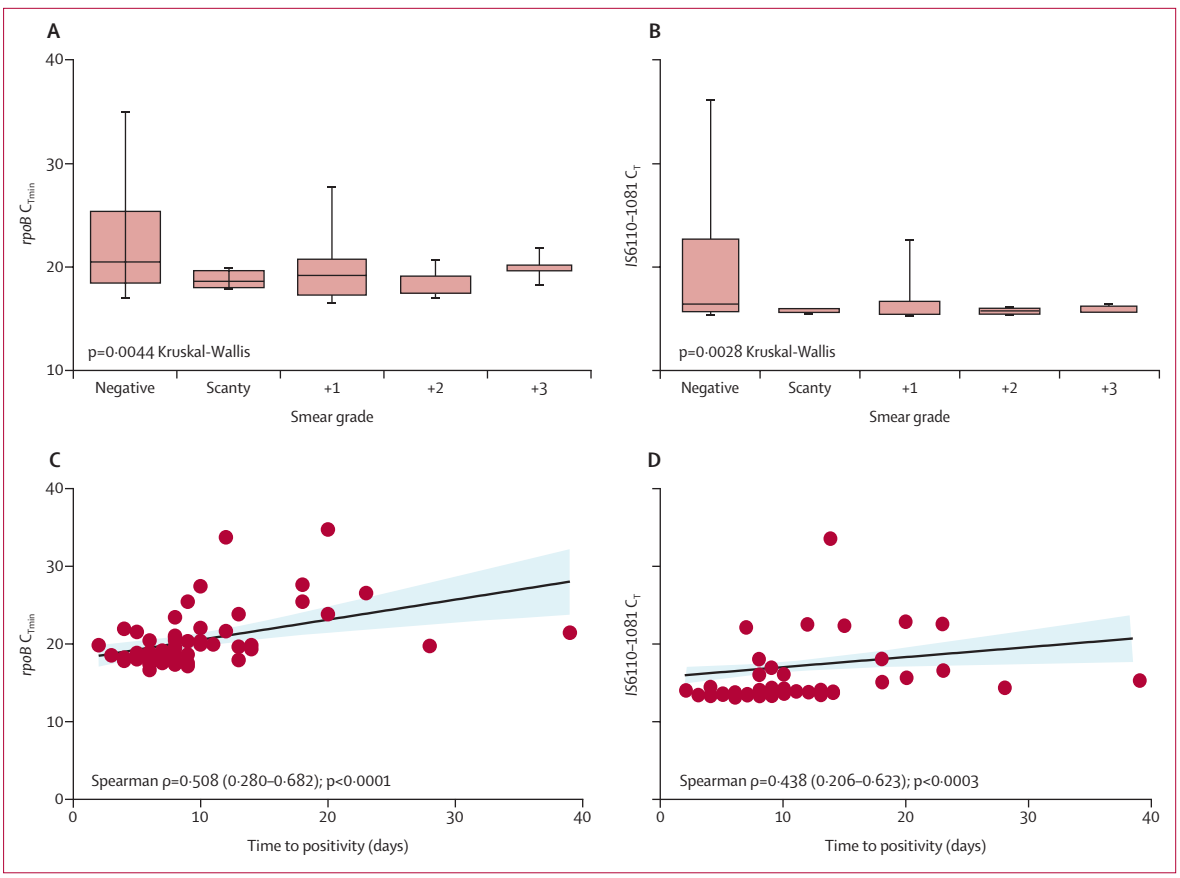


Figure 2: **Quantitative information for both *rpoB* and *IS6110–1081* probes of Ultra compared with traditional measures of bacillary load (ie, smear and culture)**
Both readouts inversely correlated with smear grade (A and B), and positively correlated with time to positivity (C and D). Datapoints are individual patient values. Dotted lines are the linear regression line and shaded areas are 95% CIs. C_T=cycle threshold. C_{Tmin}=C_T minimum. Ultra=Xpert MTB/RIF Ultra.

sensitivity and specificity of Ultra being similar to those of Xpert (overall and in patients without previous tuberculosis), the specificity of Ultra in patients who had been previously treated was still reduced compared with that of Xpert (90% [81 to 96] *vs* 100% [95 to 100]; p=0·007; table 4; appendix pp 14–15). In cohort B, the specificity of Ultra increased after re-classification (15% [8 to 22]; p<0·0001), rendering Ultra's specificity similar to that of Xpert without a significant reduction in Ultra's sensitivity (–5% [–13 to 4]; p=0·16; table 4; appendix pp 14–15). The proportion of Ultra positive samples with a trace result, by culture status, are in the appendix (p 6). For both cohorts, improvements in specificity after trace exclusion, similar to those re-classified to negative, were seen but without reductions in sensitivity (table 4). In cohort A, specificity of Ultra was similar to Xpert after trace exclusion (p=0·06).

In cohort A, 28 (10%) of 275 Ultra results and 14 (5%) of 301 Xpert results were non-actionable before re-testing (p=0·011; denominators here differ to those reported in the head-to-head diagnostic accuracy analysis, which, by definition, only included actionable results). The most frequent error code in Ultra was #2008 (ie, assay syringe pressure too high; n=18; appendix pp 18–19). The proportion of actionable and non-actionable results did not differ by patient or sample clinical characteristics (appendix p 20). Re-testing of the remaining specimen sample reagent mix was possible for 19 (68%) of 28 samples with Ultra and none for Xpert. Of those re-tested, 11 (58%) of 19 became actionable. In cohort B, in which samples were pre-processed, the proportion of results that were non-actionable for Ultra was 3% (five of 173) and for Xpert was 1% (one of 179; p=0·09).

In cohort A, no MTBDR*plus*-rifampicin resistant cases were observed (appendix pp 6–7) and sensitivity of this measure could not be calculated. Two (3%) of 69 samples that were Xpert positive were rifampicin resistant as detected by Xpert and both of these were MTBDR*plus*-rifampicin susceptible (one sample was Ultra negative and one was Ultra positive and rifampicin susceptible as detected by Ultra). Specificity of Ultra for detection of rifampicin resistance was 100% (54 of 54) versus 96% (51 of 53; p=0·15) for Xpert. In cohort B, of six patients who were Ultra positive with an MTBDR*plus*-rifampicin resistant result, five (83%) were rifampicin resistant as detected by Ultra (one patient was rifampicin susceptible

as detected by Ultra and MTBDR*plus* rifampicin resistant). More details on drug susceptibility results are in the appendix (p 7). In cohort A, 14 (17%) of 80 patients that were Ultra positive were rifampicin indeterminate, of whom 13 (93%) were trace (we included all results from Ultra in this analysis, not just those in the head-to-head comparison). Most patients who were rifampicin indeterminate were culture negative (ten [71%] of 14); precluding confirmatory drug resistance testing. Xpert detected rifampicin susceptibility in three patients who were rifampicin indeterminate according to Ultra. In cohort B, 21 (28%) of 76 samples that were Ultra positive were rifampicin indeterminate (all trace) and, like cohort A, most were culture negative (19 [90%] of 21).

When analysing by previous tuberculosis treatment status, using cohort A sensitivity and specificity estimates, we projected that, by contrast with the higher predictive values in patients without previous tuberculosis (figure 3A), 45% (4500 of 10 000) of patients who were Ultra positive with previous tuberculosis (figure 3B) would be culture positive at a 15% pre-test probability (meaning 55% of results that are Ultra positive will be culture negative). Similarly, in cohort B at the same pre-test probability, 33% of results that are Ultra positive would be culture positive (hence 67% would be culture negative; figure 3C). Using the tuberculosis and previous tuberculosis treatment prevalences, the proportion of results that are Ultra positive that are also culture positive will increase to 49% if all trace positive patients were re-classified or excluded. Estimates were similar when sensitivities from the Ultra multicentre evaluation[6] were used (appendix p 8). When analysing projected PPVs among patients with recent previous tuberculosis (figure 4), the area of shading reflecting PPVs of more than 70% is mostly absent for Ultra but appears across a greater prevalence range for Xpert. For example, in a scenario where 10% of patients who have been previously treated have recent previous tuberculosis, PPVs for Ultra that are 70% or higher will only be attained at high prevalences (≥30%; figure 4A). PPVs were boosted by approximately 10% across most scenarios by trace re-classification (figure 4B) but remained reduced compared with Xpert (figure 4C). By contrast, when data from the Ultra multicentre evaluation were used (in which specificity, including in patients with recent previous tuberculosis, was higher than in our study),[6] estimated PPVs were, as expected, generally higher for Ultra and Xpert (appendix p 9) in our modelling exercise, which highlights the need for future research.

Discussion

In our assessment of the accuracy of Ultra and Xpert in an HIV-endemic setting with a high burden of previous tuberculosis, our key finding is that specificity and PPV of Ultra are lower than those of Xpert in patients who have been previously treated for tuberculosis.

	All patients (n=346)				HIV-negative patients (n=158)				HIV-positive patients (n=121)			
	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
Xpert												
All samples	47/51 (92%; 81–98)	107/127 (84%; 78–91)	47/67 (70%; 61–78)	107/111 (96%; 91–99)	15/16 (94%; 70–99)	43/53 (81%; 68–91)	15/25 (60%; 39–79)	43/44 (98%; 88–99)	12/13 (92%; 64–99)	50/58 (86%; 75–94)	12/20 (60%; 30–70)	50/51 (98%; 90–99)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·88	p=0·47	p>0·99	p=0·92
Smear negative samples	15/19 (79%; 54–93)	105/124 (85%; 77–90)	15/34 (44%; 27–62)	105/109 (96%; 90–99)	6/7 (86%; 42–99)	42/51 (82%; 69–91)	6/15 (40%; 17–67)	42/43 (98%; 86–99)	4/5 (80%; 30–99)	49/57 (86%; 74–93)	4/12 (33%; 11–64)	49/50 (98%; 88–99)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·79	p=0·61	p=0·72	p=0·91
Smear-negative Xpert *vs* smear-negative Ultra	p=0·86	p=0·004	p=0·07	p<0·0001	p=0·09	p=0·08	p=0·005	p=0·67	p=0·86	p=0·06	p=0·43	p=0·71
Ultra												
All samples	38/44 (86%; 73–95)	86/124 (69%; 60–77)	38/76 (50%; 43–57)	86/92 (93%; 87–97)	13/16 (81%; 54–96)	50/73 (68%; 57–79)	13/36 (36%; 21–54)	50/53 (94%; 84–99)	7/8 (87%; 47–99)	30/42 (71%; 55–84)	7/19 (37%; 16–62)	30/31 (97%; 83–99)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·69	p=0·74	p=0·96	p=0·61
Ultra *vs* Xpert	p=0·36	p=0·005	p=0·014	p=0·34	p=0·28	p=0·11	p=0·07	p=0·40	p=0·72	p=0·07	p=0·15	p=0·72
Smear negative samples	13/17 (76%; 50–92)	85/123 (69%; 60–77)	13/51 (25%; 14–39)	85/123 (95%; 88–98)	1/3 (33%; 2–87)	50/73 (68%; 57–78)	1/24 (4%; 0–23)	50/52 (96%; 86–99)	3/4 (75%; 22–99)	29/41 (71%; 54–83)	3/15 (20%; 5–49)	29/30 (97%; 81–99)
HIV positive *vs* HIV negative	..	..	..	..	..	..	..	..	p=0·27	p=0·80	p=0·11	p=0·90

Data are n/N (%, 95% CI). Data are missing from 67 samples for HIV status and for 11 samples for smear status. Likelihood ratios are in the appendix (p 12). PPV=positive predictive value. NPV=negative predictive value. Xpert=Xpert MTB/RIF. Ultra=Xpert MTB/RIF Ultra.

Table 3: **Xpert and Ultra diagnostic accuracy in cohort B, stratified by HIV status**

	Cohort A									Cohort B		
	All patients (n=239)			No previous tuberculosis (n=145)			Previous tuberculosis (n=94)			All previous tuberculosis (n=168)		
	Overall	Change in trace reclassified	Change in trace excluded	Overall	Change in trace reclassified	Change in trace excluded	Overall	Change in trace reclassified	Change in trace excluded	Overall	Change in trace reclassified	Change in trace excluded
Sensitivity	62/72 (87%; 76 to 94)	58/72 (81%; 70 to 89)	58/68 (86%; 75 to 93)	42/47 (90%; 77 to 97)	39/47 (83%; 70 to 93)	39/44 (89%; 76 to 97)	20/25 (80%; 60 to 94)	19/25 (76%; 55 to 91)	19/24 (80%; 58 to 93)	38/44 (86%; 73 to 95)	36/44 (82%; 56 to 90)	36/42 (86%; 71 to 95)
Difference after re-categorisation	..	−6% (−14 to 1)	−1% (−12 to 11)	..	−6% (−15 to 3)	−1% (−13 to 10)	..	−4% (−16 to 8)	−1% (−23 to 22)	..	−5% (−13 to 4)	−1% (−15 to 14)
Associations												
No previous tuberculosis *vs* previous tuberculosis*	..	..	..	..	..	..	p=0·27	p=0·48	p=0·29	..	..	..
Trace reclassified or trace excluded *vs* Ultra	..	p=0·046	p=0·89	..	p=0·083	p=0·91	..	p=0·32	p=0·94	..	p=0·16	p=0·93
Trace reclassified *vs* excluded	..	..	p=0·46	..	..	p=0·44	..	..	p=0·79	..	..	p=0·62
Ultra *vs* Xpert†	..	p>0·99	p=0·46	..	p=0·60	p=0·44	..	p=0·48	p=0·66	..	p=0·13	p=0·32
Specificity	149/167 (90%; 84 to 94)	158/167 (95%; 91 to 98)	149/158 (95%; 90 to 98)	92/98 (94%; 88 to 98)	96/98 (98%; 93 to 100)	92/94 (98%; 93 to 100)	57/69 (83%; 72 to 91)	62/69 (90%; 81 to 96)	57/64 (90%; 79 to 96)	86/124 (69%; 60 to 77)	105/124 (85%; 78 to 91)	86/105 (82%; 73 to 89)
Difference after re-categorisation	..	5% (1 to 9)	5% (−1 to 11)	..	4% (−1 to 9)	4% (−2 to 10)	..	7% (0 to 15)	6% (−5 to 18)	..	15% (8 to 22)	13% (2 to 24)
Associations												
No previous tuberculosis *vs* previous tuberculosis*	..	..	..	..	..	..	p=0·021	p=0·022	p=0·019	..	..	..
Trace reclassified or trace excluded *vs* Ultra	..	p=0·0030	p=0·10	..	p=0·046	p=0·17	..	p=0·025	p=0·29	..	p<0·0001	p=0·029
Trace reclassified *vs* excluded	..	..	p=0·90	..	..	p=0·17	..	..	p=0·88	..	..	p=0·57
Ultra *vs* Xpert†	..	p=0·078	p=0·062	..	p=0·65	p=0·68	...	p=0·007	p=0·005	..	p=0·93	p=0·63
PPV	62/80 (78%; 67 to 87)	58/67 (87%; 77 to 94)	58/67 (87%; 77 to 94)	42/48 (88%; 75 to 96)	39/41 (96%; 84 to 100)	39/41 (96%; 84 to 100)	20/32 (63%; 44 to 79)	19/26 (74%; 53 to 89)	19/26 (74%; 53 to 89)	38/76 (50%; 43 to 57)	36/55 (65%; 51 to 78)	36/55 (65%; 51 to 78)
Difference after re-categorisation	..	9% (−3 to 21)	9% (−3 to 21)	..	8% (−4 to 19)	8% (−4 to 19)	..	11% (−13 to 34)	11% (−13 to 34)	..	15% (−1 to 32)	15% (−1 to 32)
Associations												
No previous tuberculosis *vs* previous tuberculosis*	..	..	..	..	..	..	p=0·0090	p=0·010	p=0·010	..	..	..
Trace reclassified or trace excluded *vs* Ultra	..	p=0·16	p=0·16	..	p=0·21	p=0·21	..	p=0·39	p=0·39	..	p=0·078	p=0·078
Trace reclassified *vs* excluded	..	..	p>0·99	..	..	p>0·99	..	..	p>0·99	..	..	p>0·99
Ultra *vs* Xpert†	..	p=0·099	p=0·099	..	p=0·62	p=0·62	..	p=0·010	p=0·010	..	p=0·58	p=0·58

(Table 4 continues on next page)

Additionally, we found that these measures decreased further in patients with recent previous tuberculosis to the extent that half of patients who were Ultra positive were culture negative for *M tuberculosis*, and this result was only partly improved by trace re-categorisation. Furthermore, PPV will likely remain suboptimal across a wide prevalence range in patients who have been previously treated. Patients who are Ultra trace positive, who comprise a meaningful proportion of all patients who are Ultra positive (16% in cohort A and 28% in cohort B) and who are indeterminate for rifampicin resistance according to Ultra, are often culture negative—ruling out further drug-resistance testing. Finally, Ultra more frequently gave non-actionable results than did Xpert. Our data have important implications for clinical decision making and settings considering Ultra implementation.

Notably, in cohort A the specificity and PPV of Ultra were lower than those of Xpert in patients who had been previously treated: 17% of patients who were culture negative were Ultra positive, and 63% of patients who were Ultra positive were culture negative. These metrics decreased further in patients with recent previous tuberculosis (cohort B), probably due to old tuberculosis DNA that the patient's body has had less opportunity to clear. This diminished specificity is a trade-off resulting from the improved limit of detection of Ultra compared with Xpert and we expect that similar effects will be observed for other emerging molecular tuberculosis assays with high sensitivity.

Specificity of Ultra was unlikely to have been underestimated due to imperfect sensitivity of the reference standard in our study because, even when we re-classified the highest permissible number of Ultra false-positives to true-positive, specificity did not significantly increase in the analysis with the four-culture reference standard. Previously, we described how most symptomatic patients who were Xpert positive and culture negative, after exhaustive sampling including bronchoalveolar lavage and clinical follow-up, were unlikely to have active tuberculosis.[25] Our cohort A follow-up data support this interpretation, with almost two-thirds of patients who were Ultra false-positive transitioning to true negative during a year without treatment. However, some patients who were initiated on treatment programmatically in the interim or later became culture positive, suggesting further research is needed to better understand how these patients should be managed. Thus, in our setting, a positive Ultra result in patients who have been previously treated is of little diagnostic use due to the high likelihood of false positives and low PPV.[17] From our PPV estimates at different pre-test probabilities, this finding likely holds true across most settings with substantial burdens of recent previous tuberculosis, which usually correlates with tuberculosis prevalence. Our PPV estimates also differed to those obtained when data from a multicentre evaluation across multiple settings were used. This

	Cohort A							Cohort B						
	All patients (n=239)			No previous tuberculosis (n=145)			Previous tuberculosis (n=94)			All previous tuberculosis (n=168)				
	Overall	Change in trace reclassified	Change in trace excluded	Overall	Change in trace reclassified	Change in trace excluded	Overall	Change in trace reclassified	Change in trace excluded	Overall	Change in trace reclassified	Change in trace excluded
(Continued from previous page)												
NPV	149/159 (94%; 89 to 97)	158/172 (92%; 87 to 96)	149/159 (94%; 89 to 97)	92/97 (95%; 89 to 99)	96/104 (93%; 86 to 97)	92/97 (95%; 89 to 99)	57/62 (92%; 83 to 98)	62/68 (92%; 82 to 97)	57/62 (92%; 83 to 98)	86/92 (93%; 87 to 97)	105/113 (93%; 86 to 97)	86/92 (93%; 87 to 97)
Difference after re-categorisation	..	−2% (−7 to 4)	0% (−5 to 5)	..	−3% (−9 to 4)	0% (−6 to 6)	..	−1% (−10 to 9)	0% (−10 to 10)	..	−1% (−7 to 6)	0% (−7 to 7)
Associations												
No previous tuberculosis *vs* previous tuberculosis*	..	..	..	..	..	..	p=0·46	p=0·79	p=0·46	..	..	..
Trace reclassified or trace excluded *vs* Ultra	..	p=0·52	p>0·99	..	p=0·46	p>0·99	..	p=0·88	p>0·99	..	p=0·88	p>0·99
Trace reclassified *vs* excluded	..	..	p=0·52	..	..	p=0·46	..	..	p=0·88	..	..	p=0·87
Ultra *vs* Xpert†	..	p=0·92	p=0·57	..	p=0·64	p=0·24	..	p=0·44	p=0·55	..	p=0·25	p=0·34

Data are n/N (%; 95% CI), difference with 95% CI in parentheses or p value. Both the re-calculated estimate and the change are shown after re-categorisation. NPV=negative predictive value. PPV=positive predictive value. Xpert=Xpert MTB/RIF. Ultra=Xpert MTB/RIF Ultra. *McNemar's test or test of proportion. †Xpert values in table 2 and the appendix (p 14) for cohort A and table 3 for cohort B.

Table 4: **Effects of previous tuberculosis and different trace re-categorisation strategies on Ultra diagnostic accuracy in cohorts A (head-to-head data) and B**

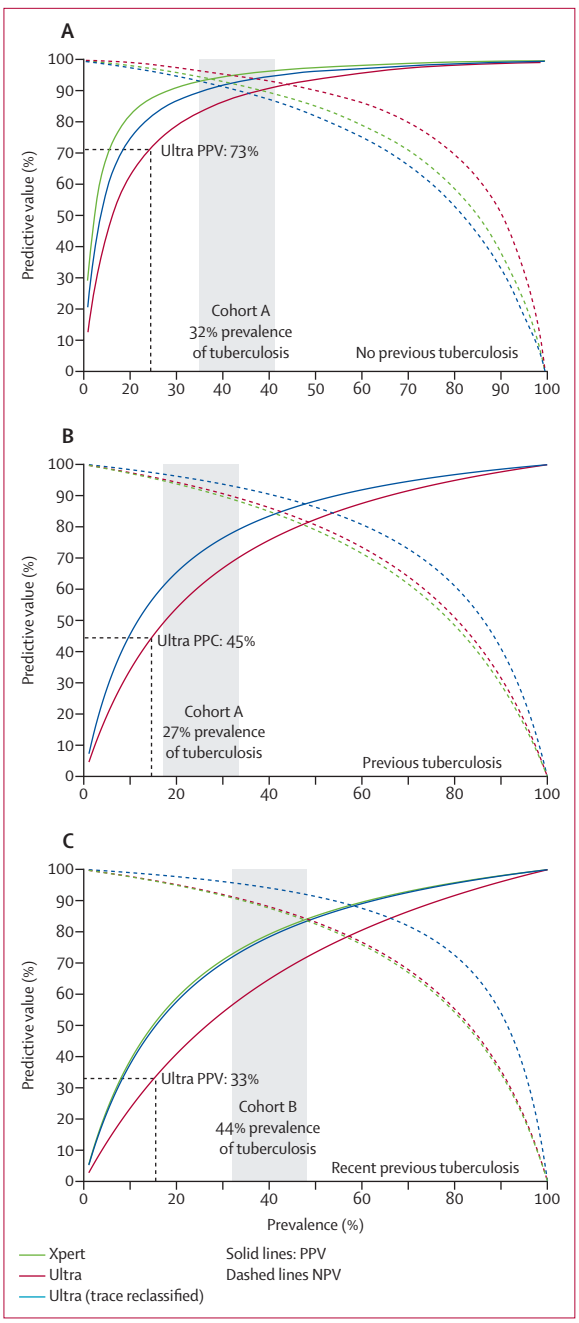


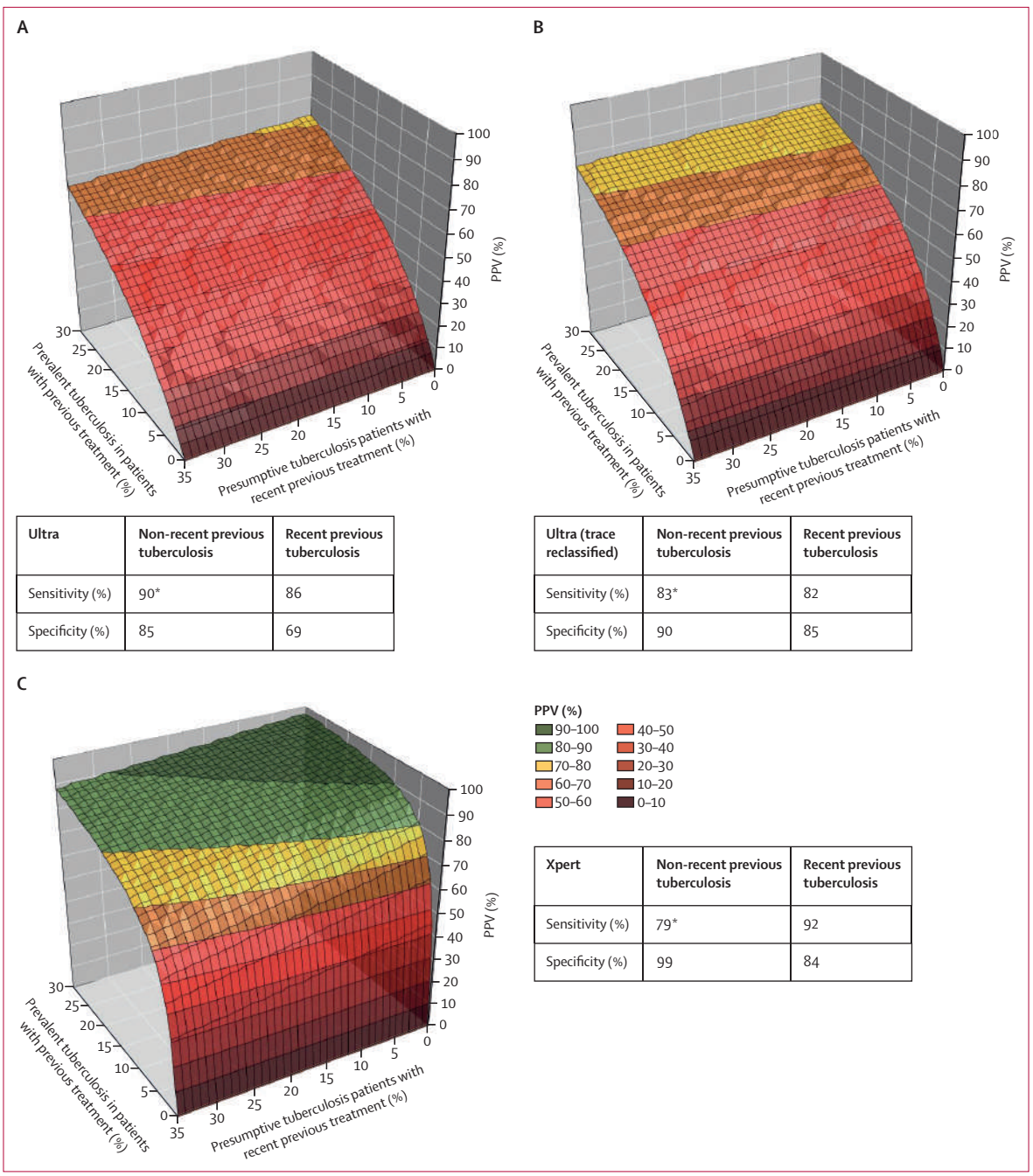
Figure 3: **Xpert and Ultra PPVs and NPVs by pre-test probability, after stratification by tuberculosis history and recency within 2 years**
(A) Analysis using cohort A sensitivity and specificity estimates in patients without previous tuberculosis and (B) analysis using estimates from patients who had been previously treated for tuberculosis. (C) Analysis using cohort B sensitivity and specificity estimates. PPVs for Ultra (trace excluded) are not shown because they are near identical to Ultra (trace reclassified), which themselves resemble those of Xpert. Xpert in panel B had 100% specificity and PPV is not shown. Our observed rate of culture positivity is shown with 95% CI as a grey column. Dotted black lines that meet the Ultra PPV curve are illustrative examples of the expected PPV of Ultra at a culture-positive prevalence of 15%. NPV=negative predictive value. PPV=positive predictive value. Xpert=Xpert MTB/RIF. Ultra=Xpert MTB/RIF Ultra.

difference is probably driven by the differing proportions of patients with previous tuberculosis in the two studies. Our findings might not be applicable to a setting with a low prevalence of tuberculosis where, notably, guidance recommends that a confirmatory culture is done in patients who are Ultra positive.[26]

Ultra-positive non-trace results are, by definition, probably detectable by Xpert. Thus, interpretation of Ultra's potentially biggest source of improvement (trace) needs to be carefully considered and not discarded by programmes. We showed that both re-categorisation strategies (re-classification and exclusion) result in similar changes in specificity for relatively small sensitivity costs, with the highest specificity gain in patients with recent previous tuberculosis treatment. However, in cohort A, trace re-categorisation did not fully restore specificity of Ultra to the level of Xpert, indicating that trace is probably not the only cause of Ultra's diminished specificity (modifications to improve sensitivity other than the IS*6110* and IS*1081* probes probably have a role[5]). For example, after re-categorisation, 26% of Ultra positive patients with previous tuberculosis treatment will still be culture negative in cohort A. However, further research is needed because, in cohort B, trace re-categorisation restored the specificity of Ultra to the same level as Xpert.

We observed high indeterminate rates for detection of rifampicin resistance by Ultra among patients who were Ultra positive: 17% in cohort A and 28% in cohort B. Almost all indeterminate samples were culture negative. Thus, countries with a high burden of drug-resistant tuberculosis need to prepare to manage large numbers of patients with increased risks of resistance (patients who have been previously treated) and unknown rifampicin susceptibilities—notably, some of these cases would not have been identified as tuberculosis by Xpert in the first place. Research is needed on the appropriate drug susceptibility testing algorithm in such patients; however, downstream tests like MTBDR*plus* or FluroType MTBDR[27] are unlikely to work given the low bacillary load of the samples. Furthermore, indirect confirmatory testing is unlikely to be useful given the absence of conventionally culturable bacilli.

Importantly, improvements in Ultra sensitivity will be partially undermined by the higher rate of non-actionable results than Xpert, which reduces operational sensitivity.[28] In cohort A, this rate was higher than in the previous multicentre evaluation,[6] but, in line with that study, double the rate of Xpert. Although iterative design improvements to Xpert reduced initially high non-actionable rates, non-actionable result rates increase with time under programmatic conditions.[29] Whether future improvements to Ultra reduce the rate of non-actionable results remains to be seen, but safeguards to minimise their occurrence should be strengthened.[30,31] Additionally, we showed that by re-testing remnant sputum sample buffer mix, the



Ultra	Non-recent previous tuberculosis	Recent previous tuberculosis
Sensitivity (%)	90*	86
Specificity (%)	85	69

Ultra (trace reclassified)	Non-recent previous tuberculosis	Recent previous tuberculosis
Sensitivity (%)	83*	82
Specificity (%)	90	85

Xpert	Non-recent previous tuberculosis	Recent previous tuberculosis
Sensitivity (%)	79*	92
Specificity (%)	99	84

Figure 4: **Projections of PPVs for Ultra (A), Ultra (trace reclassified; B), and Xpert (C) for varying (assumed) levels of prevalent tuberculosis (pre-test probability) and proportions of patients with recent previous tuberculosis (≤2 years), among presumptive patients who have been previously treated for tuberculosis**

Horizontal coloured bands from the right wall show PPV ranges, lines from the left wall show pre-test probability, and vertical lines from the base that intersect coloured rows show the proportion of patients who were previously treated for tuberculosis with recent previous tuberculosis treatment. PPVs were calculated using estimates of test specificity among patients with non-recent previous tuberculosis (derived from cohort A) and among patients with recent tuberculosis (derived from cohort B). PPV=positive predictive value. Xpert=Xpert MTB/RIF. Ultra=Xpert MTB/RIF Ultra. *Sensitivity estimates from patients with no previous tuberculosis in cohort A were used.

proportion of patients who will not receive an Ultra diagnosis due to non-actionable results will be improved from approximately 10% to about 6% after re-testing; an approach that is, to our knowledge, not widely adopted.

Finally, we detected a trend towards higher sensitivity with Ultra than with Xpert but had large uncertainty in our estimates. These large uncertainty intervals are probably due to an insufficient number of cases that

were smear negative to detect the relatively small sensitivity increases offered by Ultra in this specific population who are self-presenting and symptomatic. By contrast, Ultra might have larger sensitivity increments than Xpert in patients with earlier stage disease than we assessed here. More data are thus needed on whether programmes seeking to streamline their previously Xpert-orientated diagnostic algorithms should remove the previous recommendation for culture in symptomatic patients who are HIV positive and Ultra negative. Such additional testing should be considered in the context that it is useful for clinical decision making, the appropriateness and extent of empirical treatment,[32] and whether diagnostic algorithms with added complexity and delay are adequately adhered to by health workers.[33]

This study has several limitations. We tested patients with a high prevalence of previous tuberculosis treatment (39% in cohort A) and patients with a recent history of tuberculosis treatment (cohort B), so our findings are mostly applicable to these or similar populations. Nevertheless, our findings on specificity among patients with a history of tuberculosis are probably generalisable. We used a single culture reference standard but, notably, the four-culture reference standard with a best-case Ultra simulation did not significantly affect sensitivity and specificity estimates; suggesting conclusions would remain unchanged under such conditions. In cohort A, Ultra and Xpert testing were done on separate sputum samples. Given that sputum quality and mycobacterial load can differ between samples, this step is a potential source of bias. We also used induced sputum and results might have differed if samples were expectorated, and respiratory rate was not incorporated into TBscore II as per usual practice. Although we followed up patients who were Ultra positive and culture negative in cohort A, a need exists for a longitudinal study that follows up and regularly re-samples patients with false-positive Ultra results, especially those with a trace result. Such follow-up would inform whether such patients might still benefit from a form of tuberculosis treatment. Also, we did multiple statistical tests that increase type I error. Finally, a strength of our study is that our eligibility criteria are aligned to those used by the local tuberculosis programme in South Africa. Thus, our patients resemble those seen routinely; however, a negative consequence of this aspect is that, in cohort B, some patients were missing data that are programmatically obtained, such as HIV status.

In summary, Ultra has large reductions in specificity in patients who have been previously treated for tuberculosis and other emerging high-sensitivity molecular tuberculosis assays will likely be similarly affected. Hence, diagnostic algorithms will need to become more complex, requiring consideration of previous tuberculosis treatment status, how trace results are interpreted, and how additional testing for rifampicin resistance can be implemented in patients with trace amounts of *M tuberculosis*. Clinical decision making and requirements for laboratory and health worker training will also be affected, and the choice between use of Xpert or Ultra requires careful consideration.

Contributors

GT conceived the study and PDvH acquired funding. HM, BWPR, and GT did analyses and wrote the first draft of the manuscript. FM led the analysis of PPV projections and generated the relevant figures. All authors critiqued analyses and revised the manuscript.

Declaration of interests

GT received in-kind donations (GeneXpert cartidges and machines) from Cepheid and FIND. Cepheid had no role in study design or interpretation of results. BWPR received travel support from Cepheid to attend a conference and present unrelated data. All other authors declare no competing interests.

Data sharing

The de-identified patient data, study protocol, informed consent form, and datasets generated during and/or analysed during the study are available from the corresponding author on reasonable request.

Acknowledgments

We thank Charlotte Lawn, Dandelene Sylvester, and Marina Ferreira for their diligent sample collection and dedication to the improvement of patient care; Happy Tshivhula and Brigitta Derendinger from the laboratory team for processing and testing patient samples; and Kim Stanley and Gian van der Spuy for RedCap support. We also thank the patients that consented to participate in the study. GT acknowledges funding for this study from South African Medical Research Council (SAMRC Flagship Project MRC-RFA-IFSP-01-2013), the EDCTP2 program supported by the European Union (grant SF1401, OPTIMAL DIAGNOSIS), the Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. BWPR and HM acknowledge funding from the Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.

References

1 Peter JG, Theron G, Singh N, Singh A, Dheda K. Sputum induction to aid diagnosis of smear-negative or sputum-scarce tuberculosis in adults in HIV-endemic settings. *Eur Respir J* 2014; **43:** 185–94.

2 Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2014; **1:** CD009593.

3 García-Basteiro AL, DiNardo A, Saavedra B, et al. Point of care diagnostics for tuberculosis. *Pulmonology* 2018; **24:** 73–85.

4 WHO. WHO meeting report of a technical expert consultation: non-inferiority analysis of Xpert MTB/RIF Ultra compared to Xpert MTB/RIF. Geneva: World Health Organization, 2017.

5 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.

6 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.

7 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.

8 Berhanu RH, David A, da Silva P, et al. Performance of Xpert MTB/RIF, Xpert Ultra, and Abbott Real Time MTB for the diagnosis of pulmonary tuberculosis in a high HIV burden setting. *J Clin Microbiol* 2018; **56:** e00560–18.

9 FIND. Report for WHO: a multicentre non-inferiority diagnostic accuracy study of the Ultra assay compared to the Xpert MTB/RIF assay. Version 1.8. Geneva: FIND, February, 2017. https://www.finddx.org/publication/ultra-report/ (accessed Dec 5, 2019).

10 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.

11 Marx FM, Dunbar R, Enarson DA, et al. The temporal dynamics of relapse and reinfection tuberculosis after successful treatment: a retrospective cohort study. *Clin Infect Dis* 2014; **58:** 1676–83.

12 Nunn AJ, Phillips PP, Mitchison DA. Timing of relapse in short-course chemotherapy trials for tuberculosis. *Int J Tuberc Lung Dis* 2010; **14:** 241–42.

13 Marx FM, Yaesoubi R, Menzies NA, et al. Tuberculosis control interventions targeted to previously treated people in a high-incidence setting: a modelling study. *Lancet Glob Health* 2018; **6:** e426–35.

14 Kendall EA, Schumacher SG, Denkinger CM, Dowdy DW. Estimated clinical impact of the Xpert MTB/RIF Ultra cartridge for diagnosis of pulmonary tuberculosis: a modeling study. *PLoS Med* 2017; **14:** e1002472.

15 García-Basteiro AL, Saavedra B, Cobelens F. The good, the bad and the ugly of the next-generation Xpert Mtb/Rif Ultra test for tuberculosis diagnosis. *Arch Bronconeumol* 2017; **53:** 665–66.

16 WHO. Global tuberculosis report. Geneva: World Health Organization, 2019.

17 Wood R, Lawn SD, Caldwell J, Kaplan R, Middelkoop K, Bekker L-G. Burden of new and recurrent tuberculosis in a major South African city stratified by age and HIV-status. *PLoS One* 2011; **6:** e25098.

18 WHO. Tuberculosis (TB). Systematic screening for active tuberculosis: principles and recommendations. Geneva: World Health Organization, 2019. https://www.who.int/tb/tbscreening/en/ (accessed Dec 5, 2019).

19 WHO. Definitions and reporting framework for tuberculosis: 2013 revision, updated December 2014. Geneva: World Health Organization, 2013 (accessed on Aug 19, 2019).

20 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.

21 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42:** 377–81.

22 Rieder HL, Martin Chonde T, Myking H, et al. The public health service national tuberculosis reference laboratory and the national laboratory network: minimum requirements, role and operation in a low-income country. Paris: International Union Against Tuberculosis and Lung Disease, 1998.

23 Shenai S, Ronacher K, Malherbe S, et al. Bacterial loads measured by the Xpert MTB/RIF assay as markers of culture conversion and bacteriological cure in pulmonary TB. *PLoS One* 2016; **11:** e0160062.

24 Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016; **6:** e012799.

25 Theron G, Venter R, Smith L, et al. False-positive Xpert MTB/RIF results in retested patients with previous tuberculosis: frequency, profile, and prospective clinical outcomes. *J Clin Microbiol* 2018; **56:** e01696–17.

26 Xpert MTB/RIF Ultra(Package insert). Cepheid. Sunnyvale, California, USA. May, 2018. https://www.cepheid.com/uk/cepheid-solutions/clinical-ivd-tests/critical-infectious-diseases/xpert-mtb-rif-ultra (accessed Jan 15, 2020).

27 de Vos M, Derendinger B, Dolby T, et al. Diagnostic accuracy and utility of FluoroType MTBDR, a new molecular assay for multidrug-resistant tuberculosis. *J Clin Microbiol* 2018; **56:** e00531–18.

28 Simel DL, Feussner JR, DeLong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med Decis Making* 1987; **7:** 107–14.

29 Albert H, Nathavitharana RR, Isaacs C, Pai M, Denkinger CM, Boehme CC. Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J* 2016; **48:** 516–25.

30 Gumma V, DeGruy K, Bennett D, et al. Impact of external quality assurance on the quality of Xpert MTB/RIF testing in Viet Nam. *J Clin Microbiol* 2019; **57:** e01669–18.

31 McKinsey & Company. Independent evaluation of the Stop TB Partnership. Final report. April 21, 2008. http://www.stoptb.org/assets/documents/resources/publications/achieve_eval/Stop%20TB%20Evaluation%20Final%20Report_Exhibits_Annexes.pdf (accessed Jan 15, 2020).

32 Theron G, Peter J, Dowdy D, Langley I, Squire SB, Dheda K. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? *Lancet Infect Dis* 2014; **14:** 527–32.

33 Datta S, Saunders MJ, Tovar MA, Evans CA. Improving tuberculosis diagnosis: Better tests or better healthcare? *PLoS Med* 2017; **14:** e1002406.