



Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment

Sven O Friedrich*, Andrea Rachow*, Elmar Saathoff, Kasha Singh, Chacha D Mangu, Rodney Dawson, Patrick P J Phillips, Amour Venter, Anna Bateson, Catharina C Boehme, Norbert Heinrich, Robert D Hunt, Martin J Boeree, Alimuddin Zumla, Timothy D McHugh, Stephen H Gillespie, Andreas H Diacon, Michael Hoelscher, on behalf of the Pan African Consortium for the Evaluation of Anti-tuberculosis Antibiotics (PanACEA)

Summary

Background An accurate biomarker is urgently needed to monitor the response to treatment in patients with pulmonary tuberculosis. The Xpert MTB/RIF assay is a commercially available real-time PCR that can be used to detect *Mycobacterium-tuberculosis*-specific DNA sequences in sputum samples. We therefore evaluated this assay with serial sputum samples obtained over 26 weeks from patients undergoing treatment for tuberculosis.

Methods We analysed sputum samples from 221 patients with smear-positive tuberculosis enrolled at two sites (Cape Town, South Africa, and Mbeya, Tanzania) of a multicentre randomised clinical trial REMoxTB of antituberculosis treatment on a weekly basis (weeks 0 to 8), then at weeks 12, 17, 22, and 26 after treatment initiation. The Xpert MTB/RIF results over time were compared with the results of standard smear microscopy and culture methods.

Findings We obtained and analysed 2741 sputum samples from 221 patients. The reduction in positivity rates with Xpert MTB/RIF were slower than those with the standard methods. At week 8, positive results were obtained for 62 (29%) of 212 sputum samples with smear microscopy, 46 (26%) of 175 with solid culture (Löwenstein-Jensen medium), 77 (42%) of 183 with liquid culture (Bactec MGIT960 system), and 174 (84%) of 207 with Xpert MTB/RIF; at 26 weeks, positive results were obtained for ten (5%) of 199, four (3%) of 157, seven (4%) of 169, and 22 (27%) of 83 sputum samples, respectively. The reduction in detection of quantitative *M tuberculosis* DNA with Xpert MTB/RIF correlated with smear grades ($\rho=-0.74$; $p<0.0001$), solid culture grades ($\rho=-0.73$; $p<0.0001$), and time to liquid culture positivity ($\rho=0.73$; $p<0.0001$). Compared with the combined binary smear and culture results as a reference standard, the Xpert MTB/RIF assay had high sensitivity (97.0%, 95% CI 95.8–97.9), but poor specificity (48.6%, 45.0–52.2).

Interpretation The poor specificity precludes the use of the Xpert MTB/RIF assay as a biomarker for monitoring tuberculosis treatment, and should not replace standard smear microscopy and culture.

Funding Global Alliance for TB Drug Development, Bill & Melinda Gates Foundation, UK Medical Research Council, German Ministry of Science and Technology.

Introduction

Substantial progress has been made in global tuberculosis control but these gains are threatened by the global spread of drug-resistant tuberculosis.¹ The urgency of this situation calls for better treatment markers that can be used to detect insufficient treatment—the main cause of the emergence of resistance—early and easily. Additionally, the clinical development of the emerging new portfolio of tuberculosis drugs² could be accelerated if phase 2 and 3 trials could be greatly shortened and reliable surrogate markers used for the measurement of treatment success.³

Current biomarkers that have been used as predictors of non-relapsing cure are sputum-smear microscopy and sputum-culture conversion,⁴ both of which have shortcomings as markers for the measurement of treatment response. Sputum-smear microscopy is insensitive, operator dependent, and has been shown to be a poor predictor of treatment outcome.⁵ It cannot be used to differentiate between *Mycobacterium tuberculosis*

and non-tuberculosis mycobacteria reliably⁶ and between live and dead bacilli. Sputum culture takes a long time, is prone to contamination, is expensive, and is rarely available in high-burden settings. In the absence of an ideal marker to measure treatment success in individuals, WHO currently recommends smear microscopy at month 2 to identify individuals at risk of poor treatment outcome. Also, in recent clinical studies of bedaquiline (TMC207)^{7,8} and delamanid (OPC-67683),⁹ both biomarkers have been used to predict the therapeutic value of early drug regimens. Sputum-culture conversion at 8 weeks has some predictive value for the measurement of the sterilising activity of a novel regimen.¹⁰ Potentially more promising, but not yet sufficiently validated, is the time to culture positivity measured with semiautomated liquid culture.^{11–13}

Although several other *M tuberculosis* and human biomarkers have been studied over the past decade,^{14,15} including *M tuberculosis* DNA-based and RNA-based

Lancet Respir Med 2013;
1: 462–70

Published Online

July 1, 2013

[http://dx.doi.org/10.1016/S2213-2600\(13\)70119-X](http://dx.doi.org/10.1016/S2213-2600(13)70119-X)

See [Comment](#) page 427

*Both authors contributed equally

Division of Medical Physiology, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa (S O Friedrich PhD, Prof A H Diacon MD); National Institute of Medical Research–Mbeya Medical Research Centre, Mbeya, Tanzania (A Rachow MD, C D Mangu MD); Division of Infectious Diseases and Tropical Medicine, Medical Centre of the University of Munich (LMU), Munich, Germany (A Rachow, E Saathoff PhD, N Heinrich MD, Prof M Hoelscher FRCP); DZIF German Centre for Infection Research, Munich, Germany (A Rachow, Prof M Hoelscher); Centre for Clinical Microbiology, Division of Infection and Immunity, University College London, London, UK (K Singh MD, A Bateson PhD, R D Hunt PhD, Prof A Zumla FRCP, Prof T D McHugh PhD); Radboud University Nijmegen Medical Centre/UCCZ Dekkerswald, Nijmegen (M J Boeree MD); Division of Pulmonology, Department of Medicine, University of Cape Town Lung Institute, Groote Schuur Hospital, Cape Town, South Africa (Prof R Dawson MD); Medical Research Centre (MRC) Clinical Trials Unit, London, UK (P P J Phillips PhD); MRC Centre for Molecular and Cellular Biology, Department of Biomedical Sciences, Faculty

assays,^{16–18} no new specific and accurate biomarkers of disease activity and relapse have yet been validated. Therefore, an urgent need exists for biomarkers that can be used to accurately measure bacillary clearance and identify individuals receiving insufficient treatment with high risk of tuberculosis recurrence. Furthermore, for innovative trial designs in which the aim is to reduce the size and duration of clinical trials, the use of faster surrogate endpoints for treatment efficacy, disease activity, cure, and relapse will be indispensable.³

WHO approved the Xpert MTB/RIF assay (Cepheid, Sunnyvale, CA, USA) for sputum-based rapid diagnosis of pulmonary tuberculosis and multidrug-resistant (MDR) tuberculosis. The assay can be used to accurately measure the *M tuberculosis* load beyond the detection limit of 131 organisms per mL in an in-vitro suspension.^{19,20} The Xpert MTB/RIF assay in patients with suspected tuberculosis and newly diagnosed cases of tuberculosis has been evaluated in several studies.^{21–24} The results of these studies suggested that the Xpert MTB/RIF assay is better than sputum-smear microscopy; its quantitative readouts correlate well with results of conventional solid and liquid cultures.^{21–24} Specifically, for the detection of tuberculosis and MDR-tuberculosis in HIV-positive individuals, a cost benefit was noted compared with conventional smear microscopy.^{25,26} In this respect, detection of *M tuberculosis* DNA in serial sputum samples over time with the Xpert MTB/RIF assay could have the potential to replace conventional microbiological tests for monitoring response to tuberculosis treatment regimens. We evaluated the Xpert MTB/RIF assay in serial sputum samples obtained over 26 weeks of tuberculosis treatment from patients enrolled in the multicentre randomised clinical trial REMoxTB (ClinicalTrials.gov, NCT00864383).

Methods

Patients and study design

In this substudy, we obtained sputum samples from a subset of 221 participants enrolled in the multicentre, randomised clinical drug trial REMoxTB at two of the 47 clinical trial sites in Cape Town, South Africa, and Mbeya, Tanzania, between January, 2008, and April, 2010. In the REMoxTB trial, 1931 patients with previously untreated, drug-sensitive, smear-positive, pulmonary tuberculosis were enrolled between January, 2008, and January, 2012. HIV-positive patients with a CD4-cell count of fewer than 2.5×10^8 per L or those already on antiretroviral treatment were excluded to avoid concomitant antiretroviral treatment, according to current practice at the time the protocol was prepared. All participants were treated for tuberculosis either with one of two treatment-shortening regimens of 4 months, in which moxifloxacin was substituted for either isoniazid or ethambutol or with the standard 6 month regimen (2 months of isoniazid, rifampicin, ethambutol, and pyrazinamide followed by 4 months of isoniazid and rifampicin). Every patient was followed up for a total of

18 months. To avoid interference with the endpoints of the main treatment trial, the sample analysis with Xpert MTB/RIF was restricted until week 26 and all investigators were still masked to the treatment allocations. The results of the main study will become available in 2014.

Ethical and regulatory approval, granted by all local and national ethics committees and regulatory authorities, included the permission to undertake studies to improve treatment and diagnosis of tuberculosis.

Sputum samples

All participants were invited for regular visits to the study clinics and asked to provide one early morning sputum sample, which was spontaneously expectorated at home or at the clinics before treatment (week 0) and then weekly until week 8, followed by monthly sputum sample collection (weeks 12, 17, and 22) until week 26 after treatment initiation. Sputum specimens were processed by two study laboratories (Department of Biomedical Sciences, University of Stellenbosch, and the tuberculosis laboratory at the National Institute for Medical Research-Mbeya Medical Research Centre) in accordance with the standardised REMoxTB laboratory manual at both laboratories. Briefly, a maximum of 10 mL of sample was homogenised and decontaminated with the N-acetyl-L-cysteine-sodium hydroxide procedure.²⁷ After centrifugation for 15 min at 4°C and 3000×g, the resulting pellet was stored at 4°C for further analyses.

30 µL of the concentrated sputum pellet was used for Ziehl-Neelsen staining followed by smear microscopy; 1.5 mL phosphate-buffered saline (pH 6.8; Becton Dickinson, Sparks, MD, USA) was added and the pellet was resuspended. 500 µL of the suspension was used for evaluation in liquid culture with the Bactec MGIT960 system (MGIT; Becton Dickinson) and 150 µL in solid culture on Löwenstein-Jensen medium (Becton Dickinson) with standard protocols.

The remaining sputum pellet was mixed with 2 mL of the phosphate-buffered saline and vortexed. 1 mL was processed for evaluation in the Xpert MTB/RIF assay according to the manufacturer's instructions by use of the GeneXpert Dx software (version 2.1 and 4.0) for automation of all PCR processing, reaction, and detection steps. The signals for fluorescence were recorded,

of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa (A Venter NatDipMicro); Foundation of Innovative New Diagnostics, Geneva, Switzerland (C C Boehme MD); and University of St Andrews School of Medicine, North Haugh, St Andrews, Scotland, UK (Prof S H Gillespie FRCP)

Correspondence to: Prof Stephen H Gillespie, University of St Andrews School of Medicine, North Haugh St Andrews, Fife KY16 9TF, Scotland, UK shg3@st-andrews.ac.uk

	Both sites (n=221)	Mbeya (n=90)	Cape Town (n=131)	p value*
Age (years)	30 (23–40)	30 (25–39)	30 (23–41)	0.43
Sex, male	144 (65%)	57 (63%)	87 (66%)	0.64
HIV-positive status	20 (9%)	16 (18%)	4 (3%)	<0.0001
CD4-cell count $\times 10^8$ per L in participants who are HIV positive	3.62 (2.91–4.52)	3.59 (2.91–4.96)	3.92 (3.10–4.31)	0.56

Data are number (%) or median (IQR), unless otherwise indicated. *For difference between sites, from Wilcoxon rank-sum test (binary variables) or two-sample test of proportions (continuous variables).

Table 1: Demographic and clinical characteristics of study participants

analysed, and reported as semiquantitative readouts based on the minimal cycle threshold (C_{min}), which is defined as one of five probes specific for *M tuberculosis* complex becoming positive for detection.²

Statistical analysis

A negative result for smear microscopy was reported if no acid-fast bacilli were detected in at least 100 observation fields. Cultures in MGIT and on Löwenstein-Jensen media were judged to have a negative result if no mycobacterial growth was seen until 6 weeks and 8 weeks after incubation, respectively. A tuberculosis negative result for the Xpert MTB/RIF assay was generated automatically following an interpretative algorithm with the GeneXpert Dx software. Apart from calculating binary (positive, negative) variables for all test outcomes, results were processed to analyse them quantitatively: the intensity grades for smear (0, 1+, 2+, 3+, 4+)²⁸ and Löwenstein-Jensen (negative, <20 colonies, 1+, 2+, 3+) were coded into five categories from 0 (negative) to 4. Time to positivity of MGIT and C_{min} of the Xpert MTB/RIF assay were used as quantitative outcomes and both were coded as 50 days and 50 cycles, respectively, if negative. Diagnostic test performance (sensitivity, specificity, negative predictive value, positive predictive value) and Spearman rank correlations were calculated with standard formulae. In the absence of a perfect and accurate marker for tuberculosis disease activity, treatment response, or treatment outcome, either the combined result for smear microscopy, Löwenstein-Jensen culture, and MGIT culture, or each individual test was used as reference standard for the calculation of accuracy of the Xpert MTB/RIF assay. Differences between groups were tested for significance with the two-sample proportion Z test for binary variables and the non-parametric Wilcoxon rank sum test for continuous variables because none of the results conformed to a normal distribution. Longitudinal

C_{min} values over time were analysed by use of the non-linear least-squares estimation and non-linear mixed-effects models to better account for between-patient variability. The same biexponential model was used previously to fit to these type of data.^{18,29} Stata software (version 12.1) was used for all statistical analyses with the exception of the non-linear mixed-effects models, which were fit using the nlme package in R (version 2.9.0).

Role of the funding source

The study was a collaboration between the Global TB Alliance and the Pan African Consortium for the Evaluation of Anti-tuberculosis Antibiotics (PanACEA). This study was undertaken with permission from the REMox trial sponsor, but the study design, data collection, analysis, and interpretation, and writing of the report were done independently. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

2741 sputum specimens were obtained and analysed from 221 patients (131 in Cape Town, 90 in Mbeya) from weeks 0–26 after initiation of antituberculosis treatment. Table 1 shows the demographic characteristics of the patients.

Figure 1 shows the qualitative data and table 2 the quantitative data for all tests at baseline and follow-up visits. Valid smear results were available for 2732 samples and valid Xpert MTB/RIF assay results for 2000 samples. After exclusion of acid-fast-bacillus-negative, but contaminated, cultures and those with no speciation result, 2324 and 2383 results for Löwenstein-Jensen and MGIT cultures, respectively, were included in the qualitative analysis (figure 1). There were 262 (11%) of 2383 acid-fast-bacillus-positive MGIT cultures showing contamination; these were excluded from the calculation

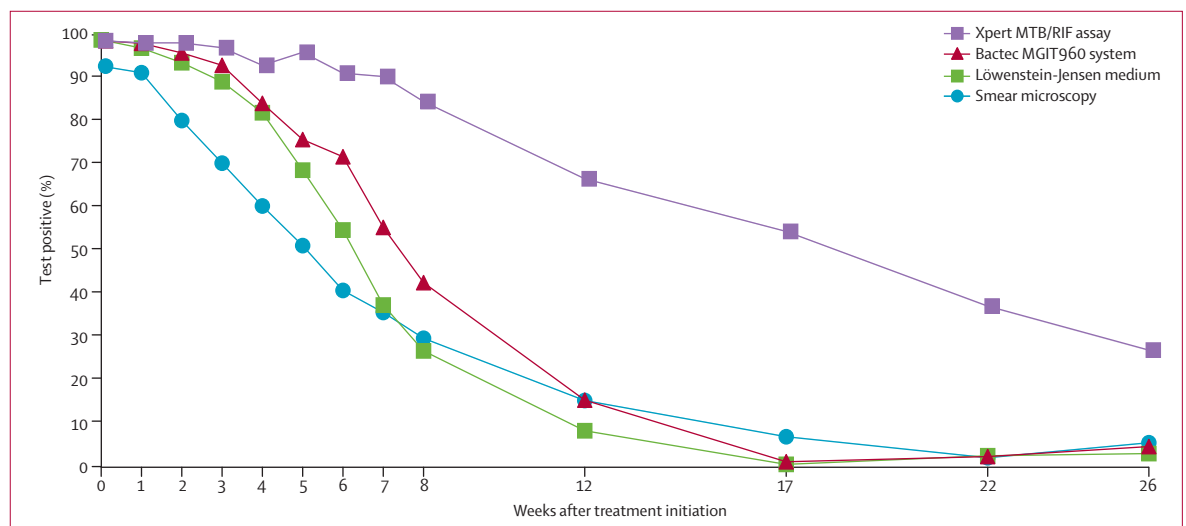


Figure 1: Qualitative data for all tests at baseline and follow-up visits

of the mean time to positivity (table 2). Because only patients with smear-positive tuberculosis were included in the trial after screening, 92% and 98% of baseline values for smear microscopy and all other diagnostic tests, respectively, were positive for *M tuberculosis* (figure 1; table 2). Similarly, at week 0, quantitative measurements were lowest for Xpert MTB/RIF C_{min} and MGIT time to positivity, and highest for smear and Löwenstein-Jensen grade (figure 1; table 2). The sputum load of *M tuberculosis* measured with smear and culture tests fell rapidly and consistently after induction of treatment (figure 1; table 2). A delayed, less rapid, and almost linear fall in the mycobacterial load was noted with the Xpert MTB/RIF assay (figure 1). For instance, 84% of the sputum samples were still positive with the Xpert MTB/RIF assay at 8 weeks after treatment initiation whereas 29% were smear positive (table 2). Furthermore, at the end of the observation period (week 26), 27% of the samples were positive with the Xpert MTB/RIF assay but only 3% and 4% with the Löwenstein-Jensen and MGIT cultures, respectively (table 2).

To further elaborate the findings, we used three different statistical models to calculate the increase in C_{min} with the Xpert MTB/RIF assay (figure 2). In model 1,

all negative results from the analysis were excluded. In model 2, a patient's first negative result was imputed with a C_{min} of 50 and the subsequent negative results were excluded. In model 3, all negative results were imputed with a C_{min} of 50. The raw means (excluding negative results) with 95% CIs and raw medians were plotted by visit week. Figure 2 shows the fitted line for each model with the raw means and medians for C_{min} . During the first 7 weeks, when there were few negative results with the Xpert MTB/RIF assay, the lines for the mean and median values were close, indicating that the distribution of C_{min} was symmetrical. Assuming that the underlying distribution of C_{min} remains symmetrical throughout the observation period, the median is judged to be a better summary measure than the mean for calculations with the symmetrically distributed data for C_{min} because no assumptions are needed for values greater than the limit of detection, as is the case for models 1–3. Therefore, although, the increase in mean C_{min} suggested a shallower slope after week 8, the median continued to increase almost linearly until week 17. The lines for all models had linear slopes with slightly flattened curves after week 12, depending on the inclusion of the negative results as determined by the model. Numerical values for the

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 12	Week 17	Week 22	Week 26
Smear microscopy, positive results	204/221 (92%)	198/218 (91%)	173/217 (80%)	150/215 (70%)	127/212 (60%)	105/207 (51%)	83/206 (40%)	74/210 (35%)	62/212 (29%)	31/210 (15%)	13/202 (6%)	3/203 (1%)	10/199 (5%)
Smear microscopy, grade	3.3 (3.1–3.4)	2.9 (2.7–3.1)	2.3 (2.1–2.5)	1.8 (1.6–2.0)	1.5 (1.3–1.7)	1.2 (1.0–1.3)	0.8 (0.6–1.0)	0.7 (0.5–0.8)	0.5 (0.4–0.6)	0.3 (0.2–0.4)	0.1 (0.1–0.2)	0.0 (0.0–0.1)	0.1 (0.0–0.2)
Löwenstein-Jensen medium, positive results	188/191 (98%)	195/202 (97%)	176/189 (93%)	172/194 (89%)	145/178 (81%)	122/179 (68%)	100/184 (54%)	65/176 (37%)	46/175 (26%)	14/179 (8%)	0/168 (0%)	3/152 (2%)	4/157 (3%)
Löwenstein-Jensen medium, grade	3.6 (3.5–3.7)	3.0 (2.8–3.1)	2.5 (2.3–2.7)	2.0 (1.8–2.2)	1.7 (1.5–1.8)	1.2 (1.1–1.4)	0.9 (0.8–1.1)	0.6 (0.4–0.7)	0.5 (0.3–0.6)	0.2 (0.1–0.3)	0.0 (0.0–0.0)	0.1 (0.0–0.1)	0.1 (0.0–0.2)
Bactec MGIT960 system, positive results	215/219 (98%)	205/210 (98%)	186/195 (95%)	173/187 (93%)	154/184 (84%)	128/170 (75%)	124/174 (71%)	95/173 (55%)	77/183 (42%)	27/181 (15%)	1/168 (<1%)	3/170 (2%)	7/169 (4%)
Bactec MGIT960 system, time to positivity (days)	6.0 (5.0–7.0)	10.2 (9.1–11.2)	13.5 (12.0–15.0)	16.0 (14.1–17.8)	21.3 (18.9–23.6)	24.9 (22.2–27.6)	29.2 (26.4–32.0)	35.0 (32.3–37.7)	39.0 (36.5–41.4)	46.0 (44.1–47.8)	49.7 (49.2–50.3)	49.2 (48.4–50.1)	48.5 (47.3–49.7)
Xpert MTB/RIF assay, positive results	215/219 (98%)	214/219 (98%)	85/87 (98%)	85/88 (97%)	196/212 (92%)	85/89 (96%)	78/86 (91%)	80/89 (90%)	174/207 (84%)	131/198 (66%)	107/198 (54%)	72/196 (37%)	22/83 (27%)
Xpert MTB/RIF assay, C_{min}	19.0 (18.1–19.8)	20.4 (19.6–21.3)	20.2 (19.0–21.4)	21.7 (20.3–23.1)	25.1 (23.9–26.3)	24.2 (22.7–25.7)	26.1 (24.3–28.0)	27.1 (25.2–29.0)	29.2 (27.9–30.6)	34.6 (32.9–36.2)	38.4 (36.9–40.0)	42.4 (40.9–43.8)	44.2 (42.1–46.4)

Binary data are n/N (%) and quantitative data are mean (95% CI).

Table 2: Quantitative data for all tests at baseline and follow-up visits

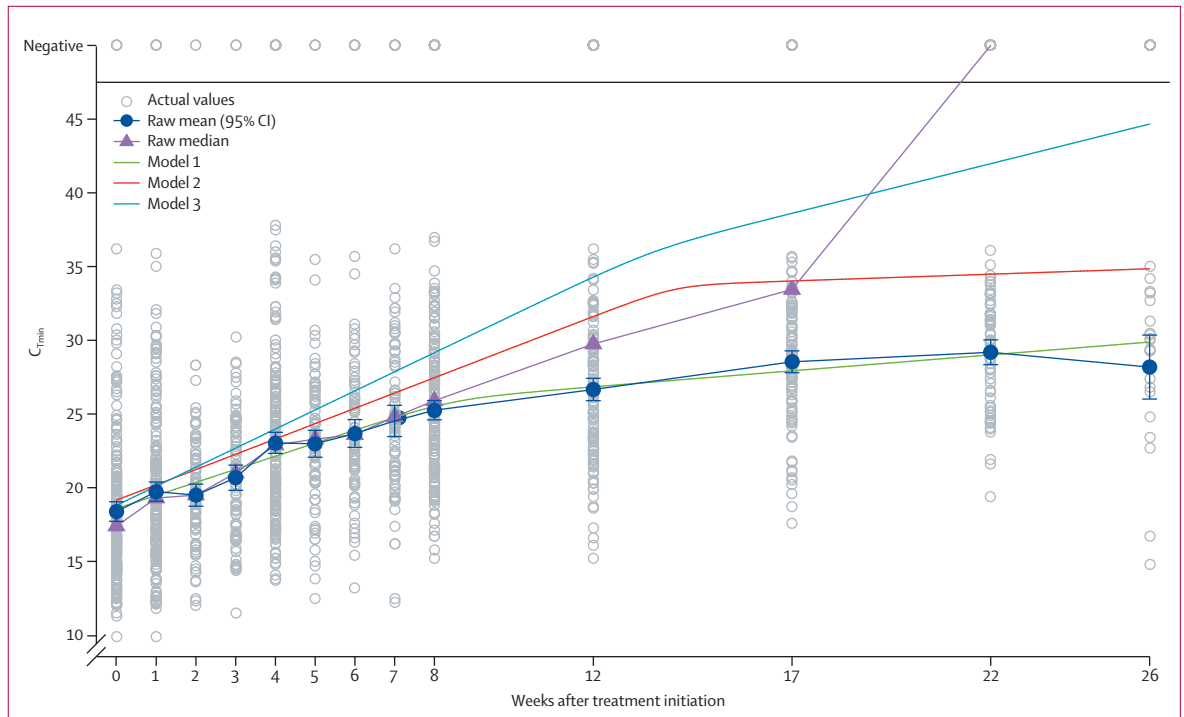


Figure 2: Increase in C_{min} during tuberculosis treatment fitted with three different statistical models
 Note break in y axis. The horizontal line is the positivity cutoff value. C_{min} =minimal cycle threshold.

	Number of comparisons	Sensitivity (true positives/false negatives; %, 95% CI)	Specificity (true negatives/false positives; %, 95% CI)	Positive predictive value (%; 95% CI)	Negative predictive value (%; 95% CI)
Week 0	219	214/3 (98.6, 96.0–99.7)	1/1 (50.0, 1.3–98.7)	99.5 (97.4–100.0)	25.0 (0.6–80.6)
Week 1	216	210/4 (98.1, 95.3–99.5)	1/1 (50.0, 1.3–98.7)	99.5 (97.4–100.0)	20.0 (0.5–71.6)
Week 2	85	83/1 (98.8, 93.5–100.0)	1/0 (100.0, 2.5–100.0)	100.0 (95.7–100.0)	50.0 (1.3–98.7)
Week 3	87	82/1 (98.8, 93.5–100.0)	2/2 (50.0, 6.8–93.2)	97.6 (91.7–99.7)	66.7 (9.4–99.2)
Week 4	206	178/3 (98.3, 95.2–99.7)	11/14 (44.0, 24.4–65.1)	92.7 (88.1–96.0)	78.6 (49.2–95.3)
Week 5	87	78/1 (98.7, 93.2–100.0)	3/5 (37.5, 8.5–75.5)	94.0 (86.5–98.0)	75.0 (19.4–99.4)
Week 6	86	65/2 (97.0, 89.6–99.6)	6/13 (31.6, 12.6–56.6)	83.3 (73.2–90.8)	75.0 (34.9–96.8)
Week 7	86	44/4 (91.7, 80.0–97.7)	5/33 (13.2, 4.4–28.1)	57.1 (45.4–68.4)	55.6 (21.2–86.3)
Week 8	200	101/6 (94.4, 88.2–97.9)	25/68 (26.9, 18.2–37.1)	59.8 (52.0–67.2)	80.6 (62.5–92.6)
Week 12	196	44/8 (84.6, 71.9–93.1)	57/87 (39.6, 31.5–48.1)	33.6 (25.6–42.4)	87.7 (77.2–94.5)
Week 17	191	12/1 (92.3, 64.0–99.8)	86/92 (48.3, 40.8–55.9)	11.5 (6.1–19.3)	98.9 (93.8–100.0)
Week 22	179	3/1 (75.0, 19.4–99.4)	115/60 (65.7, 58.2–72.7)	4.8 (1.0–13.3)	99.1 (95.3–100.0)
Week 26	80	3/0 (100.0, 29.2–100.0)	59/18 (76.6, 65.6–85.5)	14.3 (3.1–36.3)	100.0 (93.9–100.0)
Overall	1918	1117/35 (97.0, 95.8–97.9)	372/394 (48.6, 45.0–52.2)	73.9 (71.6–76.1)	91.4 (88.2–93.9)

Data for sensitivity, specificity, positive predictive value, and negative predictive value are calculated against the reference standard. Samples with a negative smear result and missing culture result (Löwenstein-Jensen medium and Bactec MGIT960 system) were excluded from the analysis, and samples with a positive smear result and missing culture result (Löwenstein-Jensen medium and Bactec MGIT960 system) were included in the analysis.

Table 3: Diagnostic performance of Xpert MTB/RIF assay compared with the combined reference standard (smear microscopy, Löwenstein-Jensen medium, and Bactec MGIT960 system)

medians at weeks 22 and 26 could not be calculated because more than 50% of patients had negative results in the Xpert MTB/RIF assay (figure 2).¹

To investigate the relation between C_{min} and the quantitative readouts for the smear, Löwenstein-Jensen, and MGIT assays, the Spearman rank correlation

coefficient was calculated for each study timepoint. The magnitudes of the overall correlation coefficients (ρ) over the entire 26 weeks were almost identical for smear (-0.74 ; $p < 0.0001$), Löwenstein-Jensen (-0.73 ; $p < 0.0001$), and MGIT (0.73 ; $p < 0.0001$). Correlations were variable at discrete timepoints. C_{min} values correlated moderately

with smear microscopy, Löwenstein-Jensen solid culture, and MGIT liquid culture in the first 8 weeks but showed poorer correlation thereafter.

We analysed the Xpert MTB/RIF assay for prediction of the binary smear and culture results at consecutive timepoints during treatment. Although the overall sensitivity of the assay was 97%, specificity was low compared with the reference standard, which combined smear microscopy, MGIT, and Löwenstein-Jensen results (table 3). At baseline and early follow-up visits, the putative poor specificity of Xpert MTB/RIF assay was misleading because it was caused by false-negative microbiological results with the reference methods and a very low number of patients. Of greater relevance was the low specificity at weeks 6–12, indicating a delayed Xpert MTB/RIF assay conversion compared with smear and culture methods (figure 1; table 3). The diagnostic performance of the Xpert MTB/RIF assay versus smear or culture methods alone did not show substantially different values (table 4). Notably, compared with MGIT, the Xpert MTB/RIF assay sensitivity and negative predictive value for culture positivity were both 100% from week 17 onwards (table 4; appendix).

We also investigated whether any change in C_{min} over time or absolute C_{min} could be used to predict a patient's conversion to a negative smear or culture result during treatment and did not find it to be predictive (data not shown).

Discussion

The important findings of our study are that quantitative Xpert MTB/RIF assay readouts for sputum *M tuberculosis* DNA as a biomarker correlate well with *M tuberculosis* smear and culture results during antituberculosis treatment. Positivity rates for sputum *M tuberculosis* DNA detection with Xpert MTB/RIF assay decline more slowly than do those with conventional sputum smear microscopy, solid culture, and liquid culture and seemed to be linear throughout treatment. Compared with a combined reference standard of smear and culture methods, the Xpert MTB/RIF assay had high sensitivity (97%) but poor specificity (49%; table 3). To our knowledge, this study is the first evaluation of the capacity of the Xpert MTB/RIF assay as an early sputum biomarker of response for monitoring tuberculosis treatment, and

See Online for appendix

	Smear microscopy				Löwenstein-Jensen medium				Bactec MGIT 960 system			
	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Week 0	98.5% (95.7–99.7)	5.9% (0.2–28.7)	92.6% (88.2–95.7)	25.0% (0.6–80.6)	98.4% (95.4–99.7)	33.3% (0.8–90.6)	98.9% (96.2–99.9)	25.0% (0.6–80.6)	98.6% (95.9–99.7)	25.0% (0.6–80.6)	98.6% (95.9–99.7)	25.0% (0.6–80.6)
Week 1	100.0% (98.1–100.0)	25.0% (8.7–49.1)	92.9% (88.6–96.0)	100.0% (47.8–100.0)	100.0% (98.1–100.0)	57.1% (18.4–90.1)	98.5% (95.6–99.7)	100.0% (39.8–100.0)	98.0% (95.0–99.5)	20.0% (0.5–71.6)	98.0% (95.0–99.5)	20.0% (0.5–71.6)
Week 2	98.5% (92.1–100.0)	5.3% (0.1–26.0)	78.8% (68.6–86.9)	50.0% (1.3–98.7)	98.7% (92.7–100.0)	33.3% (0.8–90.6)	97.3% (90.7–99.7)	50.0% (1.3–98.7)	98.7% (93.2–100.0)	25.0% (0.6–80.6)	96.3% (89.6–99.2)	50.0% (1.3–98.7)
Week 3	98.2% (90.5–100.0)	6.3% (0.8–20.8)	64.7% (53.6–74.8)	66.7% (9.4–99.2)	98.6% (92.4–100.0)	20.0% (2.5–55.6)	89.7% (80.8–95.5)	66.7% (9.4–99.2)	98.6% (92.6–100.0)	28.6% (3.7–71.0)	93.5% (85.5–97.9)	66.7% (9.4–99.2)
Week 4	99.2% (95.7–100.0)	18.1% (10.5–28.1)	64.8% (57.6–71.5)	93.8% (69.8–99.8)	98.6% (95.0–99.8)	30.3% (15.6–48.7)	85.9% (79.6–90.8)	83.3% (51.6–97.9)	98.7% (95.3–99.8)	30.0% (14.7–49.4)	87.7% (81.8–92.2)	81.8% (48.2–97.7)
Week 5	97.0% (84.2–99.9)	5.4% (1.1–14.9)	37.7% (27.4–48.8)	75.0% (19.4–99.4)	98.3% (91.1–100.0)	12.5% (2.7–32.4)	73.8% (62.7–83.0)	75.0% (19.4–99.4)	98.3% (90.9–100.0)	17.7% (3.8–43.4)	80.6% (69.5–88.9)	75.0% (19.4–99.4)
Week 6	100.0% (85.2–100.0)	12.7% (5.7–23.5)	29.5% (19.7–40.9)	100.0% (63.1–100.0)	95.7% (85.5–99.5)	17.7% (6.8–34.5)	61.6% (49.5–72.8)	75.0% (34.9–96.8)	98.2% (90.5–100.0)	23.8% (8.2–47.2)	77.5% (66.0–86.5)	83.3% (35.9–99.6)
Week 7	95.8% (78.9–99.9)	12.3% (5.5–22.8)	28.8% (19.2–40.0)	88.9% (51.8–99.7)	96.0% (79.7–99.9)	14.0% (5.8–26.7)	35.8% (24.5–48.5)	87.5% (47.4–99.7)	90.9% (78.3–97.5)	12.9% (3.6–29.8)	59.7% (47.0–71.5)	50.0% (15.7–84.3)
Week 8	98.4% (91.2–100.0)	21.4% (15.0–29.0)	34.5% (27.5–42.1)	96.9% (83.8–99.9)	95.6% (84.9–99.5)	16.8% (10.7–24.5)	29.3% (22.1–37.3)	91.3% (72.0–98.9)	94.7% (86.9–98.5)	22.6% (14.9–31.9)	47.3% (39.1–55.6)	85.2% (66.3–95.8)
Week 12	100.0% (88.1–100.0)	39.6% (32.2–47.4)	22.1% (15.4–30.2)	100.0% (94.6–100.0)	71.4% (41.9–91.6)	35.9% (28.4–44.0)	9.1% (4.5–16.1)	93.3% (83.8–98.2)	76.9% (56.4–91.0)	34.5% (26.9–42.7)	17.1% (10.8–25.2)	89.5% (78.5–96.0)
Week 17	91.7% (61.5–99.8)	48.9% (41.5–56.4)	10.5% (5.4–18.0)	98.9% (94.0–100.0)	100.0% (2.5–100.0)	45.1% (37.4–53.1)	1.1% (0.0–6.0)	100.0% (95.1–100.0)
Week 22	66.7% (9.4–99.2)	63.7% (56.5–70.5)	2.8% (0.3–9.7)	99.2% (95.6–100.0)	100.0% (29.2–100.0)	68.1% (59.8–75.6)	6.1% (1.3–16.9)	100.0% (96.3–100.0)	100.0% (29.2–100.0)	64.8% (56.9–72.1)	5.0% (1.0–13.9)	100.0% (96.6–100.0)
Week 26	100.0% (15.8–100.0)	75.3% (64.5–84.2)	9.1% (1.1–29.2)	100.0% (94.1–100.0)	100.0% (15.8–100.0)	77.1% (65.6–86.3)	11.1% (1.4–34.7)	100.0% (93.4–100.0)	100.0% (15.8–100.0)	76.8% (65.1–86.1)	11.1% (1.4–34.7)	100.0% (93.3–100.0)
Overall	98.7% (97.7–99.3)	36.8% (34.0–39.7)	53.7% (51.1–56.2)	97.4% (95.4–98.7)	98.0% (96.9–98.9)	40.1% (37.6–44.4)	63.5% (60.8–66.1)	95.2% (92.4–97.2)	97.3% (96.0–98.2)	43.5% (39.9–47.1)	68.9% (66.4–71.4)	92.5% (89.3–95.0)

Data are percentage (95% CI). Data for sensitivity, specificity, positive predictive value, and negative predictive value were calculated against individual reference standards

Table 4: Diagnostic performance of Xpert MTB/RIF assay compared with smear microscopy, Löwenstein-Jensen medium, and Bactec MGIT960 system

as a alternative for standard smear microscopy and culture (panel).

The Xpert MTB/RIF assay is assumed to be specific; it only detects DNA from intact *M tuberculosis* bacilli and contamination from free DNA is thought to be removed by a washing step.^{20,30} The assay, however, does not differentiate between viable, dormant,³¹ and non-viable intact *M tuberculosis* bacilli that are shed during effective antituberculosis treatment. Furthermore, our findings—high rates of positive results at the end of the 6 month treatment—suggest that even DNA fragments from lysed or damaged bacteria could have been detected by use of Xpert MTB/RIF. Detection of these bacteria would explain the lengthy and constant slope, indicating a linear decay of DNA and dead mycobacteria, and is in contrast to the biphasic curves reported for culture or RNA-based assays.^{12,18} By contrast with DNA, the presence and amount of *M tuberculosis* RNA, particularly mRNA, correlated well with culture positivity and viable colony counts, respectively,^{17,18,32} whereas DNA-based assays were also positive beyond culture conversion or even the end of successful treatment.^{16,33} The results of a recent study suggest that viable and non-viable *M tuberculosis* bacilli can be distinguished by use of propidiummonoazide, which enters damaged or non-viable mycobacteria, covalently binds to DNA after light-exposure, and prevents DNA amplification, such as in the Xpert MTB/RIF assay.³⁴

The data show that detection of *M tuberculosis* DNA in sputum with the Xpert MTB/RIF assay in its current format cannot be used as a biomarker of disease activity

and cannot replace conventional smear culture for the monitoring of patients undergoing treatment for tuberculosis. This is in agreement with the results of a recently reported study in which patients taking part in an early bactericidal activity trial were monitored by use of the Xpert MTB/RIF assay, and liquid and solid cultures; the Xpert MTB/RIF assay was the least suitable for the assessment of drug efficacy.⁸ Hence, several aspects of the capacity of the assay for use in monitoring tuberculosis treatment require further study: whether the specificity of the Xpert MTB/RIF assay can be improved without affecting sensitivity through the prevention of DNA amplification from dead bacteria; whether a positive or negative Xpert MTB/RIF assay result at the end of 6 months of tuberculosis treatment can be used to provide prognostic information about treatment failure and relapse; and the prognostic relevance of baseline quantitative values or the timepoint of conversion to negativity, which was shown for time to positivity with MGIT.¹³

This study has several advantages because it was nested in a clinical trial. A large number of samples (n=2741) were obtained and processed under standardised and highly monitored conditions, as shown by the robust and high-quality data generated at both study sites for the Xpert MTB/RIF assay, and smear and culture methods. The similarity of the data between different sites suggests that our findings could not be attributed to technical factors. The number of patients lost to follow-up was very low and the falling numbers of available results during follow-up are usually the result of patients being unable to produce sputum samples at the end of an effective treatment.

A limitation of this study is that the results of the Xpert MTB/RIF assay could not be linked with clinical outcomes such as cure and relapse. Since the REMoxTB trial is still in progress, information about the allocated tuberculosis treatment or the presence of drug resistance in relation to the patients' responses to treatment was not available for data analysis in this study. However, because the Xpert MTB/RIF assay is widely available in many countries, clinicians might be starting to use this assay speculatively to monitor patients, as is commonly done with microscopy. Therefore, the results of our study are important because they show that the Xpert MTB/RIF assay is less responsive for the monitoring of tuberculosis treatment than is smear microscopy.

Notably, however, the absence of an adequate reference standard, and the consecutive use of smear microscopy and culture as imperfect surrogates, might have distorted the results of the diagnostic performance of the Xpert MTB/RIF assay. The results are further limited by our study being undertaken in only two sites in Africa and in a population of patients who were not severely ill.

Our data indicate that the Xpert MTB/RIF assay in its current format is not suitable for the monitoring of

Panel: Research in context

Systematic review

We did an online search of the PubMed database on Jan 31, 2011, to identify scientific articles about the Xpert MTB/RIF assay published since Jan 1, 2010, using the terms "Xpert" or "GeneXpert" and "tuberculosis". Only articles published in English were selected. Six articles were reviewed. The results of these studies showed the ability of this testing platform as a primary diagnostic method for detection of *Mycobacterium tuberculosis* mainly in expectorated sputum samples from patients with untreated tuberculosis.²⁰⁻²⁴ Although several studies have been planned, there are no reports of the use of the Xpert MTB/RIF assay to monitor tuberculosis treatment in the context of a clinical trial.

Interpretation

Previous research has shown that the Xpert MTB/RIF assay has a high specificity in the diagnosis of *M tuberculosis* with improved sensitivity compared with smear microscopy and culture. Thus, the available evidence shows that the Xpert MTB/RIF assay is useful as a primary diagnostic aid. In view of the need for improved biomarkers, researchers have speculated that the assay might be used to monitor treatment response and the samples available to us provided a unique opportunity to test this hypothesis. As a result, the Xpert MTB/RIF assay was compared with the standard method to assess its usefulness for treatment monitoring and its potential as a biomarker. We concluded that, in its current format, the Xpert MTB/RIF assay cannot be used to monitor antituberculosis treatment due to its slow decline in positivity compared with smear microscopy and culture. Future technological change that would eliminate the signal from dead bacilli could overcome this problem and provide a useful biomarker for use in resource-poor settings.

tuberculosis treatment and, thus, cannot replace standard smear microscopy and culture. Therefore, technical improvements in the Xpert MTB/RIF assay that prevent the assay from detecting DNA from non-viable or even damaged bacteria will be needed before the test is suitable for monitoring patients in settings where it has already replaced smear microscopy as a diagnostic test. Larger studies with longer follow-up are needed to clarify the prognostic relevance of the Xpert MTB/RIF assay results for the prediction of treatment failure or relapse.

Contributors

All authors are part of PanACEA and the REMoxTB trial and contributed to the Article. TDM, SHG, AHD, and MH designed the study and protocols. Data were gathered by SOF, AR, AV, CDM, RD, CCB, and NH. SOF, AR, ES, PPJP, RDH, AHD, and MH undertook the management and the analyses of the data. SOF, AR, MH, AHD, SHG, TDM, MJB, and AZ wrote the first and final drafts. KS and AB contributed to the work through management and analysis of the data. All authors participated in writing the Article.

PanACEA

Medical Centre of the University of Munich, Munich, Germany (Sonja Henne, Anna Maria Mekota, Norbert Heinrich, Andrea Rachow, Anke Kohlenberg, Elmar Saathoff, Michael Hoelscher); University of St Andrews, St Andrews, UK (Stephen Gillespie); Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands (Georgette Plemper van Balen, Marloes Weijers, Rob Aarnoutse, Martin J Boeree); University College of London, London, UK (Anna Bateson, Timothy McHugh, Kasha Singh, Robert Hunt, Alimuddin Zumla); Medical Research Council Clinical Trials Unit, London, UK (Andrew Nunn, Patrick Phillips); University of Cape Town, Cape Town, South Africa (Rodney Dawson, Kim Narunsky); University of Stellenbosch, Cape Town, South Africa (Andreas Diacon, Jeannine du Bois, Amour Venter, Sven Friedrich); University of the Witwatersrand, Johannesburg, South Africa (Ian Sanne, Karla Mellet, Eefje de Jong); Aurum Institute, Johannesburg, South Africa (Gavin Churchyard, Salome Charalambous); University of Zambia, Lusaka, Zambia (Peter Mwaba); National Institute of Medical Research-Mbeya Medical Research Centre, Mbeya, Tanzania (Nyanda Elias, Chacha Mangu, Gabriel Rojas-Ponce, Bariki Mtafya, Leonard Maboko); Ifakara Health Institute-Bagamoyo Research and Training Centre, Bagamoyo, Tanzania (Klaus Reither, Levan Jugheli); Kilimanjaro Clinical Research Institute, Moshi, Tanzania (Noel Sam, Gibson Kibiki, Hadija Semvua, Stella Mpagama); Medical Research Unit—Albert Schweitzer Hospital, Lambarene, Gabon (Abraham Alabi, Ayola Akim Adegnikia); Kenya Medical Research Institute, Nairobi, Kenya (Evans Amukoye); Makerere University, Kampala, Uganda (Alphonse Okwera).

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

PanACEA is funded by the European and Developing Countries Clinical Trials Partnership (grants numbers CT.2004.32011.001, IP.2007.32011.011, IP.2007.32011.012, IP.2007.32011.013), Global Alliance for TB Drug Development, Bill & Melinda Gates Foundation, UK Medical Research Council, and German Ministry of Science and Technology (grant number 01KA0901). We thank all patients who took part in this study.

References

- Raviglione M, Marais B, Floyd K, et al. Scaling up interventions to achieve global tuberculosis control: progress and new developments. *Lancet* 2012; **379**: 1902–13.
- Zumla A, Hafner R, Lienhardt C, Hoelscher M, Nunn A. Advancing the development of tuberculosis therapy. *Nat Rev Drug Discov* 2012; **11**: 171–72.
- Phillips PP, Gillespie SH, Boeree M, et al. Innovative trial designs are practical solutions for improving the treatment of tuberculosis. *J Infect Dis* 2012; **205** (suppl 2): S250–57.
- Wallis RS, Perkins MD, Phillips M, et al. Predicting the outcome of therapy for pulmonary tuberculosis. *Am J Respir Crit Care Med* 2000; **161** (4 pt 1): 1076–80.
- Yeager H LJ, Smith LR, LeMaistre CA. Quantitative studies of mycobacterial populations in sputum and saliva. *Am Rev Respir Dis* 1967; **95**: 998–1004.
- Wright PW, Wallace RJ Jr, Wright NW, Brown BA, Griffith DE. Sensitivity of fluorochrome microscopy for detection of *Mycobacterium tuberculosis* versus nontuberculous mycobacteria. *J Clin Microbiol* 1998; **36**: 1046–49.
- Diacon AH, Dawson R, Von Groote-Bidlingmaier F, et al. Randomized dose-ranging study of the 14-day early bactericidal activity of bedaquiline (TMC207) in patients with sputum microscopy smear-positive pulmonary tuberculosis. *Antimicrob Agents Chemother* 2013; **57**: 2199–203.
- Kayigire XA, Friedrich SO, Venter A, et al. Direct comparison of Xpert MTB/RIF assay with liquid and solid mycobacterial culture for quantification of early bactericidal activity. *J Clin Microbiol* 2013; **51**: 1894–98.
- Gler MT, Skripconoka V, Sanchez-Garavito E, et al. Delamanid for multidrug-resistant pulmonary tuberculosis. *N Engl J Med* 2012; **366**: 2151–60.
- Mitchison DA. Assessment of new sterilizing drugs for treating pulmonary tuberculosis by culture at 2 months. *Am Rev Respir Dis* 1993; **147**: 1062–63.
- Pheiffer C, Carroll NM, Beyers N, et al. Time to detection of *Mycobacterium tuberculosis* in BACTEC systems as a viable alternative to colony counting. *Int J Tuberc Lung Dis* 2008; **12**: 792–98.
- Diacon AH, Maritz JS, Venter A, van Helden PD, Dawson R, Donald PR. Time to liquid culture positivity can substitute for colony counting on agar plates in early bactericidal activity studies of antituberculosis agents. *Clin Microbiol Infect* 2012; **18**: 711–17.
- Hesseling AC, Walzl G, Enarson DA, et al. Baseline sputum time to detection predicts month two culture conversion and relapse in non-HIV-infected patients. *Int J Tuberc Lung Dis* 2010; **14**: 560–70.
- Wallis RS, Pai M, Menzies D, et al. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. *Lancet* 2010; **375**: 1920–37.
- Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A. Immunological biomarkers of tuberculosis. *Nat Rev Immunol* 2011; **11**: 343–54.
- Kennedy N, Gillespie SH, Saruni AO, et al. Polymerase chain reaction for assessing treatment response in patients with pulmonary tuberculosis. *J Infect Dis* 1994; **170**: 713–16.
- Desjardin LE, Perkins MD, Wolski K, et al. Measurement of sputum *Mycobacterium tuberculosis* messenger RNA as a surrogate for response to chemotherapy. *Am J Respir Crit Care Med* 1999; **160**: 203–10.
- Honeyborne I, McHugh TD, Phillips PP, et al. Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum *Mycobacterium tuberculosis* bacillary load during treatment. *J Clin Microbiol* 2011; **49**: 3905–11.
- van Zyl-Smit RN, Binder A, Meldau R, et al. Comparison of quantitative techniques including Xpert MTB/RIF to evaluate mycobacterial burden. *PLoS One* 2011; **6**: e28815.
- Helb D, Jones M, Story E, et al. Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. *J Clin Microbiol* 2010; **48**: 229–37.
- Theron G, Peter J, van Zyl-Smit R, et al. Evaluation of the Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in a high HIV prevalence setting. *Am J Respir Crit Care Med* 2011; **184**: 132–40.
- Blakemore R, Nabeta P, Davidow AL, et al. A multi-site assessment of the quantitative capabilities of the Xpert(R) MTB/RIF assay. *Am J Respir Crit Care Med* 2011; **184**: 1076–84.
- Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010; **363**: 1005–15.
- Rachow A, Zumla A, Heinrich N, et al. Rapid and accurate detection of *Mycobacterium tuberculosis* in sputum samples by Cepheid Xpert MTB/RIF assay—a clinical validation study. *PLoS One* 2011; **6**: e20458.
- Weyer K, Mirzayev F, Migliori G, et al. Rapid molecular TB diagnosis: evidence, policy-making and global implementation of Xpert(R)MTB/RIF. *Eur Respir J* 2012; published online Nov 22. DOI:10.1183/09031936.00157212.

- 26 Pantoja A, Fitzpatrick C, Vassall A, Weyer K, Floyd K. Xpert MTB/RIF for diagnosis of TB and drug-resistant TB: a cost and affordability analysis. *Eur Respir J* 2012; published online Dec 20. DOI:10.1183/09031936.00147912.
- 27 Kent PT, Kubica GP. Public health mycobacteriology: a guide for the level III laboratory. Atlanta, GA: US Department of Public Health and Human Services, 1985.
- 28 American Thoracic Society. Diagnostic standards and classification of tuberculosis in adults and children. *Am J Respir Crit Care Med* 2000; 4: 1376–95.
- 29 Rustomjee R, Lienhardt C, Kanyok T, et al. A phase II study of the sterilising activities of ofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2008; 12: 128–38.
- 30 Blakemore R, Story E, Helb D, et al. Evaluation of the analytical performance of the Xpert MTB/RIF assay. *J Clin Microbiol* 2010; 48: 2495–501.
- 31 Garton NJ, Waddell SJ, Sherratt AL, et al. Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. *PLoS Med* 2008; 5: e75.
- 32 Li L, Mahan CS, Palaci M, et al. Sputum *Mycobacterium tuberculosis* mRNA as a marker of bacteriologic clearance in response to antituberculosis therapy. *J Clin Microbiol* 2010; 48: 46–51.
- 33 Hellyer TJ, Fletcher TW, Bates JH, et al. Strand displacement amplification and the polymerase chain reaction for monitoring response to treatment in patients with pulmonary tuberculosis. *J Infect Dis* 1996; 173: 934–41.
- 34 Miotto P, Bigoni S, Migliori GB, Matteelli A, Cirillo DM. Early tuberculosis treatment monitoring by Xpert(R) MTB/RIF. *Eur Respir J* 2012; 39: 1269–71.