EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Diagnostic accuracy of centralised assays for TB detection and detection of resistance to rifampicin and isoniazid: A systematic review and meta-analysis

Mikashmi Kohli, Emily MacLean, Madhukar Pai, Samuel G. Schumacher, Claudia M. Denkinger

Please cite this article as: Kohli M, MacLean E, Pai M, *et al*. Diagnostic accuracy of centralised assays for TB detection and detection of resistance to rifampicin and isoniazid: A systematic review and meta-analysis. *Eur Respir J* 2020; in press (https://doi.org/10.1183/13993003.00747-2020).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

Diagnostic accuracy of centralized assays for TB detection and detection of resistance to rifampicin and isoniazid: A systematic review and meta-analysis

Mikashmi Kohli[1,2]; Emily MacLean[1,2]; Madhukar Pai[1,2]; Samuel G. Schumacher[3]; Claudia M. Denkinger[3,4]

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada;
2 McGill International TB Centre, McGill University, Montreal, Canada;
3 Foundation for Innovative New Diagnostics, Geneva, Switzerland;
4 Centre for Infectious Diseases, University Heidelberg, Heidelberg, Germany

Corresponding author:
Samuel G. Schumacher, MSc. PhD
Deputy Head of TB
FIND, Geneva
Email: samuel.schumacher@finddx.org

Abstract

Background

Various diagnostic companies have developed high throughput molecular assays for TB and resistance detection for rifampicin and isoniazid. We performed a systematic review and meta-analyses to assess the diagnostic accuracy of five of these tests for pulmonary specimens. The tests included were Abbott RealTime MTB, Abbott RealTime RIF/INH, FluoroType MTB, FluoroType MTDBR and BD Max MDR-TB assay.

Methods

A comprehensive search of six databases for relevant citations was performed. Cross-sectional, case-control, cohort studies, and randomized controlled trials of any of the index tests were included. Respiratory specimens (such as sputum, bronchoalveolar lavage, tracheal aspirate, etc.) or their culture isolates.

Results

A total of 21 included studies contributed 26 datasets. We could only meta-analyse data for three of the five assays identified, as data were limited for the remaining two. For TB detection, the included assays had a sensitivity of 91% or more and the specificity ranged from 97%- 100%. For rifampicin resistance detection, all the included assays had a sensitivity of more than 92%, with a specificity of 99-100%. Sensitivity for isoniazid resistance detection varied from 70-91%, with higher specificity of 99-100% across all index tests. Studies that included head-to-head comparisons of these assays with Xpert MTB/RIF for detection of TB and rifampicin resistance suggested comparable diagnostic accuracy.

Conclusion

In people with symptoms of pulmonary TB, the centralized molecular assays demonstrate comparable diagnostic accuracy for detection of TB, rifampicin and isoniazid resistance to Xpert MTB/RIF assay, a WHO recommended molecular test.

Introduction:

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* complex (MTBC), has surpassed HIV/AIDS as the world's leading infectious cause of death. WHO estimates that in 2018, 10 million people became ill with TB, and approximately 1.45 million died of the disease. In 2018, only half of all confirmed TB patients underwent drug susceptibility testing[1].

The introduction and rollout of nucleic acid amplification tests (NAATs) has significantly improved the area of TB diagnosis by providing rapid TB and drug resistance detection (WHO 2010). The principal behind these assays is amplification of a targeted region of the *Mtb* genome by polymerase chain reaction (PCR). NAATs are used for both TB detection (particularly the Xpert MTB/RIF) and identification of mutations that confer resistance to anti-TB drugs (for example, Bruker-Hain and Nipro line probe assays (LPAs), most commonly rifampicin (RIF) and isoniazid (INH)[2,3]. Globally, INH mono-resistant TB is more prevalent than multidrug-resistant TB (MDR-TB), and WHO guidelines advocate for universal testing for both RIF and INH resistance before commencing TB treatment[4].

Recently, several companies have developed molecular tests for TB and RIF/INH resistance detection on centralized platforms, many of which have already been established as multi-disease platforms, primarily for detection of human immunodeficiency virus (HIV), human papillomavirus, and hepatitis C virus.

This systematic review intended to evaluate the diagnostic accuracy of five of these tests for *Mtb* and RIF/INH resistance detection to assess their diagnostic accuracy. The tests included were Abbott RealTime MTB, Abbott RealTime RIF/INH, FluoroType MTB, FluoroType MTDBR and BD Max MDR-TB assay.

Methods:

Search strategy, information sources, and eligibility criteria

We followed standard guidelines and methods for systematic review and meta-analyses of diagnostic test accuracy [5,6]. A comprehensive search of the following databases (PubMed, EMBASE, BIOSIS, Web of Science, LILACS, Cochrane) for relevant citations, without language restrictions was performed. An example search strategy is provided in Supplementary Methods. The time period was restricted to January 2009 to June 2018 and another scoping search was done till May 2020 to look for published studies for these platforms. We also contacted the developers of these tests to provide available data and lists of studies they are aware of. Cross-sectional, case-control, cohort studies, and randomized controlled trials of any of the index tests (listed above) were included if at least 25 specimens were tested. Abstracts and unpublished studies were excluded. Patients of all age groups with presumed or confirmed pulmonary TB or MDR-TB, in all settings and any country, were included.
Our search strategy also included terms for two assays by Roche and Bioneer that are comparable to the assays reviewed, however, we did not find any studies for these assays.

Citation screening and study selection

Two authors (MK, EM) independently screened and reviewed the full texts. Any discrepancies were resolved by discussion, and in case of disagreement, a third author was consulted (CMD). If a study contributed data to more than one analysis (e.g. two different index tests in one study), it was considered as two or more datasets. Disagreements in extracted information were resolved by discussion with third author (CMD). Study authors were contacted in cases of missing data. In cases of papers without extractable diagnostic accuracy data, the study was excluded if after three attempts the study author did not reply.

Reference standards

For TB detection, solid or liquid culture was the reference standard. For resistance detection, phenotypic drug susceptibility testing (DST) was the primary reference standard. However, if the studies provided information on sequencing, we analyzed the data using a phenotypic DST reference standard, a sequencing reference standard, and a composite reference standard (CRS). For a CRS, if phenotypic DST showed drug sensitivity but sequencing identified mutations recognized to be associated with resistance, the CRS was considered resistant when the mutations were associated with high or moderate confidence of resistance as per Miotto et al.[7] If phenotypic DST showed resistance but sequencing did not identify mutations associated with resistance, the CRS was considered resistant (as mutations could be outside of the region sequenced).

Head-to- head comparisons

When possible, the index tests were also compared to other well-characterised, WHO-recommended molecular test: Xpert MTB/RIF for both TB detection and rifampicin resistance. Such head-to-head comparisons are preferred, as using a WHO-recommended comparator test with known diagnostic accuracy serves as an easily understood benchmark for the index test's performance [8]. It can allow flagging of studies with particularly strong or weak results for the index test, which may help explain some between-study heterogeneity.

Assessment of methodological quality

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool, a validated quality assessment tool for diagnostic studies[9], was used to assess the included studies' risk of bias.

Statistical analysis and data synthesis

For each index test, meta-analyses were performed of sensitivity and specificity of TB detection, as well as RIF resistance and INH resistance when at least 4 studies were available. Studies were pooled using bivariate random effect hierarchical models to calculate sensitivity and specificity, with associated 95% confidence intervals, of each index test against the relevant reference standard. When there were fewer than 4 studies for an index test or evident heterogeneity between studies, a descriptive analysis only was performed.

Results:

From the literature search, 750 citations were identified, 81 full-text articles were reviewed, and 21 studies were included in the systematic review (see Figure 1). The 21 studies contributed 26 datasets, as four provided data for more than one index test. All studies were conducted in central level laboratories, which was expected as these assays require sophisticated laboratory infrastructure and skilled laboratory workers. As most studies were laboratory-based, there was limited demographics data available, such as age, HIV status, and past TB history of the included patient population. Tables 1a and 1b show the results of all the index tests analysed separately for both TB detection and resistance detection. Table 2 provides data for head-to-head comparisons of the index tests with Xpert MTB/RIF.

Risk of bias by QUADAS-2 assessment

The overall methodological quality of the included studies for each index test is summarised in Supplementary Figures S1- S10. For all assays except BD Max MDR-TB, the studies had applicability concerns in the domain of participant selection, as the studies were not conducted in high TB or MDR-TB burden settings. Similarly, for risk of bias, some studies had concerns in the patient selection domain and also the reference standard domains. In all other domains, risk of bias was low.

Abbott RealTime MTB:

Ten studies with 4858 respiratory specimens were included in the meta-analysis that evaluated Abbott RealTi*m*e MTB assay for TB detection [10-19] (Figure 2). In all studies, the assay was run directly on specimens, as opposed to positive culture isolates. Most studies (6/10) used fresh specimens, while four used frozen specimens. The median sample size was 389 (interquartile range [IQR]: 242 to 599). In individual studies, the sensitivity point estimates of Abbott MTB assay varied from 79% to 100% with specificity varied from 84% to 99% (Figure 2). Pooled sensitivity and specificity were 96.2% (95% Confidence Interval (CI): 90.2- 98.6) and 97.1% (95%CI: 93.7–98.7), respectively.

<u>Comparator test for TB detection: Xpert MTB/RIF</u>

In addition to the RealTime MTB assay, three studies [10,14,19] performed Xpert MTB/RIF on the same specimens [10,19] or on different specimens obtained from the patient on the same visit [14] (Figure 3a). In the study by Wang et al. [19], a lower overall specificity was observed for both Xpert (90%) and RealTime MTB (84%) than would be expected. In contrast, Scott et al, [14] showed Xpert specificity of 98%, and specificity for RealTime MTB was 92% in the study. Berhanu et al[10] also evaluated Xpert Ultra on same specimens. The study showed an increased Xpert Ultra sensitivity of 89%, but with a trade-off for lower specificity of 96% (Figure 3b).

Sub-group analyses: smear status

All ten studies provided data allowing stratification by smear status. For smear-positive specimens, the sensitivity of RealTime MTB assay varied from 95% to 100%. Pooled sensitivity was 99.0% (95%CI: 97.7-100) (10 studies, 765 specimens).

For smear-negative specimens, the sensitivity in these specimens varied from 41% to 100%. Pooled sensitivity was 88.4% (95%CI: 74.0-99.3) (10 studies, 4056 specimens). The study by Berhanu et al.[10] demonstrated very low sensitivity of 41.0% (95%CI: 18.0-67.0), which may partially be explained by the high prevalence of HIV in their study population, meaning that a high proportion of cases suffered from paucibacillary disease. We were not able to explore test performance by HIV status further as most of the studies (60%) did not report HIV prevalence.

Abbott RealTime MTB RIF/INH:

Seven studies provided data for RIF and INH resistance detection by RealTime MTB RIF/INH, with phenotypic DST as the reference standard in both use cases [10, 13-15, 20-22]. Six studies performed the index test directly on known TB positive specimens or as an accompanying drug susceptibility test with RealTime MTB. One study [20] used TB positive culture isolates for the index test specimen. Four studies used fresh specimens while others used bio-banked specimens.

RIF resistance detection

The pooled sensitivity and specificity for RIF resistance were 94% (95%CI: 89.0-99.0) and 100% (95%CI: 99.0- 100), respectively, from seven studies and a total of 1008 specimens (Figure 4). There was little heterogeneity across studies.

Additionally, three studies provided sequencing data for RIF resistance, so we compared RealTime MTB RIF/INH performance against sequencing and a composite reference standard (CRS) in these instances (Figure S11). In the paper by Hoffman-Thiel et al., three specimens were classified as resistant by RealTime MTB RIF/INH due to a L511P mutation in the *rpoB* gene, but were sensitive on phenotypic DST [20]. These three specimens were reclassified as true positives with CRS. In the same study, 10 specimens that were susceptible to RIF by index test were resistant by both sequencing and culture (6 specimens with the high confidence mutation H526R and 4 with the moderate confidence mutation L533P mutations). In the paper by Kostera et al., four specimens were classified as susceptible wildtype by the index test and sequencing, but were classified as false negatives by the CRS, as phenotypically they were resistant to RIF [21]. In the smaller study by Tam et al., the index test and reference standards had complete concordance[15]. Thus overall, given the limited number of discordances between the phenotypic and genotypic DST, the results in reference to the different reference standards hardly changed (Figure S11).

INH resistance detection

For INH resistance detection, the pooled sensitivity and specificity were 89% (95%CI: 86.0-92.0) and 99% (95%CI: 98.0-100), respectively, from seven studies and a total of 1013 specimens (Figure 5). There was little heterogeneity across studies.

The same three studies provided data for INH resistance against sequencing. RealTime MTB RIF/INH displayed better accuracy when compared against the sequencing reference standard than against the phenotypic DST. For Hofmann-Thiel, there were 18 specimens that were susceptible by index test but resistant by phenotypic DST[20]. These 18 specimens did not show any mutations in the *katG* or *inhA* target regions using sequencing, so by the CRS we classified them as resistant, since these mutations could have been outside the target regions. Hence the accuracy estimates with CRS in the study were identical to the phenotypic DST. In the study by Kostera et al. 2016, seven discordant specimens that were classified as susceptible phenotypically but INH resistant by index test were confirmed to be resistant by sequencing. This was due to the presence of the *katG* mutation S315T in three cases and an *inhA* protomer region mutation, c -15t, in four cases. These 7 specimens were correctly identified as resistant by the index test but were missed by conventional phenotypic DST (Figure S12).

FluoroType MTB:

Five studies with 2660 respiratory specimens were included in the meta-analysis [13, 23-26]. Median sample size was 608 (IQR: 296 to 661). The assay was performed directly on specimens in all studies, with all but one (4/5, 80%) studies reporting use of fresh specimens. One study used biobanked specimens[13]. Individual sensitivities ranged from 87% to 95%, while specificities ranged from 60% to 100% (Figure 6). Pooled sensitivity and specificity were 92.1% (95%CI: 87.6-93.3) and 98.9% (95%CI: 64.0-99.9), respectively. Obasanya et al observed relatively low specificity of 60% (95%CI: 53.0-66.0), which may be partially explained by the study being conducted in a low resource setting with higher potential for sample contamination, the use of Petroff's method for sputum decontamination, and Löwenstein-Jensen solid culture as the reference standard[26].

Comparator test for TB detection: Xpert MTB/RIF

In assessing Xpert as a comparator test in the same study[26], a substantially higher specificity was observed (94% for Xpert versus 60% for the FluoroType) (Figure 7). However, the specificity of Xpert was lower than the observed specificity of the test for PTB in a large meta-analysis[27]. This study observed Xpert MTB/RIF sensitivity of 79% and FluoroType MTB sensitivity of 89%.

FluoroType MTBDR:

Two studies[28, 29] evaluated FluoroType MTBDR for TB detection using 782 frozen specimens (Table 2). The study by de Vos et al reported a sensitivity of 96% (95%CI: 93-98) and a specificity of 100% (95%CI: 97-100)[28]. Haasis et al reported a sensitivity of 91% (95%CI: 82-97) and specificity of 100% (95%CI: 98-100)[29]. The de Vos study only included Xpert-positive specimens, which could have introduced spectrum bias and an inflated sensitivity estimate.

RIF resistance detection

Two studies[29, 30] assessed the performance of the test for RIF resistance detection using a phenotypic DST. Hillemann et al[30] used culture isolates for FluoroType MTBDR while Haasis performed the testing directly on specimens[29]. Sensitivity was 97% (95%CI: 82.0-100) for Haasis and 99% (95%CI: 96.0-100) for Hillemann and specificity was 100% in both studies. No comparison to sequencing was performed.

INH resistance detection

For isoniazid resistance detection, phenotypic culture was also the reference standard. In Haasis[29] and Hillemann[30], sensitivities were 70% (95%CI: 46.0-88.0) and 92% (95%CI: 84.0-97.0), respectively, and specificity was 100% in both studies. No comparison to sequencing was performed.
For the Hillemann et al[30] study, the use of culture isolates for testing might have resulted in better resistance detection than in Haasis et al[29].

BD Max MDR-TB:

One recently published multicentre study provided data for this assay[31]. The assay was run on fresh sputum specimens. It reported a sensitivity of 93% (95%CI: 89.0-96.0) with specificity of 97% (95%CI: 96.0-98.0) on raw sputum specimens. For decontaminated sputum specimens, the sensitivity was 91% (95%CI: 87.0-94.0) and specificity was 95% (95%CI: 93.0-97.0).

Comparator test for TB detection: Xpert MTB/RIF

The study performed Xpert on the same processed sputum specimens as a comparator test. It reported similar sensitivities of 91% and 90% and specificities of 96% and 98% for BD Max and Xpert, respectively (Figure 8).

RIF resistance detection

For RIF resistance, the sensitivity and specificity with phenotypic DST as reference standard were 90% (95%CI: 55-100) and 95% (95%CI: 91-97), respectively (1 study, 232 specimens).

However, six of eleven specimens classified as false positives by phenotypic DST had *rpoB* mutations identified by Sanger sequencing. Two specimens each had D516Y and L511P mutations, while one specimen each had D516F and L533P mutations, all of which are considered to confer resistance with high or moderate confidence[7]. Based on this reclassification, specificity increased from 95% (211/222) against phenotypic DST to 98% (211/216) with the sequencing and CRS reference standards[31].

INH resistance detection

For INH resistance, the sensitivity and specificity were 82% (95%CI: 63.0-92.0) and 100% (95%CI: 98.0-100), respectively, against phenotypic DST.

Discussion

In this systematic review, we summarise the performance of five diagnostic test for TB and RIF/INH resistance detection: Abbott RealTime MTB, Abbott RealTime MTB RIF/INH, FluoroType MTB, FluoroType MTBDR, and BD Max MDR-TB. Overall, the tests show similar performance to tests currently recommended by the WHO.

Sensitivity across tests was in the range of 90% and above with markedly low observed variability for all assays. For specificity in TB detection, there was more variability across studies and tests and further research needs to be conducted to understand whether this variability is related to test characteristics. For some studies, accuracy estimates were low for both the index test and the comparator (Xpert), which helped in understanding that decreased accuracy could be due to some confounders or study characteristics not stated explicitly[19,26]. Contrastingly, other studies were well conducted and there was more confidence in the diagnostic accuracy of the index tests as the comparators had accuracy estimates which were in-line with the WHO estimates [14, 31]

Conceivably, the different tests might perform differently when it comes to detection of viable and non-viable bacteria depending on the extraction methods and the methods used to enrich whole cell bacteria (e.g. filters)[32, 33]. Therefore, studies recruiting individuals with recent TB history that compare index tests to existing WHO recommended tests (such as Xpert MTB/RIF) would be useful. As well, manual extraction methods, such as those employed by Obasanya et al[26] for Fluorotype MTB, in the hands of less experienced users might have contributed to contamination and thus false positive results.

For RIF and INH resistance detection, the sensitivity and specificity estimates were also in the range of the published accuracy estimates for Xpert[27] and LPA[34]. Although data was limited and variability was observed, which might relate to how the tests were performed (e.g. from isolates or sample) or the study populations.

Three assays were evaluated for the detection of RIF and INH resistance. Abbott RealTi*m*e RIF/INH assay was the only assay that had sufficient data for meta-analysis, with pooled

sensitivity and specificity for RIF resistance of 94% and 100%, respectively, and for INH resistance 89% and 99%, respectively. For the other two assays, data was insufficient to meta-analyze, but overall diagnostic accuracy for RIF and INH resistance detection at this point appeared comparable to that of the WHO-recommended LPA test (90%)[34]. The use of CRS increased the specificity in some studies due to the identification of disputed mutations by sequencing that went undetected by phenotypic DST[7]. All studies that provided sequencing information performed targeted Sanger sequencing, which is a limitation as only targeted sequences can be identified, compared to whole genome sequencing which would provide information on the entire genome and thus identify resistance conferring mutations outside of target regions such as rpoB. A concerning finding to be noted was that in a study[20] of RealTime MTB RIF/INH where six specimens were identified as susceptible to RIF by index test despite the presence of the high confidence mutation H526R. This finding needs to be further assessed in additional studies. S315T is a frequent *katG* mutation and arises typically before all other drug mutations. It is also one of the mutations termed as "harbinger mutations". Its early detection may help in preventing multidrug resistance transmission[35]. In the current systematic review, Abbott RealTime RIF/INH assay picked up this mutation correctly in three specimens in comparison to the phenotypic DST[21]. There was insufficient data to assess these mutations in other assays included in the review.

Only for the BD MAX MDR-TB a single well-conducted study provided information across a well-characterized and representative population. For other tests, HIV status, gender, TB history, and TB treatment status were not available for 70% of the datasets included in the analyses, making generalizability to specific settings difficult. For these tests, additional studies are needed that provide more demographic information for the samples tested to allow for further generalizability of the data.

Operational characteristics are also a critical component for the use of testing platforms in different settings. The throughput of all of the mostly automated platforms assessed in this study is large. Specifically, the number of specimens that can be processed in these platforms vary from 24 (BD) to 94 specimens (Abbott, Hain, Roche). The turnaround time vary from 3 to 5 hours as available from the company manufacturers' package inserts. All of the platforms can be connected to central laboratory information management systems, which is beneficial for disseminating reports to clinicians and patients without delay. Furthermore, the platforms are able to run a large portfolio of assays for different diseases, with Abbott having the largest among the tests evaluated. As such the assays are suited for centralized settings and can provide results to many patients with minimal hands-on manipulation. This limits infection risk to healthcare workers and laboratory technicians, as well as the risk of sample contamination. All tests demonstrated sensitivity for smear-negative cases comparable to Xpert MTB/RIF assay, making them good contenders for this frequently difficult-to-diagnose use case.
However, the tests are not suited for use in lower levels of the healthcare system where patients first present for care. And for the platforms to have the same impact than near-patient platforms, specimen transport needs to optimized. In addition, without reliable systems in place to deliver test results to patients, the impact of these centralized platforms will be very limited, despite their high performance.

An important strength of our systematic review and meta-analysis was that we provided head-to-head comparisons of the index test with Xpert MTB/RIF, a WHO recommended molecular test[8]. Additionally, we also used multiple reference standards for evaluating drug resistance, which provided information on the mutations captured or missed by the index tests. However, the review and meta-analysis also had some limitations. As most of these tests are very new to market, there was minimal data to perform more detailed analyses. Most of the studies were laboratory-based studies, and therefore demographic data of the included participants were not provided. Thus, the generalisability of the performances of all tests (with the exception of BD) is uncertain. Another potential concern was that most of the studies had test manufacturers' involvement.

In summary, for patients with pulmonary TB, these centralized molecular assays demonstrate promising diagnostic accuracy for TB, RIF resistance, and INH resistance detection. While data was limited, the performance of these assays appears similar to that of WHO-recommended Xpert and LPA assays. The assays might prove to have operational advantages in some settings, but further research is necessary.

Acknowledgments
We would like to thank Genevieve Gore, McGill Librarian for helping with the literature search for this review. We would also like to thank the Government of the United Kingdom for providing financial support.

References

1. World Health Organization. Global Tuberculosis Report 2018. Geneva, World Health Organization 2018.
2. UNITAID. TUBERCULOSIS: Diagnostics Technology Landscape. UNITAID 2017.
3. Dicks KV and Stout JE. Molecular Diagnostics for Mycobacterium tuberculosis Infection [Annual Review of Medicine]. 2019;70(1):77-90.
4. World Health Organization. Treatment guidelines for isoniazid-resistant tuberculosis: Supplement to the WHO treatment guidelines for drug-resistant tuberculosis 2018.
5. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097.
6. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. 2008.
7. Miotto P, Cabibbe AM, Borroni E, Degano M, Cirillo DM. Role of Disputed Mutations in the rpoB Gene in Interpretation of Automated Liquid MGIT Culture Results for Rifampin Susceptibility Testing of Mycobacterium tuberculosis. J Clin Microbiol. 2018; 56(5): e01599-17.
8. Schumacher SG, Wells WA, Nicol MP, Steingart KS, Theron G, Dorman SE et al. JID 2019: 220 (Suppl 3) • S99- S107
9. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al.. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-536.
10. Berhanu RH, David A, da Silva P, Shearer K, Sanne I, Stevens W, Scott L. Performance of Xpert MTB/RIF, Xpert Ultra, and Abbott Real Time MTB for the diagnosis of pulmonary tuberculosis in a high HIV burden setting. Journal of Clinical Microbiology 2018;56(12):e00560-18.
11. Chen, J. H.; She, K. K.; Kwong, T. C.; Wong, O. Y.; Siu, G. K.; Leung, C. C. et al. Performance of the new automated Abbott RealTime MTB assay for rapid detection of Mycobacterium tuberculosis complex in respiratory specimens. European Journal of Clinical Microbiology & Infectious Diseases 2015;34(9):1827-32.
12. Hinic, V.; Feuz, K.; Turan, S.; Berini, A.; Frei, R.; Pfeifer, K.; Goldenberger, D. Clinical evaluation of the Abbott RealTime MTB Assay for direct detection of Mycobacterium tuberculosis-complex from respiratory and non-respiratory samples. Tuberculosis 2017;104:65-9.
13. Hofmann-Thiel, S.; Molodtsov, N.; Antonenka, U.; Hoffmann, H. Evaluation of the Abbott RealTime MTB and RealTime MTB INH/RIF Assays for Direct Detection of Mycobacterium tuberculosis Complex and Resistance Markers in Respiratory and Extrapulmonary Specimens. Journal of Clinical Microbiology 2016;54(12):3022-27.
14. Scott, L.; David, A.; Noble, L.; Nduna, M.; Da Silva, P.; Black, A. et al. Performance of the Abbott RealTime MTB and MTB RIF/INH Assays in a Setting of High Tuberculosis and HIV Coinfection in South Africa. Journal of Clinical Microbiology 2017;55(8):2491-501.
15. Tam, K. K.; Leung, K. S.; To, S. W.; Siu, G. K.; Lau, T. C.; Shek, V. C. et al. Direct detection of Mycobacterium tuberculosis and drug resistance in respiratory specimen using Abbott Realtime MTB detection and RIF/INH resistance assay. Diagnostic Microbiology & Infectious Disease 2017;89(2):118-24.

16. Tang, N.; Frank, A.; Pahalawatta, V.; Lampinen, J.; Coblenz-Korte, A.; Dunn, C. et al. Analytical and clinical performance of Abbott RealTime MTB, an assay for detection of Mycobacterium tuberculosis in pulmonary specimens. Tuberculosis 2015;95(5):613-9.
17. Vinuesa, V.; Navarro, D.; Poujois, S.; Zaragoza, S.; Borras, R. Performance characteristics of the new Abbott Real Time MTB assay for detection of Mycobacterium tuberculosis complex in respiratory specimens. Diagnostic Microbiology & Infectious Disease 2016; 84(3):212-4.
18. Vinuesa, V.; Borras, R.; Briones, M. L.; Clari, M. A.; Cresencio, V.; Gimenez, E. et al. Performance of a highly sensitive mycobacterium tuberculosis complex real-time PCR assay for diagnosis of pulmonary tuberculosis in a low-prevalence setting: a prospective intervention study. Journal of clinical microbiology 2018;56(5).
19. Wang, S. F.; Ou, X. C.; Li, Q.; Zheng, H. W.; Wang, Y. F.; Zhao, Y. L.. The Abbott RealTime MTB assay and the Cepheid GeneXpert assay show comparable performance for the detection of Mycobacterium tuberculosis in sputum specimens. International Journal of Infectious Diseases 2016; 45:78-80.
20. Hofmann-Thiel S, Molodtsov N, Duffner C, Kadyrov A, Kalmambetova G, Kabirov O, et al. Capacity of Abbott RealTime MTB RIF/INH to detect rifampicin and isoniazid resistant tuberculosis. International Journal of Tuberculosis and Lung Diseases 2019 ;23(4):458-464
21. Kostera, J.; Leckie, G.; Tang, N.; Lampinen, J.; Szostak, M.; Abravaya, K, et al. Analytical and clinical performance characteristics of the Abbott RealTime MTB RIF/INH Resistance, an assay for the detection of rifampicin and isoniazid resistant Mycobacterium tuberculosis in pulmonary specimens. Tuberculosis 2016;101:137-43.
22. Kostera, J.; Leckie, G.; Abravaya, K.; Wang, H. Performance of the Abbott RealTime MTB RIF/INH resistance assay when used to test Mycobacterium tuberculosis specimens from Bangladesh. Infection and Drug Resistance 2018;11:695-699.
23. Eigner, U., Veldenzer, A., Holfelder, M. Evaluation of the FluoroType MTB assay for the rapid and reliable detection of Mycobacterium tuberculosis in respiratory tract specimens. Clinical Laboratory 2013;59(9-10):1179-81.
24. Genc GE, Satana D, Yildirim E, Erturan Z, Yegenoglu Y, Uzun M. Evaluation of FluoroType MTB for direct detection of Mycobacterium tuberculosis complex and GenoType MTBDRplus for determining rifampicin and isoniazid resistance. Biotechnology & Biotechnological Equipment 2018.
25. Hofmann-Thiel, S., Hoffmann, H. Evaluation of Fluorotype MTB for detection of Mycobacterium tuberculosis complex DNA in clinical specimens from a low-incidence country. BMC Infectious Diseases 2014;14:59.
26. Obasanya, J.; Lawson, L.; Edwards, T.; Olanrewaju, O.; Madukaji, L.; Dacombe, R. et al. FluoroType MTB system for the detection of pulmonary tuberculosis. European Respiratory Journal Open Research 2017;3.
27. Horne DJ, Kohli M, Zifodya JS, Schiller I, Dendukuri N, Tollefson D, et al. Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* 2019;**7**(6):CD009593.
28. De Vos M, Derendinger B, Dolby T, Simpson J, van Helden PD, Rice JE, et al. Diagnostic Accuracy and Utility of FluoroType MTBDR, a New Molecular Assay for Multidrug-Resistant Tuberculosis. Journal of Clinical Microbiology 2018;56(9):e00531-18.
29. Haasis C, Rupp J, Andres S, Schlüter B, Kernbach M, Hillemann D. Validation of the FluoroType® MTBDR assay using respiratory and lymph node samples. Tuberculosis113:76-80.

30. Hillemann, D.; Haasis, C.; Andres, S.; Behn, T.; Kranzer, K. Validation of the FluoroType MTBDR assay for detection of rifampin and isoniazid resistance in mycobacterium tuberculosis complex isolates. Journal of clinical microbiology 2018;56(6).
31. Shah M, Paradis S, Betz J, Beylis N, Bharadwaj R, Caceres T, et al. Multicenter Study of the Accuracy of the BD MAX Multidrug-resistant Tuberculosis Assay for Detection of *Mycobacterium tuberculosis* Complex and Mutations Associated With Resistance to Rifampin and Isoniazid, *Clinical Infectious Diseases*, 2019, ciz932, https://doi.org/10.1093/cid/ciz932
32. Theron G, Venter R, Calligaro G, Smith L, Limberis J, Meldau R, et al. Xpert MTB/RIF results in patients with previous tuberculosis: can we distinguish true from false positive results? Clinical Infectious Diseases 2016;62:995- 1001
33. Theron G, Venter R, Smith L, Esmail A, Randall P, Sood V, et al. False positive Xpert MTB/RIF results in re-tested patients with previous tuberculosis: frequency, profile, and prospective clinical outcomes. Journal of Clinical Microbiology 2018;56(3):pii: e01696-17.
34. Nathavitharana RR, Cudahy PG, Schumacher SG, Steingart KR, Pai M, Denkinger CM. Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. The European respiratory journal 2017;49(1):1601075. DOI: 10.1183/13993003.01075-2016
35. Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE, et al. 2017. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat Genet 2017; 49:395–402.

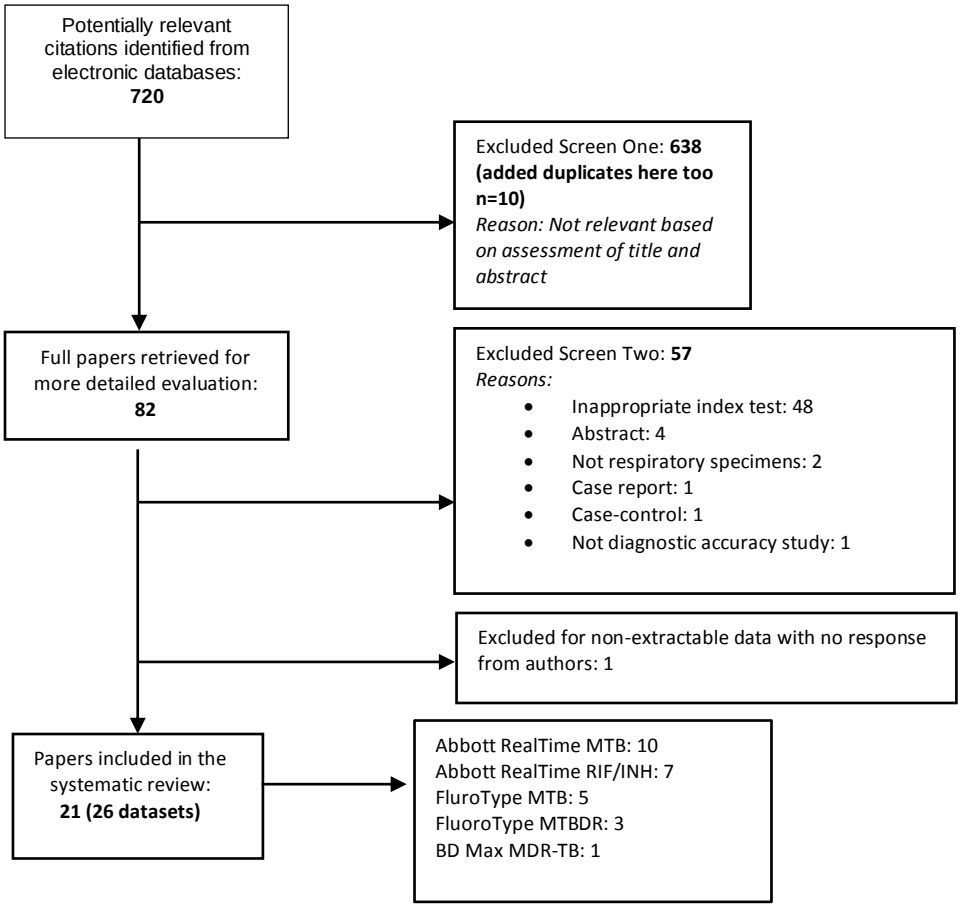

Figure 1: PRISMA diagram of included studies

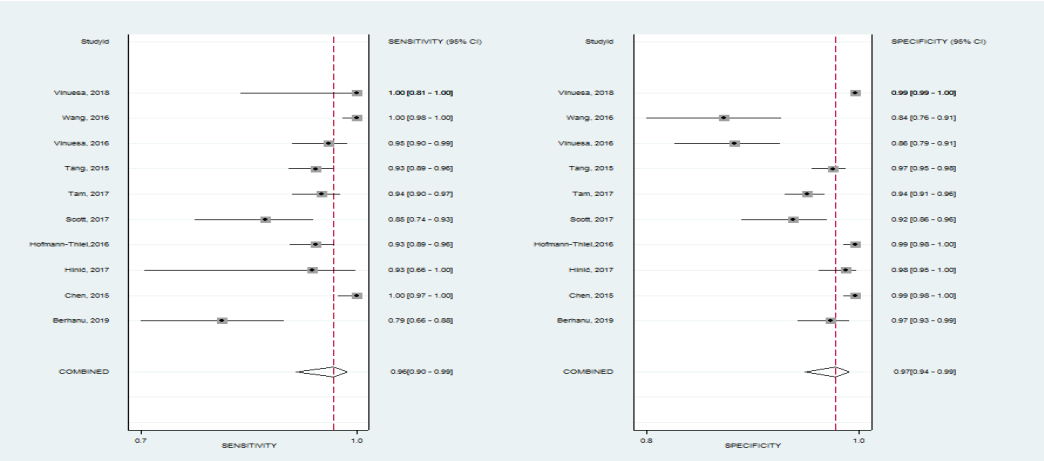

Figure 2: Forest plots for TB detection by Abbott RealTi*m*e MTB

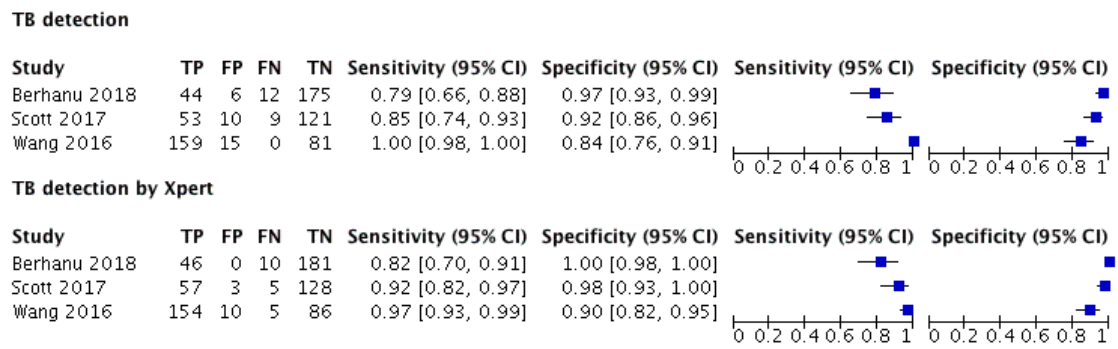

Figure 3a: Forest plot for TB detection with Abbott RealTi*m*e MTB assay and Xpert MTB/RIF with culture as reference standard

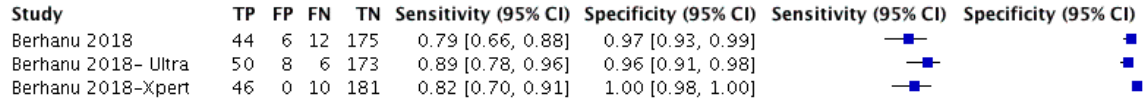

Figure 3b: Forest plot for TB detection with Abbott RealTi*m*e MTB assay, Xpert MTB/RIF and Xpert Ultra with culture as reference standard

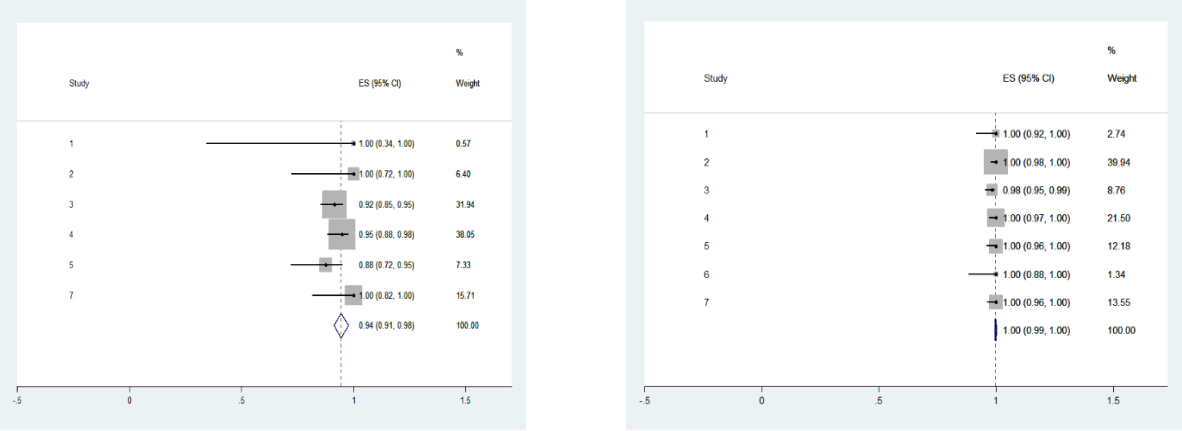

Figure 4: Forest plots for rifampicin resistance detection by Abbott RIF/INH assay using phenotypic DST as reference standard

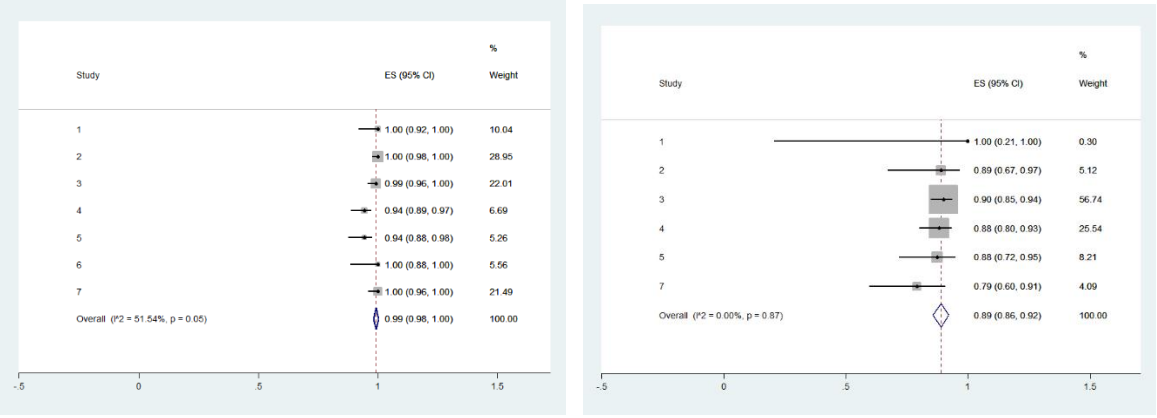

Figure 5: Forest plots for isoniazid resistance detection by Abbott RIF/INH assay using phenotypic DST as reference standard

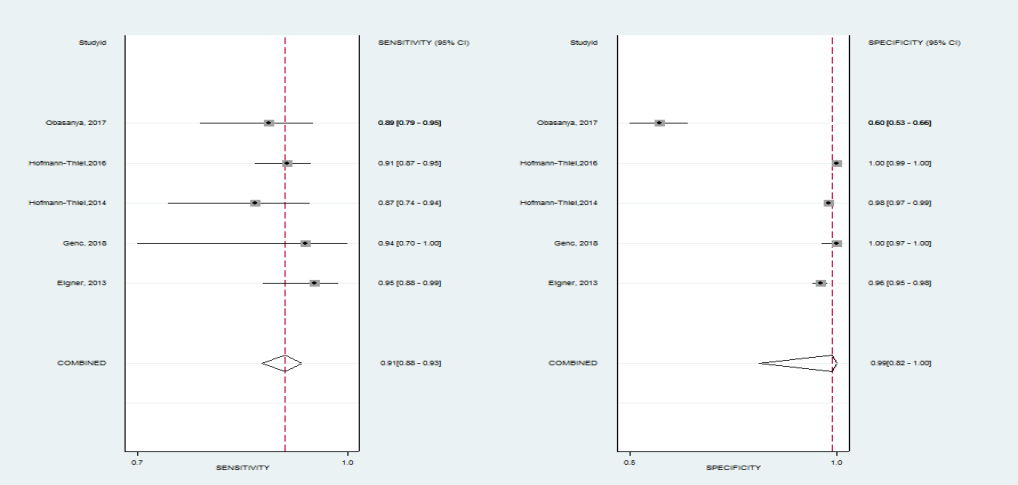

Figure 6: Forest plots for TB detection by FluoroType MTB assay

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Obasanya 2017	62	91	8	135	0.89 [0.79, 0.95]	0.60 [0.53, 0.66]		
Obasanya 2017-Xpert	55	14	15	212	0.79 [0.67, 0.87]	0.94 [0.90, 0.97]		

0 0.2 0.4 0.6 0.8 1 0 0.2 0.4 0.6 0.8 1

Figure 7: Forest plots of TB detection by FluoroType MTB and Xpert MTB/RIF

TB detection by BD Max MDRTB

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Shah 2019	249	27	25	588	0.91 [0.87, 0.94]	0.96 [0.94, 0.97]		

0 0.2 0.4 0.6 0.8 1 0 0.2 0.4 0.6 0.8 1

TB detection by Xpert

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Shah 2019	246	11	28	604	0.90 [0.86, 0.93]	0.98 [0.97, 0.99]		

0 0.2 0.4 0.6 0.8 1 0 0.2 0.4 0.6 0.8 1

Figure 8: Forest plot of TB detection by BD Max MDR-TB and Xpert MTB/RIF

Table 1a. Diagnostic accuracy of each index test- TB detection

Index test	Smear status	# Datasets (#specimens)	Sensitivity (95% CI)	Specificity (95% CI)
Abbott MTB	All	10 (4858)	96.2% (90.2-98.6)	97.1% (93.7-98.7)
	Positive	10 (765)	99.0% (97.7-100)	-
	Negative	10 (4056)	88.4% (74.0-95.3)	98.3% (96.3-99.2)
FluoroType MTB	All	5 (2660)	92.1% (87.6-93.3)	98.9% (64.0-99.9)
	Positive*	3 (174)	Range: 100% (92-100)	-
	Negative*	3 (1754)	Range: 30%-85%	Range:62%-98%
FluoroType MTBDR*	All	2 (782)	Range: 91%-96%	Range: 100% (97-100)
	Positive	2 (288)	Range: 98%-100%	-
	Negative	2 (494)	Range: 69%-98%	Range: 100% (97-100)
BD Max MDR-TB*	All	1 (892)	93% (89.0- 96.0)	97% (96.0- 98.0)
	Positive	1 (176)	100% (98–100)	-
	Negative	3 (713)	81% (73–88%)	98% (96–99%)

CI: Confidence interval; # = number of;
*These datasets were not analysed as the number of studies were less than four and could not be analyzed.

Table 1b. Diagnostic accuracy of each index test - resistance detection

Index test	# Datasets (# specimens)	Sensitivity (95% CI)	Specificity (95% CI)
Abbott RIF/INH			
Rifampicin resistance	7 (1008)	94% (89-99)	100% (99-100)
Isoniazid resistance	7 (1013)	89% (86-92)	99% (98-100)
FluoroType MTBDR*			
Rifampicin resistance	2 (231)	Range: 97%-99%	Range: 100% (85-100)
Isoniazid resistance	2 (207)	Range: 70%-92%	Range: 100% (84-100)
BD Max MDR-TB*			
Rifampicin resistance	1 (232)	90% (55-100)	95% (91-97)
Isoniazid resistance	1 (232)	82% (63–92)	100% (98-100)

CI: Confidence interval; # = number of
* These datasets were not meta-analyzed as the number of studies were less than four and could not be analyzed.

Table 2: Head to head comparisons of the index test with Xpert MTB/RIF

Index test	Smear status	# Datasets (# specimens)	Sensitivity (95% CI)	Specificity (95% CI)
Head to Head comparisons (Abbott RealTime MTB and Xpert MTB/RIF)				
Berhanu 2018				
Abbott RealTime MTB	All	1 (237)	79% (66-88)	97% (93-99)
Xpert MTB/RIF	All		82% (70-91)	100% (98-100)
Scott 2017				
Abbott RealTime MTB	All	1 (193)	85% (74-93)	92% (86-96)
Xpert MTB/RIF	All		92% (82-97)	98% (93-100)
Wang 2016				
Abbott RealTime MTB	All	1 (255)	100% (98-100)	84% (76-91)
Xpert MTB/RIF	All		97% (93-99)	90% (82-95)
Head to head comparisons (FluoroType MTB and Xpert MTB/RIF)				
Obasanya 2017				
FluoroType MTB	All	1 (296)	89% (79-95)	60% (53-66)
Xpert MTB/RIF	All		79% (67-87)	94% (90-97)
Head to head comparisons (BD Max MDR-TB and Xpert MTB/RIF)				
Shah 2019				
BD Max MDR-TB	All	1 (889)	91% (87-94)	96% (94-97)
Xpert MTB/RIF	All		90% (86-93)	98% (97-99)

Diagnostic accuracy of centralized assays for TB detection and detection of resistance to rifampicin and isoniazid: A systematic review and meta-analysis

Mikashmi Kohli[1,2]; Emily MacLean[1,2]; Claudia M. Denkinger[3,4]; Madhukar Pai[1,2]; Samuel G. Schumacher[3]

Supplementary Material

Corresponding author:
Samuel G. Schumacher, MSc, PhD
Deputy Head of TB
FIND, Geneva
Email: samuel.schumacher@finddx.org

Database: Embase <1996 to 2018 Week 26>
Search Strategy:

---

1 (rifampin* or rifampicin* or Isoniazid*).mp. (73472)

2 (MDR TB or MDRTB or RRTB or RR TB or DRTB or DR TB).mp. (4878)

3 exp tuberculosis/ or mycobacterium tuberculosis/ or tuberculosis control/ or rifampicin/ or isoniazid/ (189174)

4 (tubercul* or antitubercul* or tb).mp. (190575)

5 1 or 2 or 3 or 4 (231179)

6 (((Abbott or RealTime* or Real Time*) adj (mtb* or rif* or inh*)) or fluorotype* or bd max* or bdmax* or cobas* taqman*).mp. (1612)

7 *real time polymerase chain reaction/ (10598)

8 ((real time or realtime or rt or direct) and (pcr or polymerase chain reaction)).ti. (18450)

9 6 or 7 or 8 (22380)

10 5 and 9 (574)

11 limit 10 to yr="2009 -Current" (447)

12 limit 11 to dc=20171205-20180628 (17)

13 remove duplicates from 12 (17)

QUADAS-2 Protocol

Domain 1 Patient Selection:
Risk of Bias: Could the selection of patients have introduced bias?

- Signaling question 1: Was a consecutive or random sample of patients or specimens enrolled?
    - We scored ‘yes’ if the study enrolled a consecutive or random sample of eligible patients; ‘no’ if the study selected patients by convenience, and ‘unclear’ if the study did not report the manner of patient selection or this cannot be discerned.

- Signaling question 2: Was a case-control design avoided?
    - We scored ‘yes’ if the study enrolled only patients presumed of drug-resistant TB, including patients with confirmed TB. We scored ‘no’ if the study enrolled patients for whom resistance status was already known, and ‘unclear’ if the study did not report the design or this cannot be discerned.
- Signaling question 3: Did the study avoid inappropriate exclusions?
    - We scored ‘yes’ if no inappropriate exclusions were noted. We scored ‘no’ if studies note specific exclusions. Inappropriate exclusions could potentially occur if patients were excluded based on prior knowledge or testing about them or if the technician does not record performed test results but this was not anticipated for research studies in this review.

Applicability: Are there concerns that the included patients and setting do not match the review question?

We were interested in how the index tests (centralized molecular DST assays) performed in patients presumed of having TB who are evaluated. We judged ‘low concern’ when the specimens included in the study were from the patients with presumptive pulmonary TB and was conducted in high TB and/or high MDR-TB burden country as per the WHO list. We judged ‘high concern’ if the specimens were collected from patients in a low TB and/or MDR-TB burden country. We will judge ‘unclear concern’ if the study included specimens from both high and low TB/MDR-TB burden settings or we could not tell.

Domain 2: Index Test

Risk of Bias: Could the conduct or interpretation of the index test have introduced bias?

- Signaling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?
    - We scored ‘yes’ for all studies because all the centralized molecular DST assay results are automatically generated and the user is provided with printable test results. Thus, there was no room for subjective interpretation of test results.

- Signaling question 2: If a threshold was used, was it prespecified?
    - As the threshold is prespecified in all centralized molecular DST assay in this review, we answered this question "yes" for all studies.

Applicability: Are there concerns that the index test, its conduct, or its interpretation differ from the review question? Variations in test technology, execution, or interpretation may affect estimates of the diagnostic accuracy of a test.

We judged ‘low concern’ if the test was done as per recommendation of the manufacturer for PTB specimens. We judged ‘high concern’ it was stated and/or if additional steps were used for sample preparation and ‘unclear concern’ if we could not tell.

Domain 3: Reference Standard

Risk of Bias: Could the reference standard, its conduct, or its interpretation have introduced bias?

- Signaling question 1: Is the reference standard likely to correctly classify the target condition?
    - For detection of TB, culture is generally considered the best reference standard. We scored 'yes' if the studies used MGIT 960 as the reference standard (higher quality reference standard). We scored 'no' if the studies used only solid media-based culture (lower quality reference standard) as all these index tests are for centralized settings, we expect the laboratory settings to have liquid culture for detecting TB. LJ culture has lower diagnostic accuracy than liquid culture and would over or under-estimate the diagnostic accuracy of the index test. We scored 'unclear' if we could not tell.
    - For detection of rifampicin resistance, culture-based drug susceptibility testing (DST, also called conventional phenotypic method) is considered to be the best reference standard. As we extracted data for studies that used culture-based DST, we will score " yes" for all studies.

- Signaling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?
    - We scored 'yes' if the reference test provided was culture e.g. MGIT 960 DST where an automated result is generated (except for LJ with confirmation of MTB by a NAAT-based test), if blinding was explicitly stated, or if it was clear that the reference standard was performed at a separate laboratory and/or performed by different people. We will score 'no' if the study stated that the reference standard was interpreted with knowledge of the index test result. We scored 'unclear' if this was not stated or answered inadequately.

- Signaling question 3: (Rifampicin resistance) Were the reference standard results interpreted without knowledge of the results of the index test?
    - We added a signaling question for rifampicin resistance detection. We scored "yes" if the reference test provided an automated result (for example, MGIT 960), blinding was explicitly stated, or it was clear that the reference standard was performed at a separate laboratory or performed by different people, or both. We scored "no" if the study stated that the reference standard result was interpreted with knowledge of the index test result. We scored "unclear" if we could not tell.

Applicability: Are there concerns that the target condition as defined by the reference standard does not match the question?
We judged applicability to be of 'low concern' for all studies.

Domain 4: Flow and Timing
Risk of Bias: Could the patient flow have introduced bias?

- Signaling question 1: Was there an appropriate interval between the index test and reference standard?
    - We scored ‘‘yes’ if the tests were paired or separated by less than 48 hours after treatment initiation. We scored ‘no’ if the reference and index tests were not performed on paired specimens or were separated by more than a week. We scored ‘unclear’ if this was not stated in the paper or answered inadequately. In the majority of included studies, we expected specimens for index tests and culture to be obtained at the same time (i.e. to be performed on paired specimens for the majority of studies), when patients are presumed of having TB or MDR-TB.

- Signaling question 2: Did all patients receive the same reference standard?
    - For the diagnosis of TB, we scored this question "yes" if all participants in the study or a subset of participants in the study (for whom we will extract data) received the acceptable reference standard (solid culture, liquid culture, or both), which we specified as a criterion for inclusion in the review. However, we acknowledge that it is possible that some specimens could undergo solid culture and others liquid culture as the reference standard. This variation was recorded.
    - For rifampicin resistance detection, we scored "yes" if all participants received the same reference standard (either culture-based DST or MTBDR*plus*), "no" if not all participants received the same reference standard, and "unclear" if we could not tell.

- Signaling question 3: Were all patients included in the analysis?
- The answer to this question was determined by comparing the number of patients enrolled with the number of patients included in the two-by-two tables. We noted if authors record the number of indeterminate results. We scored ‘yes’ if the number of participants enrolled was clearly stated and corresponded to the number presented in the analysis or if exclusions were adequately described. We scored 'no’ if there were participants missing or excluded from the analysis and there was no explanation given; and 'unclear ' if not enough information was given to assess whether participants were excluded from the analysis

QUADAS-2 summaries – Risk of bias and applicability concerns

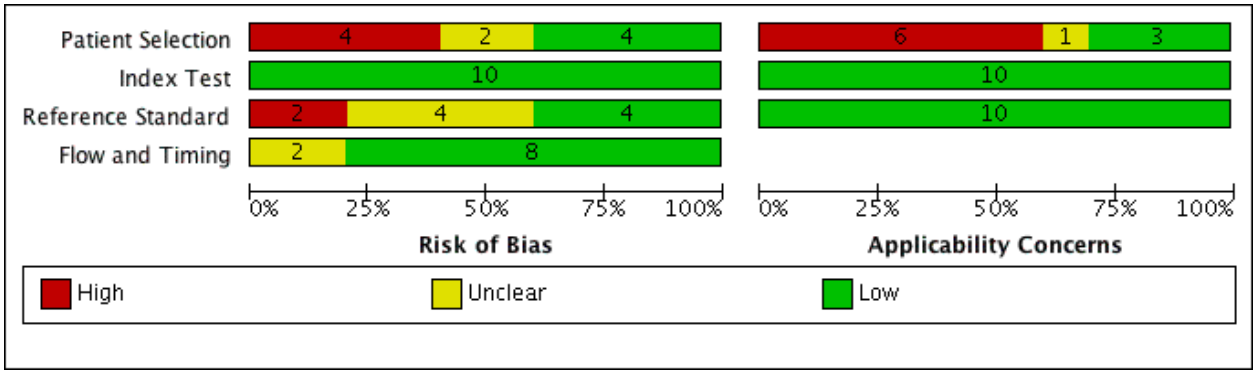

Figure S1. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain for Abbott RealTime MTB assay

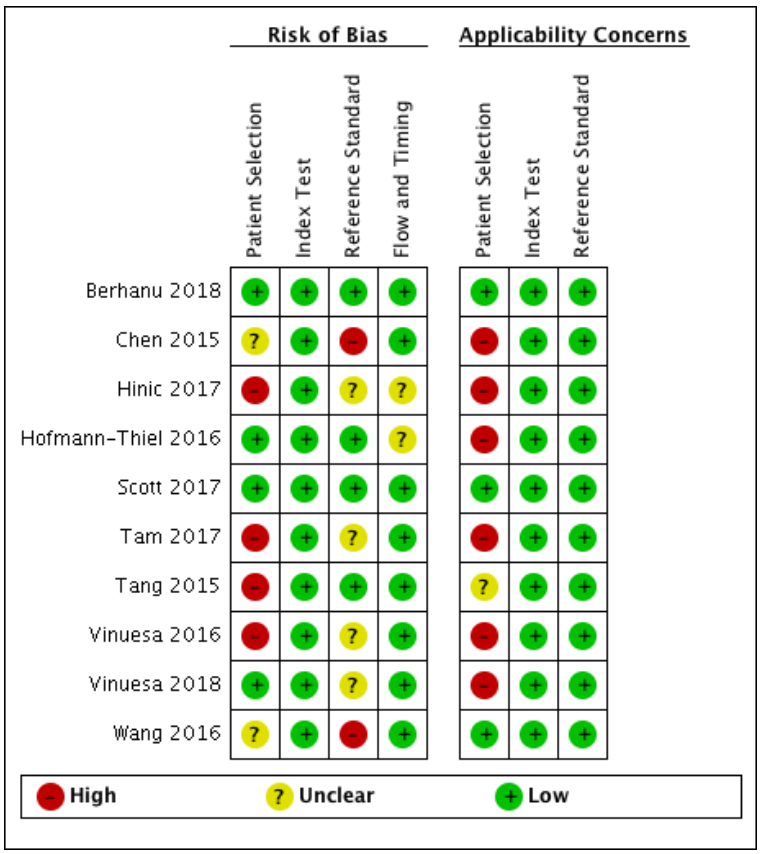

Figure S2. Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating Abbott RealTime MTB assay

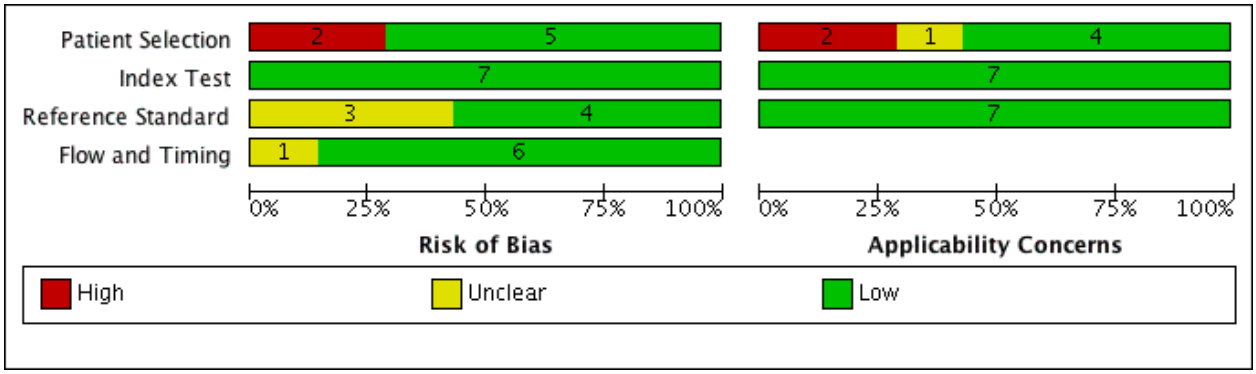

Figure S3. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for Abbott RealTime MTB RIF/INH assay

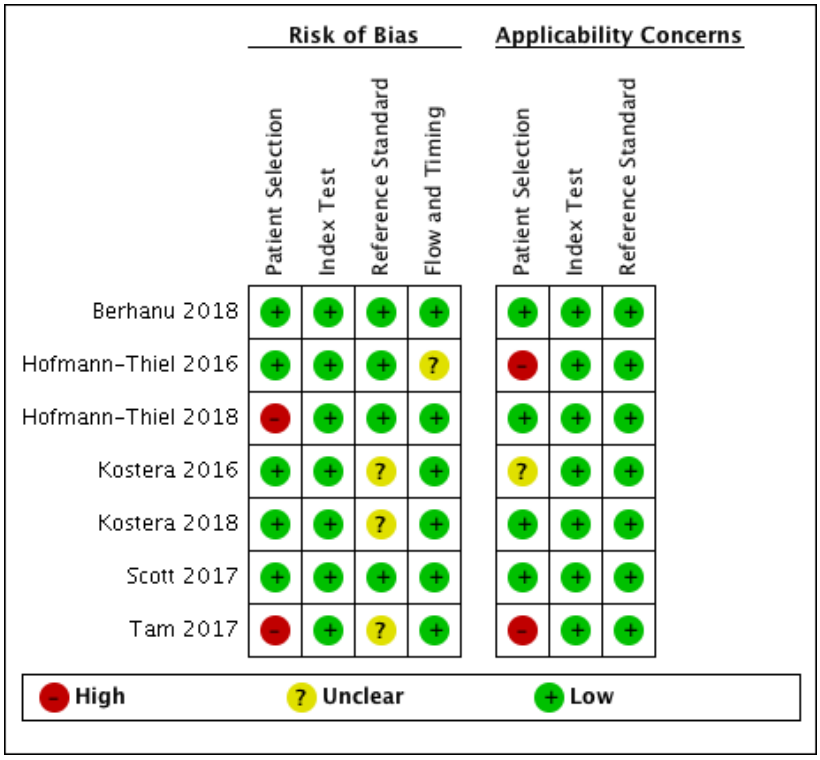

Figure S4. Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating Abbott RealTime MTB RIF/INH assay

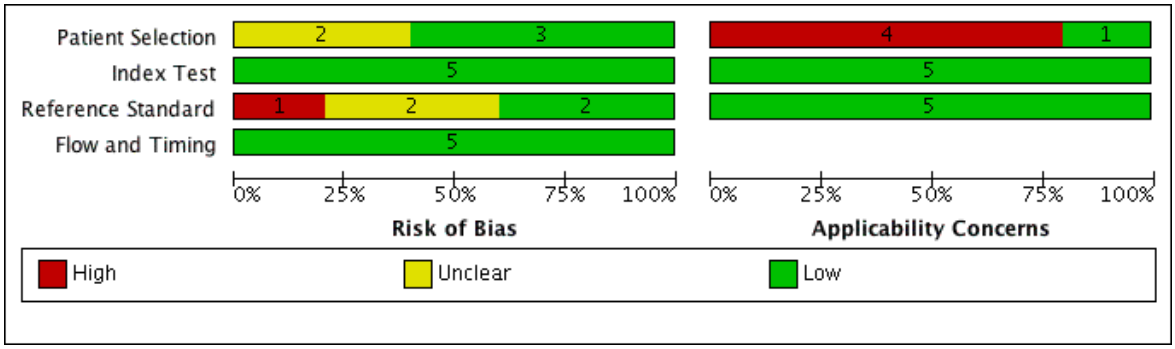

Figure S5. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for FluoroType MTB assay

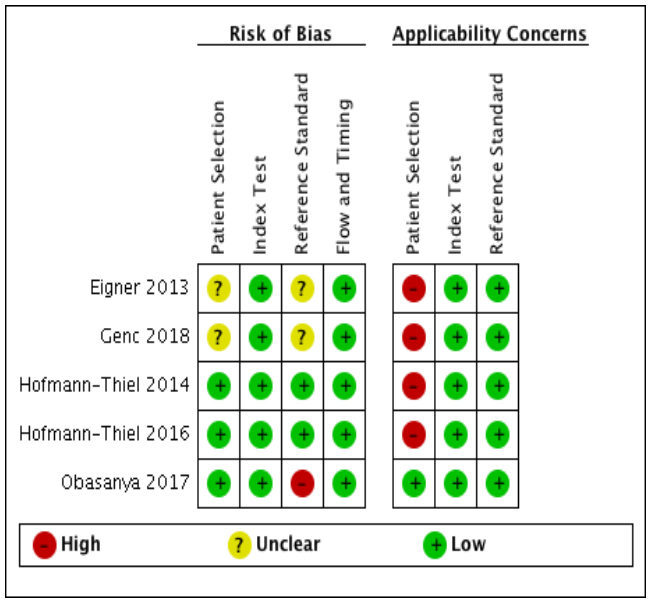

Figure S6: Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating FluoroType MTB assay

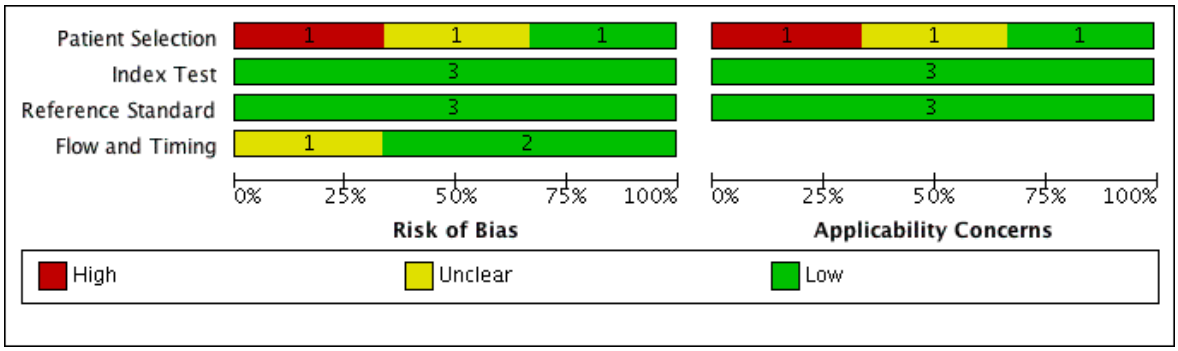

Figure S7. Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for FluoroType MTBDR assay

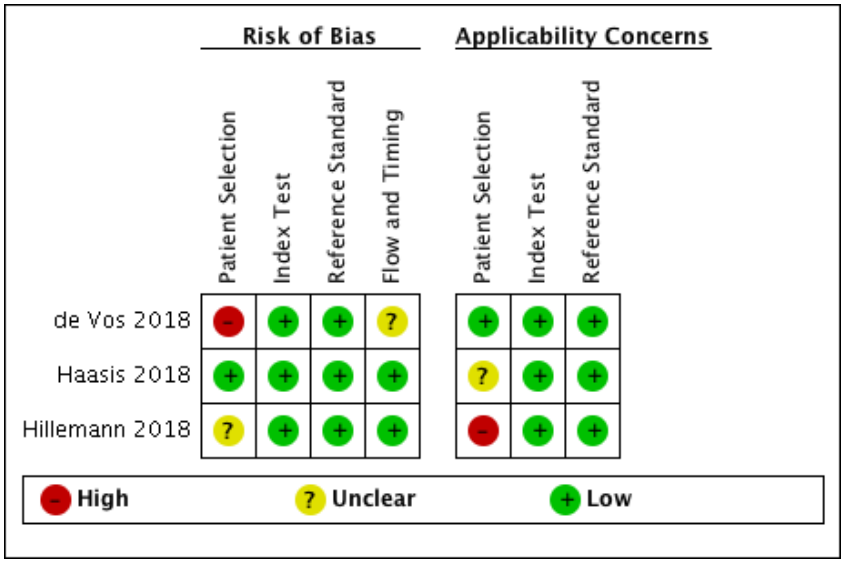

Figure S8. Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating FluoroType MTBDR assay

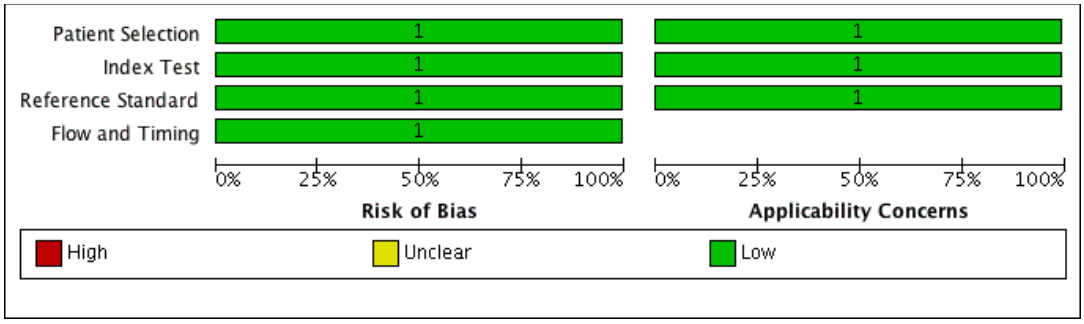

Figure S9: Risk of Bias and Applicability Concerns summary about each QUADAS-2 domain presented as percentages for BD Max MDR-TB assay

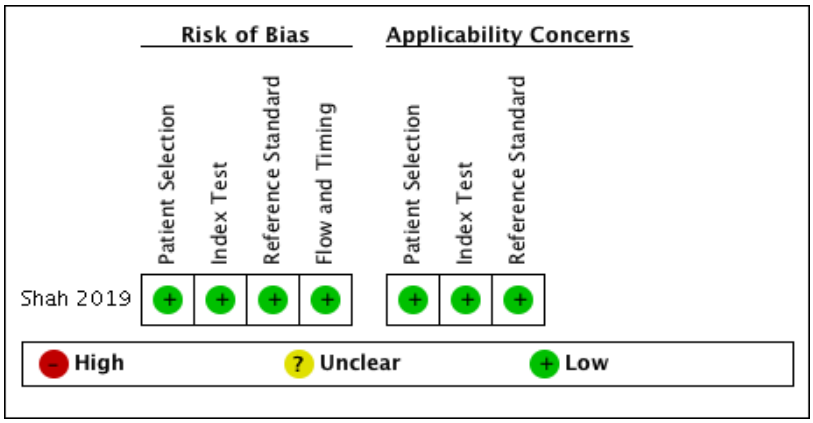

Figure S10: Risk of Bias and Applicability Concerns summary for QUADAS-2 domains in each study evaluating BD Max MDR-TB assay

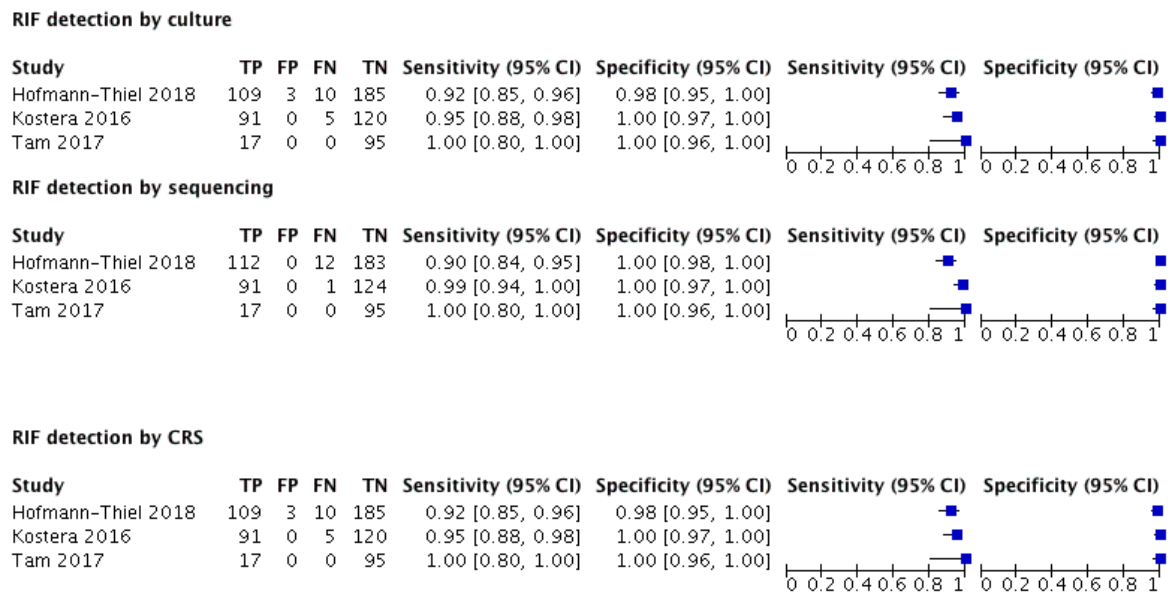

Figure S11: Forest plots for rifampicin resistance detection by Abbott RIF/INH assay using phenotypic DST, sequencing and composite reference standard

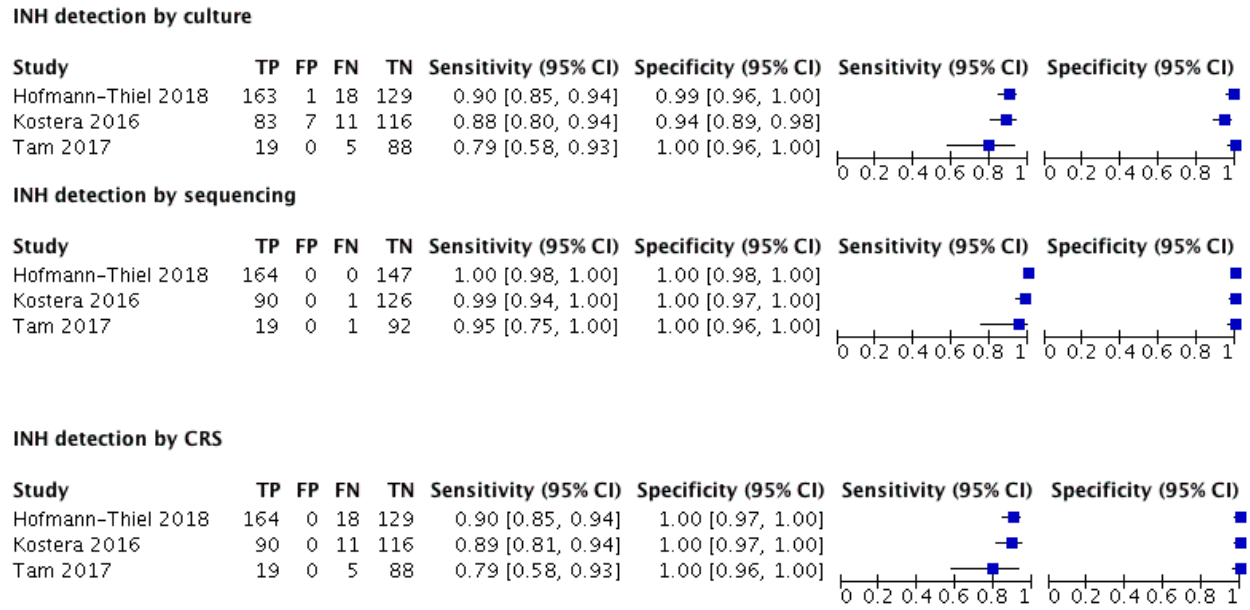

Figure S12: Forest plots for isoniazid resistance detection by Abbott RIF/INH assay using phenotypic DST, sequencing and composite reference standard