



**Cochrane**  
**Library**

Cochrane Database of Systematic Reviews

## GenoType® MTBDRsl assay for resistance to second-line anti-tuberculosis drugs (Review)

Theron G, Peter J, Richardson M, Warren R, Dheda K, Steingart KR

Theron G, Peter J, Richardson M, Warren R, Dheda K, Steingart KR.  
GenoType® MTBDRsl assay for resistance to second-line anti-tuberculosis drugs.  
*Cochrane Database of Systematic Reviews* 2016, Issue 9. Art. No.: CD010705.  
DOI: 10.1002/14651858.CD010705.pub3.

[www.cochranelibrary.com](http://www.cochranelibrary.com)

## TABLE OF CONTENTS

|  |     |
|--|-----|
| HEADER . . . . .   | 1   |
| ABSTRACT . . . . .   | 1   |
| PLAIN LANGUAGE SUMMARY . . . . .   | 3   |
| SUMMARY OF FINDINGS FOR THE MAIN COMPARISON . . . . .                        | 5   |
| BACKGROUND . . . . .   | 7   |
| Figure 1. . . . .  | 9   |
| Figure 2. . . . .  | 10  |
| OBJECTIVES . . . . .   | 11  |
| METHODS . . . . .  | 11  |
| RESULTS . . . . .  | 15  |
| Figure 3. . . . .  | 16  |
| Figure 4. . . . .  | 17  |
| Figure 5. . . . .  | 18  |
| Figure 6. . . . .  | 20  |
| Figure 7. . . . .  | 23  |
| Figure 8. . . . .  | 24  |
| Figure 9. . . . .  | 26  |
| Figure 10. . . . .   | 28  |
| Figure 11. . . . .   | 30  |
| Figure 12. . . . .   | 31  |
| Figure 13. . . . .   | 33  |
| Figure 14. . . . .   | 35  |
| Figure 15. . . . .   | 36  |
| ADDITIONAL SUMMARY OF FINDINGS . . . . .                                     | 37  |
| DISCUSSION . . . . .   | 42  |
| AUTHORS' CONCLUSIONS . . . . .   | 44  |
| ACKNOWLEDGEMENTS . . . . .   | 45  |
| REFERENCES . . . . .   | 45  |
| CHARACTERISTICS OF STUDIES . . . . .   | 51  |
| DATA . . . . .   | 102 |
| Test 1. Indirect, FQ, culture. . . . .                                       | 105 |
| Test 2. Indirect, ofloxacin, culture. . . . .                                | 106 |
| Test 3. Indirect, moxifloxacin, culture. . . . .                             | 107 |
| Test 4. Indirect, levofloxacin, culture. . . . .                             | 107 |
| Test 5. Indirect, ofloxacin, WHO critical concentration used. . . . .        | 108 |
| Test 6. Indirect, ofloxacin, WHO critical concentration not used. . . . .    | 108 |
| Test 7. Indirect, SLID, culture. . . . .                                     | 109 |
| Test 8. Indirect, amikacin, culture. . . . .                                 | 110 |
| Test 9. Indirect, kanamycin, culture. . . . .                                | 111 |
| Test 10. Indirect, capreomycin, culture. . . . .                             | 112 |
| Test 11. Indirect, amikacin, WHO critical concentration used. . . . .        | 112 |
| Test 12. Indirect, capreomycin, WHO critical concentration used. . . . .     | 113 |
| Test 13. Indirect, amikacin, WHO critical concentration not used. . . . .    | 113 |
| Test 14. Indirect, capreomycin, WHO critical concentration not used. . . . . | 114 |
| Test 15. Indirect, XDR, culture. . . . .                                     | 114 |
| Test 16. Indirect, FQ, sequencing. . . . .                                   | 115 |
| Test 17. Indirect, SLID, sequencing. . . . .                                 | 115 |
| Test 18. Indirect, XDR, sequencing. . . . .                                  | 116 |
| Test 19. Indirect, FQ, sequencing and culture. . . . .                       | 116 |
| Test 20. Indirect, SLID, sequencing and culture. . . . .                     | 117 |
| Test 21. Indirect, XDR, sequencing and culture. . . . .                      | 117 |

|  |     |
|--|-----|
| Test 22. Indirect, FQ, culture followed by sequencing of discrepant.   | 118 |
| Test 23. Indirect, SLID, culture followed by sequencing of discrepant. | 118 |
| Test 24. Direct, FQ, culture.  | 119 |
| Test 25. Direct, ofloxacin, culture.                                   | 119 |
| Test 26. Direct, moxifloxacin, culture.                                | 120 |
| Test 27. Ofloxacin, smear positive.                                    | 120 |
| Test 28. Ofloxacin, smear negative.                                    | 121 |
| Test 29. Ofloxacin, smear grade = scanty.                              | 121 |
| Test 30. Ofloxacin, smear grade = 1+.                                  | 122 |
| Test 31. Ofloxacin, smear grade $\geq$ 2+.                             | 122 |
| Test 32. Moxifloxacin, smear positive.                                 | 123 |
| Test 33. Moxifloxacin, smear negative.                                 | 123 |
| Test 34. Moxifloxacin, smear grade = scanty.                           | 124 |
| Test 35. Moxifloxacin, smear grade = 1+.                               | 124 |
| Test 36. Moxifloxacin, smear grade $\geq$ 2+.                          | 125 |
| Test 37. Direct, SLID, culture.  | 125 |
| Test 38. Direct, amikacin, culture.                                    | 126 |
| Test 39. Direct, kanamycin, culture.                                   | 126 |
| Test 40. Direct, capreomycin, culture.                                 | 127 |
| Test 41. Amikacin, smear positive.                                     | 127 |
| Test 42. Amikacin, smear negative.                                     | 128 |
| Test 43. Amikacin, smear grade = scanty.                               | 128 |
| Test 44. Amikacin, smear grade = 1+.                                   | 129 |
| Test 45. Amikacin, smear grade $\geq$ 2+.                              | 129 |
| Test 46. Kanamycin, smear positive.                                    | 130 |
| Test 47. Kanamycin, smear negative.                                    | 130 |
| Test 48. Kanamycin, smear grade = scanty.                              | 131 |
| Test 49. Kanamycin, smear grade = +1.                                  | 131 |
| Test 50. Kanamycin, smear grade $\geq$ 2+.                             | 132 |
| Test 51. Capreomycin, smear positive.                                  | 132 |
| Test 52. Capreomycin, smear negative.                                  | 133 |
| Test 53. Capreomycin, smear grade = 1+.                                | 133 |
| Test 54. Capreomycin, smear grade = scanty.                            | 134 |
| Test 55. Capreomycin, smear grade $\geq$ 2+.                           | 134 |
| Test 56. Direct, XDR, culture.   | 135 |
| Test 57. Direct, FQ, culture followed by sequencing of discrepant.     | 135 |
| Test 58. Direct, SLID, culture followed by sequencing of discrepant.   | 136 |
| Test 59. Direct, XDR, culture followed by sequencing of discrepant.    | 136 |
| Test 60. V2, Indirect, FQ, culture.                                    | 137 |
| Test 61. V2, Direct, FQ, smear positive.                               | 137 |
| Test 62. V2, Direct, FQ, smear negative.                               | 137 |
| Test 63. V2, Indirect, ofloxacin, culture.                             | 138 |
| Test 64. V2, Indirect, moxifloxacin, culture.                          | 138 |
| Test 65. V2, Indirect, SLID, culture.                                  | 139 |
| Test 66. V2, Direct, SLID, smear positive.                             | 139 |
| Test 67. V2, Direct, SLID, smear negative.                             | 139 |
| Test 68. V2, Indirect, amikacin, culture.                              | 140 |
| Test 69. V2, Indirect, kanamycin, culture.                             | 140 |
| Test 70. V2, Indirect, capreomycin, culture.                           | 141 |
| Test 71. V2, Ofloxacin, smear positive.                                | 141 |
| Test 72. V2, Ofloxacin, smear negative.                                | 142 |
| Test 73. V2, Ofloxacin, smear grade = scanty.                          | 142 |
| Test 74. V2, Ofloxacin, smear grade = 1+.                              | 143 |

|  |     |
|--|-----|
| Test 75. V2, Ofloxacin, smear grade $\geq 2+$ . . . . .    | 143 |
| Test 76. V2, Moxifloxacin, smear positive. . . . .         | 144 |
| Test 77. V2, Moxifloxacin, smear grade $\geq 2+$ . . . . . | 144 |
| Test 78. V2, Levofloxacin, smear positive. . . . .         | 145 |
| Test 79. V2, Levofloxacin, smear grade = scanty. . . . .   | 145 |
| Test 80. V2, Levofloxacin, smear grade = 1+. . . . .       | 146 |
| Test 81. V2, Levofloxacin, smear grade $\geq 2+$ . . . . . | 146 |
| Test 82. V2, Amikacin, smear positive. . . . .             | 147 |
| Test 83. V2, Amikacin, smear negative. . . . .             | 147 |
| Test 84. V2, Amikacin, smear grade = scanty. . . . .       | 148 |
| Test 85. V2, Amikacin, smear grade = 1+. . . . .           | 148 |
| Test 86. V2, Amikacin, smear grade $\geq 2+$ . . . . .     | 149 |
| Test 87. V2, Kanamycin, smear positive. . . . .            | 149 |
| Test 88. V2, Kanamycin, smear negative. . . . .            | 150 |
| Test 89. V2, Kanamycin, smear grade = scanty. . . . .      | 150 |
| Test 90. V2, Kanamycin, smear grade = 1+. . . . .          | 151 |
| Test 91. V2, Kanamycin, smear grade $\geq 2+$ . . . . .    | 151 |
| Test 92. V2, Capreomycin, smear positive. . . . .          | 152 |
| Test 93. V2, Capreomycin, smear negative. . . . .          | 152 |
| Test 94. V2, Capreomycin, smear grade = scanty. . . . .    | 153 |
| Test 95. V2, Capreomycin, smear grade = 1+. . . . .        | 153 |
| Test 96. V2, Capreomycin, smear grade $\geq 2+$ . . . . .  | 154 |
| Test 97. V2, Indirect, XDR, culture. . . . .               | 154 |
| Test 98. V2, Direct, XDR, smear positive. . . . .          | 154 |
| Test 99. V2, Direct, XDR, smear negative. . . . .          | 155 |
| ADDITIONAL TABLES . . . . .                                | 155 |
| APPENDICES . . . . .                                       | 161 |
| Figure 16. . . . .   | 166 |
| Figure 17. . . . .   | 172 |
| Figure 18. . . . .   | 173 |
| Figure 19. . . . .   | 174 |
| WHAT'S NEW . . . . .                                       | 175 |
| CONTRIBUTIONS OF AUTHORS . . . . .                         | 175 |
| DECLARATIONS OF INTEREST . . . . .                         | 175 |
| SOURCES OF SUPPORT . . . . .                               | 175 |
| DIFFERENCES BETWEEN PROTOCOL AND REVIEW . . . . .          | 176 |
| INDEX TERMS . . . . .                                      | 176 |

[Diagnostic Test Accuracy Review]

# GenoType® MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs

Grant Theron<sup>1</sup>, Jonny Peter<sup>2</sup>, Marty Richardson<sup>3</sup>, Rob Warren<sup>4</sup>, Keertan Dheda<sup>5</sup>, Karen R Steingart<sup>3</sup>

<sup>1</sup>DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa. <sup>2</sup>Division of Clinical Immunology and Allergology, Department of Medicine, University of Cape Town, Cape Town, South Africa. <sup>3</sup>Cochrane Infectious Diseases Group, Liverpool School of Tropical Medicine, Liverpool, UK. <sup>4</sup>DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Matieland, South Africa. <sup>5</sup>Lung Infection and Immunity Unit, Department of Medicine, University of Cape Town, Cape Town, South Africa

Contact address: Grant Theron, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa. [gtheron@sun.ac.za](mailto:gtheron@sun.ac.za).

**Editorial group:** Cochrane Infectious Diseases Group.

**Publication status and date:** New search for studies and content updated (no change to conclusions), published in Issue 9, 2016.

**Citation:** Theron G, Peter J, Richardson M, Warren R, Dheda K, Steingart KR. GenoType® MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs. *Cochrane Database of Systematic Reviews* 2016, Issue 9. Art. No.: CD010705. DOI: 10.1002/14651858.CD010705.pub3.

Copyright © 2016 The Authors. Cochrane Database of Systematic Reviews published by John Wiley & Sons, Ltd. on behalf of The Cochrane Collaboration. This is an open access article under the terms of the [Creative Commons Attribution-Non-Commercial](#) Licence, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## ABSTRACT

### Background

Genotype® MTBDR<sub>sl</sub> (MTBDR<sub>sl</sub>) is a rapid DNA-based test for detecting specific mutations associated with resistance to fluoroquinolones and second-line injectable drugs (SLIDs) in *Mycobacterium tuberculosis* complex. MTBDR<sub>sl</sub> version 2.0 (released in 2015) identifies the mutations detected by version 1.0, as well as additional mutations. The test may be performed on a culture isolate or a patient specimen, which eliminates delays associated with culture. Version 1.0 requires a smear-positive specimen, while version 2.0 may use a smear-positive or -negative specimen. We performed this updated review as part of a World Health Organization process to develop updated guidelines for using MTBDR<sub>sl</sub>.

### Objectives

To assess and compare the diagnostic accuracy of MTBDR<sub>sl</sub> for: 1. fluoroquinolone resistance, 2. SLID resistance, and 3. extensively drug-resistant tuberculosis, indirectly on a *M. tuberculosis* isolate grown from culture or directly on a patient specimen. Participants were people with rifampicin-resistant or multidrug-resistant tuberculosis. The role of MTBDR<sub>sl</sub> would be as the initial test, replacing culture-based drug susceptibility testing (DST), for detecting second-line drug resistance.

### Search methods

We searched the following databases without language restrictions up to 21 September 2015: the Cochrane Infectious Diseases Group Specialized Register; MEDLINE; Embase OVID; Science Citation Index Expanded, Conference Proceedings Citation Index-Science, and BIOSIS Previews (all three from Web of Science); LILACS; and SCOPUS; registers for ongoing trials; and ProQuest Dissertations & Theses A&I. We reviewed references from included studies and contacted specialists in the field.

---

GenoType® MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs (Review)

Copyright © 2016 The Authors. Cochrane Database of Systematic Reviews published by John Wiley & Sons, Ltd. on behalf of The Cochrane Collaboration.

1

## Selection criteria

We included cross-sectional and case-control studies that determined MTBDR<sub>sl</sub> accuracy against a defined reference standard (culture-based DST, genetic sequencing, or both).

## Data collection and analysis

Two review authors independently extracted data and assessed quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. We synthesized data for versions 1.0 and 2.0 separately. We estimated MTBDR<sub>sl</sub> sensitivity and specificity for fluoroquinolone resistance, SLID resistance, and extensively drug-resistant tuberculosis when the test was performed indirectly or directly (smear-positive specimen for version 1.0, smear-positive or -negative specimen for version 2.0). We explored the influence on accuracy estimates of individual drugs within a drug class and of different reference standards. We performed most analyses using a bivariate random-effects model with culture-based DST as reference standard.

## Main results

We included 27 studies. Twenty-six studies evaluated version 1.0, and one study version 2.0. Of 26 studies stating specimen country origin, 15 studies (58%) evaluated patients from low- or middle-income countries. Overall, we considered the studies to be of high methodological quality. However, only three studies (11%) had low risk of bias for the reference standard; these studies used World Health Organization (WHO)-recommended critical concentrations for all drugs in the culture-based DST reference standard.

### MTBDR<sub>sl</sub> version 1.0

*Fluoroquinolone resistance:* indirect testing, MTBDR<sub>sl</sub> pooled sensitivity and specificity (95% confidence interval (CI)) were 85.6% (79.2% to 90.4%) and 98.5% (95.7% to 99.5%), (19 studies, 2223 participants); direct testing (smear-positive specimen), pooled sensitivity and specificity were 86.2% (74.6% to 93.0%) and 98.6% (96.9% to 99.4%), (nine studies, 1771 participants, *moderate quality evidence*).

*SLID resistance:* indirect testing, MTBDR<sub>sl</sub> pooled sensitivity and specificity were 76.5% (63.3% to 86.0%) and 99.1% (97.3% to 99.7%), (16 studies, 1921 participants); direct testing (smear-positive specimen), pooled sensitivity and specificity were 87.0% (38.1% to 98.6%) and 99.5% (93.6% to 100.0%), (eight studies, 1639 participants, *low quality evidence*).

*Extensively drug-resistant tuberculosis:* indirect testing, MTBDR<sub>sl</sub> pooled sensitivity and specificity were 70.9% (42.9% to 88.8%) and 98.8% (96.1% to 99.6%), (eight studies, 880 participants); direct testing (smear-positive specimen), pooled sensitivity and specificity were 69.4% (38.8% to 89.0%) and 99.4% (95.0% to 99.3%), (six studies, 1420 participants, *low quality evidence*).

Similar to the original Cochrane review, we found no evidence of a significant difference in MTBDR<sub>sl</sub> version 1.0 accuracy between indirect and direct testing for fluoroquinolone resistance, SLID resistance, and extensively drug-resistant tuberculosis.

### MTBDR<sub>sl</sub> version 2.0

*Fluoroquinolone resistance:* direct testing, MTBDR<sub>sl</sub> sensitivity and specificity were 97% (83% to 100%) and 98% (93% to 100%), smear-positive specimen; 80% (28% to 99%) and 100% (40% to 100%), smear-negative specimen.

*SLID resistance:* direct testing, MTBDR<sub>sl</sub> sensitivity and specificity were 89% (72% to 98%) and 90% (84% to 95%), smear-positive specimen; 80% (28% to 99%) and 100% (40% to 100%), smear-negative specimen.

*Extensively drug-resistant tuberculosis:* direct testing, MTBDR<sub>sl</sub> sensitivity and specificity were 79% (49% to 95%) and 97% (93% to 99%), smear-positive specimen; 50% (1% to 99%) and 100% (59% to 100%), smear-negative specimen.

We had insufficient data to estimate summary sensitivity and specificity of version 2.0 (smear-positive and -negative specimens) or to compare accuracy of the two versions.

A limitation was that most included studies did not consistently use the World Health Organization (WHO)-recommended concentrations for drugs in the culture-based DST reference standard.

## Authors' conclusions

In people with rifampicin-resistant or multidrug-resistant tuberculosis, MTBDR<sub>sl</sub> performed on a culture isolate or smear-positive specimen may be useful in detecting second-line drug resistance. MTBDR<sub>sl</sub> (smear-positive specimen) correctly classified around six in seven people as having fluoroquinolone or SLID resistance, although the sensitivity estimates for SLID resistance varied. The test

rarely gave a positive result for people without drug resistance. However, when second-line drug resistance is not detected (MTBDRs/l result is negative), conventional DST can still be used to evaluate patients for resistance to the fluoroquinolones or SLIDs.

We recommend that future work evaluate MTBDRs/l version 2.0, in particular on smear-negative specimens and in different settings to account for different resistance-causing mutations that may vary by strain. Researchers should also consider incorporating WHO-recommended critical concentrations into their culture-based reference standards.

## PLAIN LANGUAGE SUMMARY

### The rapid test GenoType<sup>®</sup> MTBDRs/l for testing resistance to second-line TB drugs

#### Background

Different drugs are available to treat tuberculosis (TB), but resistance to these drugs is a growing problem. People with drug-resistant TB require second-line TB drugs that, compared with first-line TB drugs, must be taken for longer and may be associated with more harms. Detecting TB drug resistance quickly is important for improving health, reducing deaths, and decreasing the spread of drug-resistant TB.

#### Definitions

Multidrug-resistant TB (MDR-TB) is caused by TB bacteria that are resistant to at least isoniazid and rifampicin, the two most potent TB drugs.

Extensively drug-resistant TB (XDR-TB) is a type of MDR-TB that is resistant to nearly all TB drugs.

#### What test is evaluated by this review?

GenoType<sup>®</sup> MTBDRs/l (MTBDRs/l) is a rapid test for detecting resistance to second-line TB drugs. In people with MDR-TB, MTBDRs/l is used to detect additional drug resistance. The test may be performed on TB bacteria grown in culture from a patient specimen (indirect testing) or on a patient specimen (direct testing), which eliminates delays associated with culture. MTBDRs/l version 1.0 requires a specimen to be smear-positive by microscopy, while version 2.0 (released in 2015) may use a smear-positive or -negative specimen.

#### What are the aims of the review?

We wanted to find out how accurate MTBDRs/l is for detecting drug resistance; to compare indirect and direct testing; and to compare the two test versions.

#### How up-to-date is the review?

We searched for and used studies that had been published up to 21 September 2015.

#### What are the main results of the review?

We found 27 studies; 26 studies evaluated MTBDRs/l version 1.0 and one study evaluated version 2.0.

##### *Fluoroquinolone drugs*

MTBDRs/l version 1.0 (smear-positive specimen) detected 86% of people with fluoroquinolone resistance and rarely gave a positive result for people without resistance (*GRADE, moderate quality evidence*).

##### *Second-line injectable drugs*

MTBDRs/l version 1.0 (smear-positive specimen) detected 87% of people with second-line injectable drug resistance and rarely gave a positive result for people without resistance (*GRADE, low quality evidence*).

##### *XDR-TB*

MTBDRs/l version 1.0 (smear-positive specimen) detected 69% of people with XDR-TB and rarely gave a positive result for people without resistance (*GRADE, low quality evidence*).

For MTBDRs/l version 1.0, we found similar results for indirect and direct testing (smear-positive specimen).

As we identified only one study evaluating MTBDRs<sub>1</sub> version 2.0, we could not be sure of the diagnostic accuracy of version 2.0. Also, we could not compare accuracy of the two versions.

**What is the methodological quality of the evidence?**

We used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool to assess study quality. Overall, we considered the included studies to be of high quality; however, we had concerns about how the reference standard (the benchmark against which MTBDRs<sub>1</sub> was measured) was applied.

**What are the authors' conclusions?**

MTBDRs<sub>1</sub> (smear-positive specimen) identified most of the patients with second-line drug resistance. When the test reports a negative result, conventional testing for drug resistance can still be used.



## SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

**Participants:** patients with rifampicin-resistant or MDR-TB

**Prior testing:** patients who received MTBDR<sub>s/l</sub> testing may have first received smear microscopy, Xpert® MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB and Xpert® MTB/RIF, MTBDR<sub>plus</sub> version 2.0 or an alternative line-probe assay to detect first-line drug resistance

**Role:** The role of MTBDR<sub>s/l</sub> would be as the initial test, replacing culture-based drug susceptibility testing, for detecting second-line drug resistance

**Settings:** intermediate or central level laboratories

**Index (new) test:** MTBDR<sub>s/l</sub> version 1.0.\* The test was performed by direct testing on smear-positive specimens

**Reference standard:** culture-based drug susceptibility testing

**Studies:** mainly cross-sectional studies

**Limitations:** most included studies did not consistently use the World Health Organization (WHO)-recommended concentrations for drugs in the culture-based reference standard

**Pooled sensitivity (95% CI):** 86.2% (74.6% to 93.0%)

**Pooled specificity (95% CI):** 98.6% (96.9% to 99.4%)

| Test result   | Number of results per 1000 patients tested (95% CI) |                   |                   | Number of participants (studies) | Quality of the evidence (GRADE)     |
|---|---|-------------------|-------------------|----------------------------------|-------------------------------------|
|   | Prevalence of 5%                                    | Prevalence of 10% | Prevalence of 15% |                                  |                                     |
| <b>True positives</b><br>(patients correctly diagnosed with FQ resistance)              | 43 (37 to 47)                                       | 86 (75 to 93)     | 129 (112 to 140)  | 519 (9)                          | ⊕⊕⊕○ <sup>1,2,3,4</sup><br>moderate |
| <b>False negatives</b><br>(patients incorrectly classified as not having FQ resistance) | 7 (3 to 13)   | 14 (7 to 25)      | 21 (10 to 38)     |                                  |                                     |
| <b>True negatives</b><br>(patients correctly classified as not having FQ resistance)    | 937 (921 to 944)                                    | 887 (872 to 895)  | 838 (824 to 845)  | 1252 (9)                         | ⊕⊕⊕⊕ <sup>1,2,3</sup><br>high       |

|   |              |              |              |
|---|--------------|--------------|--------------|
| <b>False positives</b><br>(patients incorrectly classified as having FQ resistance) | 13 (6 to 29) | 13 (5 to 28) | 12 (5 to 26) |
|---|--------------|--------------|--------------|

Abbreviations: CI: confidence interval; DST: drug susceptibility testing; FQ: fluoroquinolone; GRADE: Grading of Recommendations, Assessment, Development and Evaluation; SLID: second-line injectable drug; TB: tuberculosis; XDR-TB: extensively drug-resistant TB.

By indirect testing, MTBDR<sub>s</sub>/sensitivity and specificity (95% CI) were 85.6% (79.2% to 90.4%) and 98.5% (95.7% to 99.5%).

\*This systematic review mainly evaluated MTBDR<sub>s</sub>/ version 1.0, which has recently been replaced with version 2.0. We considered the findings in this review to be applicable to the current version of the test.

<sup>1</sup>Eight studies used a cross-sectional study design and one study used a case-control study design.

<sup>2</sup>We used QUADAS-2 to assess risk of bias. All studies used consecutive sampling. In seven studies, the reader of the index test was blinded to results of the reference standard and in two studies information about blinding to the reference standard was not reported. Several studies used critical concentrations for the culture-based DST reference standard that differed from the concentrations recommended by the WHO. This may have lowered specificity, but this was not observed. We did not downgrade.

<sup>3</sup>We considered indirectness (applicability) from the perspective of diagnostic accuracy and had low concern. We did not downgrade.

<sup>4</sup>For individual studies, sensitivity estimates ranged from 33% to 100%. One small study with the lowest sensitivity only included three FQ-resistant patients. However, we could not explain the remaining heterogeneity by study quality or other factors. We downgraded one level for inconsistency.

## BACKGROUND

Tuberculosis (TB) is an infectious airborne disease caused by *Mycobacterium tuberculosis* bacteria. In 2014, an estimated 9.6 million people developed TB and 1.5 million people died from TB; 1.1 million among human immunodeficiency virus (HIV)-negative people and 0.4 million among HIV-positive people (WHO 2015). Although the number of TB deaths has dropped by nearly half since 1990, TB is now the most common cause of death from an infectious disease in adults, surpassing HIV/acquired immune deficiency syndrome (AIDS), which claimed 1.2 million lives. TB is a preventable and treatable disease. The World Health Organization (WHO) estimated that, since 2000, 43 million lives have been saved through effective diagnosis and treatment (WHO 2015).

TB predominantly affects the lungs (pulmonary TB) but can affect other parts of the body, such as the brain or the spine. Active TB disease is confirmed by the presence of TB bacilli grown in culture. The symptoms of pulmonary TB include a persistent cough (for at least two weeks), fever, night sweats, weight loss, chills, haemoptysis, and fatigue. TB that is drug sensitive (also referred to as drug-susceptible TB) is the most common type of TB and may be effectively treated with a standardized regimen of first-line anti-TB drugs (WHO 2015). However, TB bacilli may become drug resistant, meaning that first-line anti-TB drugs can no longer kill the bacilli. Drug resistance usually develops because of inappropriate or incorrect use of first-line drugs, but new cases are increasingly caused by person-to-person transmission (Streicher 2011; Zhao 2012).

The emergence of drug-resistant TB threatens to destabilize global TB control. There are two standardized definitions of drug-resistant TB: multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB). MDR-TB is caused by *M. tuberculosis* which, when tested microbiologically in the laboratory, is resistant to rifampicin and isoniazid. These drugs are two of the most effective and widely-used anti-TB drugs that form part of the standardized first-line regimen for drug-susceptible TB. Patients with MDR-TB are commonly treated with drugs belonging to the fluoroquinolone (FQ) and second-line injectable drug (SLID) anti-TB drug classes. The FQ drugs include ofloxacin, levofloxacin, moxifloxacin, and gatifloxacin and the SLIDs include amikacin and kanamycin (two aminoglycoside drugs) and capreomycin (a cyclic peptide drug). XDR-TB is caused by *M. tuberculosis* resistant to isoniazid, rifampicin, plus any FQ and at least one of the three SLIDs. Hence, patients with XDR-TB are resistant to both first-line and second-line drugs.

Therapy for drug-resistant TB requires treatment for more than 12 months and is toxic and expensive. A systematic review estimated only 62% (95% confidence interval (CI) 58% to 67%) of patients initiated on treatment for MDR-TB were successfully treated (defined as cured or completed treatment, Orenstein 2009). Around 10% of MDR-TB patients have XDR-TB, but this may be as high

as 30% in parts of Eastern Europe (WHO 2015). XDR-TB treatment success rates are poor (26%), with high five-year mortality (73%) in HIV-endemic settings (Pietersen 2014; WHO 2015). In South Africa in 2011, the treatment of approximately 8000 cases of drug-resistant TB, which comprised only 2.2% of the total TB burden, consumed 32% of the country's annual national TB budget of USD 218 million (Pooran 2013).

Improvements in the diagnosis of drug-resistant TB are also important for reducing transmission. In South Africa, 80% of MDR-TB is thought to be spread from person-to-person (Streicher 2011), and the same is likely true of MDR-TB and XDR-TB in China (Zhao 2012). Modelling studies have shown that, through the improvement of capacity to rapidly diagnose drug-resistant TB, patient cure rates can be improved through the earlier initiation of appropriate and effective TB treatment (Basu 2007; Basu 2009; Dowdy 2008). Importantly, this can reduce infectiousness within one to two weeks (Menzies 1997). However, the exact 'infectiousness period' for drug-resistant TB remains unclear. There is thus an urgent need for rapid tests that allow the early detection of drug resistance and the selection of appropriate drugs.

The use of conventional phenotypic culture-based drug susceptibility testing (DST) for detection of drug-resistant TB relies on the growth of TB bacteria and is therefore associated with considerable time delays (two to six months). These delays are exacerbated by the technical and infrastructure requirements of testing, the lack of standard methods for certain drugs and contamination (which cause unclear results that require repeating) (Richter 2009), as well as patient-associated difficulties, such as loss to follow-up. Once a diagnosis of MDR-TB has been established, second-line DST is typically used to diagnose second-line drug resistance. In 2015, 300,000 (of the estimated 450,000 cases) MDR-TB cases were reported, yet only 24% received second-line DST (WHO 2015).

Molecular tests for detecting drug resistance such as the GenoType® MTBDR<sub>s/l</sub> assay (henceforth called MTBDR<sub>s/l</sub>) have shown promise for the diagnosis of drug-resistant TB. These tests are rapid (around five hours), and genotypic, as they detect the presence of mutations associated with drug resistance. MTBDR<sub>s/l</sub> belongs to a category of molecular genetic tests called line probe assays. MTBDR<sub>s/l</sub> version 1.0 was the first commercial line probe assay for detection of resistance to second-line TB drugs and, since the beginning of 2016, is no longer available. MTBDR<sub>s/l</sub> version 2.0 was released in 2015. MTBDR<sub>s/l</sub> version 2.0 detects the mutations associated with FQ and SLID resistance detected by MTBDR<sub>s/l</sub> version 1.0, as well as additional mutations (described below). We have included a glossary of genetic terms in Appendix 1. The draft of this updated Cochrane review informed the WHO Guideline Development Group that met February to March 2016 to make recommendations about the use of this test. The WHO policy guidance, "The use of molecular line probe assays for the detection of resistance to second-line anti-tuberculosis drugs", was published in May 2016 (WHO 2016).

## Target condition being diagnosed

We considered the following target conditions.

1. Fluoroquinolone (FQ) resistance.
2. Second-line injectable drug (SLID) resistance.
3. XDR-TB.

## Index test(s)

The index test is MTBDR<sub>sl</sub> versions 1.0 and 2.0 (Hain Life Sciences 2015; Table 1). MTBDR<sub>sl</sub> detects specific mutations associated with resistance to the FQs (including ofloxacin, moxifloxacin, levofloxacin, and gatifloxacin) and SLIDs (including kanamycin, amikacin, and capreomycin) in *M. tuberculosis* complex species. Version 1.0 detects mutations in the *gyrA* quinolone resistance-determining region (codons 88, 90, 91, 94) and *rrs* (codons 1401, 1402, 1484). Version 2.0 additionally detects mutations in the *gyrB* quinolone resistance-determining region (codons 538, 540) and the *eis* promoter region (codons -37, -14, -12, -10, -2) (Hain Life Sciences 2015a). As mutations in these regions may cause additional resistance to the FQs or SLIDs respectively, MTBDR<sub>sl</sub> version 2.0 should have heightened sensitivity for resistance to these drug classes. Mutations in some regions (for example, the *eis* promoter region) may be responsible for causing resistance to one drug in a class more than other drugs within that class. For example, the *eis* C14T mutation is associated with kanamycin resistance in *M. tuberculosis* strains from Eastern Europe (Gikalo 2012). MTBDR<sub>sl</sub> version 1.0 also detects mutations in *embB* that may encode for resistance to ethambutol. As this is a first-line drug and was omitted from MTBDR<sub>sl</sub> version 2.0, we did not determine the accuracy for ethambutol resistance.

For the FQs, the presence of mutations in each of the genes probed by MTBDR<sub>sl</sub> has high but imperfect concordance with resistance to all drugs within that drug class. For example, a mutation in the *gyrA* gene may mean a strain is resistant to each of the FQs (for example, ofloxacin and moxifloxacin) (Sirgel 2012a). The same holds true for the *rrs* gene and the two aminoglycosides, kanamycin and amikacin (Sirgel 2012b). Evidence is mixed regarding the level

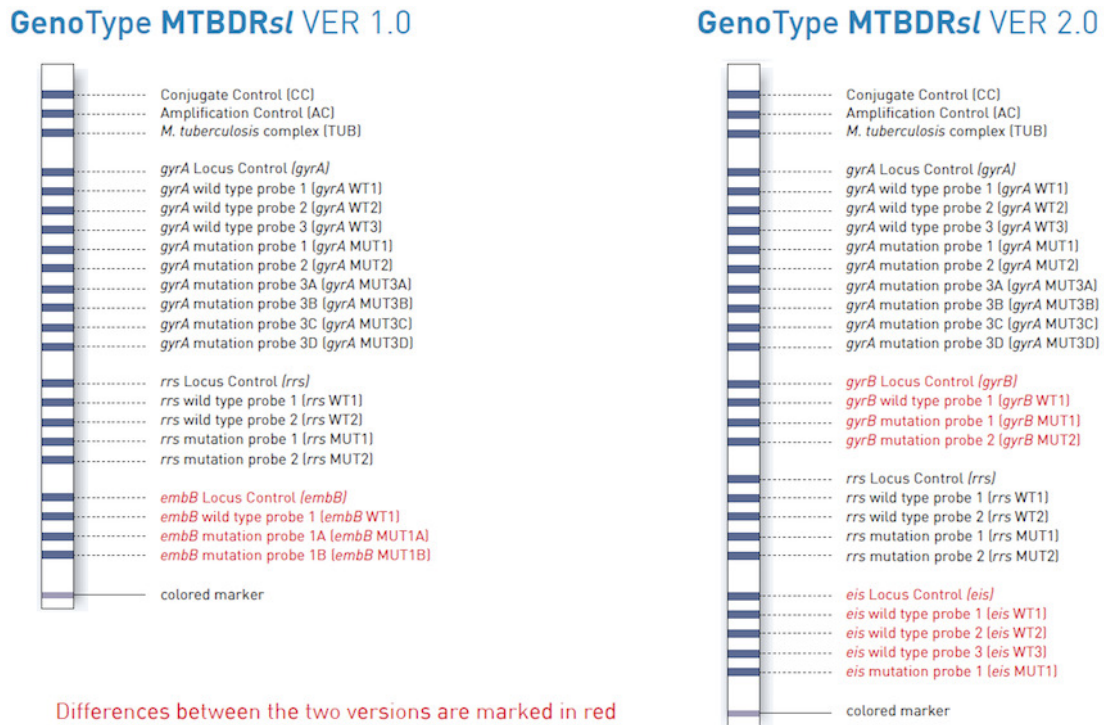
of concordance between resistance to the two aminoglycosides and capreomycin arising from mutations in the *rrs* gene. MTBDR<sub>sl</sub> reports on the presence of mutations within these genes (as well as *gyrB* and the *eis* promoter for MTBDR<sub>sl</sub> version 2.0), which are associated with resistance to a class of drugs. The presence of mutation(s) in these regions does not necessarily imply resistance to all the drugs within that class.

For MTBDR<sub>sl</sub> version 1.0, the manufacturer recommended that if the patient specimen (usually sputum) is smear-positive, the assay be performed on the specimen (direct testing) and if smear-negative, the assay be performed on the culture isolate grown from the patient specimen (indirect testing). The manufacturer states that MTBDR<sub>sl</sub> version 2.0 may be performed on a smear-positive or smear-negative specimen without the need for culture.

The assay procedure involves the following steps: 1. decontamination of the specimen; 2. isolation and amplification of DNA; 3. detection of the amplification products by reverse hybridisation; and 4. visualisation using a streptavidin-conjugated alkaline phosphatase colour reaction. The observed bands, each corresponding to a probe, can be used to determine the drug susceptibility profile of the analysed specimen. The assay can be completed in five hours.

Figure 1 shows the line probe assay strips used for MTBDR<sub>sl</sub> version 1.0 or version 2.0. A band for the detection of the *M. tuberculosis* complex (the "TUB" band) is included, as well as two internal controls (conjugate and amplification controls) and a control for each gene locus (MTBDR<sub>sl</sub> version 2.0: *gyrA*, *gyrB*, *rrs*, *eis*). The two internal controls plus each gene locus control should be positive; otherwise the assay cannot be evaluated for that particular drug. A result can be indeterminate for one locus but valid for another (on the basis of a gene-specific locus control failing). A template is supplied by the manufacturer to help read the strips, where the banding patterns are scored by eye, transcribed, and reported. In high-volume settings, the GenoScan®, an automated reader, can be incorporated to interpret the banding patterns automatically and give a suggested interpretation. If the operator agrees with the interpretation, the results are automatically uploaded, thereby reducing possible transcription errors.

**Figure 1. Comparison of version 1.0 and version 2.0 of the GenoType® MTBDR<sub>sl</sub> test (adapted from Hain Life Sciences 2015).**

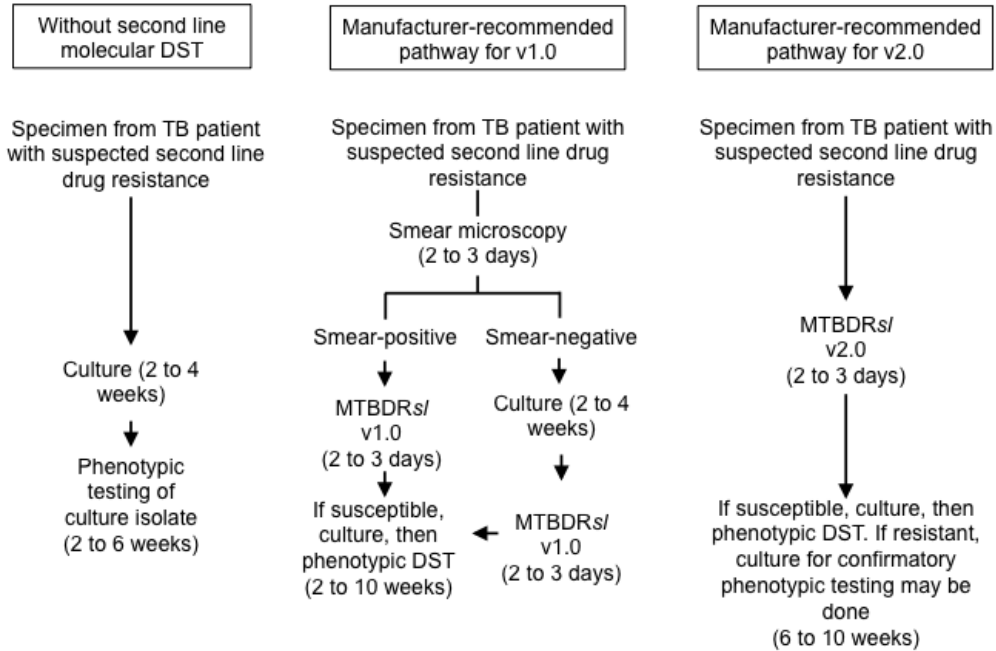


## Clinical pathway

Figure 2 illustrates the clinical pathway. Depending on the setting, DST is either performed on all patients with confirmed TB or on patients who are clinically suspected of having drug-resistant TB (for example, if the patients' symptoms have failed to improve on first-line therapy, or if they still have *M. tuberculosis* bacilli in their sputum after an extended period of treatment). DST for resistance to the second-line drugs is usually only performed if resistance to the first-line drugs is confirmed. Specifically, a patient with suspected drug-resistant TB provides a specimen (usu-

ally sputum), which is examined by smear microscopy. If smear-positive, MTBDR<sub>sl</sub> version 1.0 or version 2.0 can be performed directly on the specimen. If smear-negative, MTBDR<sub>sl</sub> version 1.0 should not be performed directly on the specimen, but rather on the culture isolate. MTBDR<sub>sl</sub> version 2.0 may be performed directly on a smear-negative specimen. A molecular test for first-line drug resistance (for example, the MTBDR<sub>plus</sub> assay) may be performed prior to testing with MTBDR<sub>sl</sub> if resistance to the first-line drugs is yet to be confirmed. Phenotypic DST may still be performed on culture-positive isolates.

**Figure 2. Clinical pathway.** A patient to be evaluated for drug-resistant tuberculosis (TB) provides a specimen (usually sputum), which is examined by smear microscopy. If smear-positive, MTBDRsl version 1.0 or version 2.0 can be performed directly on the specimen. If smear-negative, MTBDRsl version 1.0 should not be performed directly on the specimen, but rather on the culture isolate. Version 2.0 may be performed directly on a smear-negative specimen. A molecular test for first-line drug resistance (for example, the MTBDRplus assay) may be performed prior to testing with MTBDRsl if resistance to the first-line drugs is yet to be confirmed. Phenotypic (culture-based) drug susceptibility testing (DST) may still be performed on culture-positive isolates.



**Prior test(s)**

Patients who received MTBDRs/ testing may have first received smear microscopy, Xpert® MTB/RIF or another nucleic acid amplification test, and culture to diagnose TB and Xpert® MTB/RIF, MTBDRplus, or an alternative line-probe assay to detect first-line drug resistance.

**Role of index test(s)**

The role of MTBDRs/ would be as the initial test, replacing culture-based DST, for detecting second-line drug resistance.

**Alternative test(s)**

We are aware of several additional line probe assays marketed for genotypic testing for second-line drug resistance: TB Resistance Module Fluoroquinolones/Ethambutol and TB Resistance Module Kanamycin/Amikacin/Capreomycin/Streptomycin (Autoimmun Diagnostika GmbH (AID) Strassberg); MolecuTech REBA MTB-FQ®, MolecuTech REBA MTB-KM®, and MolecuTech REBA MTB-XDR® (YD diagnostics, Seoul); and NiPro LiPA FQ (NiPro Co, Osaka) (Boyle 2015). For a comprehensive review of these tests, we refer the reader to the Tuberculosis Diagnostics Technology and Market Landscape report (Boyle 2015).

## Rationale

Second-line TB drugs are used to treat patients with TB that is resistant to the most effective and widely used first-line drugs. To ensure that the most appropriate and least toxic drugs are provided to patients as quickly as possible, it is critical to know whether a patient has resistance to FQs alone, resistance to SLIDs alone, or resistance to both FQs and SLIDs (XDR-TB) as this will guide the selection of drugs. The conventional method for the diagnosis of drug resistance (culture-based DST) is vulnerable to contamination and loss of viability, meaning that the TB bacteria sometimes cannot be regrown and a culture isolate is hence not available for DST. Culture-based DST is also slow and can take several months. The resulting diagnostic delay results in unnecessary morbidity, mortality, and increased transmission, which is a major driver of new TB cases. There is a need for rapid assays to improve time-to-diagnosis and new molecular assays, such as the MTBDRs/ assay, present a promising potential solution.

## OBJECTIVES

To assess and compare the diagnostic accuracy of MTBDRs/ for: 1. fluoroquinolone resistance, 2. SLID resistance, and 3. extensively drug-resistant tuberculosis, as an indirect test on a *M. tuberculosis* isolate grown from culture or as a direct test on a patient specimen. The populations of interest were people with MDR-TB or rifampicin-resistant TB, which is considered a proxy for MDR-TB in high burden settings, WHO 2011.

## Secondary objectives

We planned to investigate heterogeneity in relation to the type of reference standard (culture-based drug susceptibility testing (DST) compared with sequencing, culture-based DST and sequencing, and culture-based DST followed by sequencing of discrepant results) and resistance to individual drugs within a drug class (for example, ofloxacin, moxifloxacin, levofloxacin, and gatifloxacin within the FQ class). We also prespecified in the protocol investigations of heterogeneity in relation to human immunodeficiency virus (HIV) status, condition of the specimen (fresh or frozen, volume of specimen), patient population (patients suspected of having MDR-TB or XDR-TB), and whether World Health Organization (WHO)-recommended critical drug concentrations were used for the culture-based DST reference standard. Subsequent to the published protocol, we added an investigation of heterogeneity in relation to microscopy smear grade.

## METHODS

## Criteria for considering studies for this review

### Types of studies

We included all studies that determined the diagnostic accuracy of the index test in comparison with a defined reference standard, including case-control designs. We only included studies from which we could extract data on true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). We excluded unpublished studies reported only in abstracts and conference proceedings.

### Participants

We included patients of any age who had rifampicin-resistant TB or MDR-TB or may have had resistance to any of the second-line TB drugs, irrespective of background burden of drug resistance and patient population.

### Index tests

The index test was MTBDRs/ version 1.0 or version 2.0.

### Target conditions

We considered the following target conditions.

1. Fluoroquinolone (FQ) resistance.
2. Second-line injectable drug (SLID) resistance.
3. XDR-TB.

### Reference standards

We included studies that used one or more of the following reference standards.

1. Culture-based drug susceptibility testing (DST): solid culture or a liquid culture.
2. Sequencing of the *gyrA* or *rrs* genes (MTBDRs/ version 1.0) or additionally the *gyrB* and *eis* promoter regions (MTBDRs/ version 2.0).
3. A composite reference standard with two components: culture-based DST and sequencing of the same samples. If a specimen was resistant according to culture-based DST or had a mutation, we classified the specimen as having the target condition. If both culture-based DST and sequencing indicated susceptibility, we classified the specimen as not having the target condition.
4. Two reference standards used sequentially: culture-based DST followed by selective testing by sequencing of samples with discrepant results. Discrepant results may be either index test positive/culture-based DST negative or index test negative/culture-based DST positive.

There are strengths and limitations to each of the reference standards. Culture-based DST is the accepted reference standard, but

it is considered to be imperfect and is dependent on the drug concentration threshold used to define resistance. Sequencing is considered to be more accurate than culture-based DST; however, this is only if it targets all known resistance-determining regions, which are not fully known for the FQs and the SLIDs. Therefore, targeted sequencing may miss mutations that cause drug resistance.

We carried out separate analyses for the different reference standards, which we have described below. In our primary analysis we used culture-based DST as the reference standard. We expected all or nearly all included studies to report results using this reference standard.

## Search methods for identification of studies

We attempted to identify all relevant studies regardless of language and publication status (published, unpublished, in press, and ongoing).

### Electronic searches

Vittoria Lutje (VL), the Information Specialist for the Cochrane Infectious Diseases Group (CIDG), performed literature searches up to 21 September 2015 without language restrictions. To identify all relevant studies, she searched the following databases using the search terms and strategy described in [Appendix 2](#): CIDG Specialized Register; MEDLINE (PubMed, 1966 to 21 September 2015); Embase OVID (1980 to 21 September 2015); Science Citation Index Expanded (SCI-EXPANDED, 1900 to 21 September 2015, Conference Proceedings Citation Index-Science (CPCI-S, 1990 to 21 September 2015), and BIOSIS Previews (1926 to 21 September 2015; all three from Web of Science); LILACS (<http://lilacs.bvsalud.org/en/>; 1982 to 21 September 2015); and SCOPUS (1995 to 21 September 2015). She also searched the ISRCTN registry (<http://isrctn.com>) and the search portal of the World Health Organization International Clinical Trials Registry Platform (ICTRP; <http://apps.who.int/trialsearch/>) to identify ongoing trials, and ProQuest Dissertations & Theses A&I to identify relevant dissertations (all websites accessed on 21 September 2015). We searched MEDION in the previous version of the review, [Theron 2014](#), but this database was unavailable in September 2015.

### Searching other resources

We reviewed reference lists of included articles and any relevant review articles identified through the above methods. We contacted researchers at FIND and other experts in the field of TB diagnostics for information on ongoing or unpublished studies.

## Data collection and analysis

### Selection of studies

Two review authors (GT and JP) independently scrutinized titles and abstracts identified by electronic literature searches to identify potentially eligible studies. We selected all citations identified as suitable during this screen for full-text review. The same two review authors then independently reviewed full-text papers for study eligibility using the predefined inclusion and exclusion criteria. For full-text articles, we resolved any discrepancies by discussion with a third review author (KRS). We maintained a list of excluded studies and their reasons for exclusion, and recorded these details in the 'Characteristics of excluded studies' table and prepared a PRISMA diagram.

### Data extraction and management

We developed a standardized data extraction form and piloted the form with two of the included studies. Based upon the pilot, we finalized the form. Then two review authors (GT and JP) independently extracted data on the following characteristics and resolved any discrepancies by discussion.

1. Details of study: first author; publication year; country where testing was performed; specimen country origin; setting (primary care laboratory, hospital laboratory, reference laboratory); study design; manner of participant selection; number of participants enrolled; number of participants for whom results available; industry sponsorship.
2. Characteristics of participants: age; HIV status; smear status; history of TB; known MDR-TB, pre-XDR-TB (defined as MDR-TB and resistance to a FQ or SLID, but not to drugs from both classes), or XDR-TB status.
3. Target conditions: resistance to FQ and SLID drug classes and XDR-TB.
4. Resistances to individual drugs: ofloxacin, moxifloxacin, levofloxacin, gatifloxacin, amikacin, kanamycin, and capreomycin.
5. Reference standards: type; percentage of patients whose reference standard was 'uninterpretable' (for example, contaminated, sequencing failed).
6. Details of specimen: type (such as expectorated sputum, induced sputum or culture isolate); condition (fresh or frozen); definition of a positive smear; type of testing (direct testing or indirect testing); smear grade (negative, scanty, 1+, 2+, 3+).
7. Details of outcomes: the number of TP, FP, FN, and TN results; number of indeterminate assay results.
8. Intra-reader and inter-reader variability.
9. Time to treatment initiation: defined as the time from specimen collection until patient starts treatment.



10. Time to diagnosis: defined as the time from specimen collection until there is an available TB result in lab or clinic, if the assay was performed in a clinic.

We assigned country income status (high, middle or low) as classified by the World Bank List of Economies (World Bank 2015). We contacted authors of primary studies for missing data or clarifications. We assigned smear grade according to the WHO definition (WHO 2014). We entered all data into a database manager (Microsoft Excel 2014).

For one study that tested the same panel of TB isolates in multiple centres, we selected one centre that provided results in the middle range (neither the best nor the worst results) (Ignatyeva 2012). One study included extrapulmonary specimens, which we excluded from the analysis (Barnard 2012). Whenever possible, we extracted data that used a single patient as the unit of analysis (one MTBDRs/ result per one specimen from one patient).

When culture-based DST was performed using more than one drug from the FQs (ofloxacin, moxifloxacin, levofloxacin, or gatifloxacin) or SLIDs (amikacin, kanamycin or capreomycin), we extracted TP, FP, FN, and TN values for each drug and for each class overall. If the reference standard indicated resistance for at least one drug in that class, we classified the sample as resistant to that class of drugs. We did not require reference standard DST results for all drugs in a class in order to classify a sample as resistant or susceptible.

In the 2 x 2 tables of TP, FP, FN, and TN, we based the results of the index test on categorical assay results defined by the visual readout of the MTBDRs/ strip.

#### **Possible results for the GenoType® MTBDRs/ assay (as defined by the product manual)**

1. Sensitive to either FQs or SLIDs (referred to as 'aminoglycosides/cyclic peptides'), or both (conjugation and amplification bands present; *Mycobacterium tuberculosis* complex-specific control (TUB) band present; gene locus band present; all wild type (*wt*) bands for each gene present; no mutation bands present). In the case of susceptibility to both drug classes, the test would indicate susceptibility for each, rather than having a single composite readout specifying XDR-TB.

2. Resistant to either FQs or SLIDs, or both (conjugation and amplification bands present; TUB band present; gene locus band present; all, none or some *wt* bands for each gene present; all, none or some mutation bands present with similar intensity to amplification control). In the case of resistance to both drug classes, the test would indicate resistance for each, rather than having a composite readout.

3. Indeterminate (faint bands) or no result (no conjugation or amplification bands present, no locus band present for the gene of interest).

4. No TB (negative for MTB complex irrespective of locus control band).

5. No result (failure of any one of the control bands, as well as the TUB band).

No studies reported on the number of 'no TB' or 'no result' results obtained from MTBDRs/, therefore we only extracted the number and percentage of 'indeterminate' results.

#### **Assignment of results to the fluoroquinolones, second-line injectable drugs, or both categories**

We were able to report accuracy estimates for individual drugs within the drug classes when that drug was used as part of the culture-based DST reference standard. For determining resistance to the drug class, we used the following approach. For a culture-based DST reference standard, one study might have used detection of ofloxacin resistance and another study, detection of moxifloxacin resistance to confirm an MTBDRs/ FQ-resistant result. In such a scenario, if culture-based DST is positive for resistance to one of the drugs in the drug class and the MTBDRs/ result is concordant, we classified the index test result as a true positive for resistance to the FQs. We adopted the same approach for the SLIDs.

For sequencing as a reference standard, if the index test reported resistance to FQs and the presence of mutations known to be associated with drug resistance to the FQs was confirmed in the same regions of the genome targeted by MTBDRs/, we recorded the test result as concordant and classified the index test as a true positive for resistance to the FQs. We adopted the same approach for the SLIDs.

#### **Assessment of methodological quality**

We appraised the quality of the included studies with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool (Whiting 2011; Appendix 3). QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing. We assessed all domains for risk of bias and the first three domains for concerns regarding applicability. We used signalling questions in each domain to form judgments about the risk of bias. One review author (GT) piloted the tool with two included studies and finalized the tool based on experience gained from the pilot testing. Three review authors (GT, JP, and KRS) then independently assessed the methodological quality of included studies with the finalized tool and finalized judgments by discussion.

#### **Statistical analysis and data synthesis**

We performed descriptive analyses for key variables (such as country income status) of the primary studies using Stata (Stata 2015), and displayed key study characteristics in the 'Characteristics of included studies' table.

We analysed data separately for MTBDRs/ version 1.0 and version 2.0. We used the reference standard culture-based DST in our primary analyses. We stratified these analyses first by target condition and second by type of MTBDRs/ testing (indirect testing or direct

testing). Within each stratum (for example, FQ resistance by indirect testing), we plotted estimates of the studies' observed sensitivities and specificities in forest plots with 95% confidence intervals (CIs) and in receiver operating characteristic (ROC) space using Review Manager (RevMan) (RevMan 2014). Where adequate data were available, we combined data using meta-analysis. We performed most meta-analyses by fitting the bivariate random-effects model (Macaskill 2010; Reitsma 2005), using Stata with the `metandi` and `xtmelogit` commands (Stata 2015). In situations with few studies, we performed meta-analysis where appropriate by reducing the bivariate model to two univariate random-effects logistic regression models by assuming no correlation between sensitivity and specificity. When we observed little or no heterogeneity on forest plots and summary receiver operating characteristic (SROC) plots, we further simplified the models into fixed-effect models by eliminating the random-effects parameters for sensitivity or specificity, or both sensitivity and specificity (Takwoingi 2015).

We compared results from studies of direct testing with results from studies of indirect testing by adding a covariate for the type of testing to the model. We assessed the significance of the differences in sensitivity and specificity estimates between studies in which MTBDRs/ was performed by direct testing or indirect testing by a likelihood ratio test comparing models with and without covariate terms. Where data were sufficient, we performed comparative analyses including only those studies that made direct comparisons between test evaluations with the same participants. Otherwise, we included all studies with available data. Comparative studies are preferred to non-comparative studies when deriving evidence of diagnostic test accuracy (Takwoingi 2013).

### **Approach to uninterpretable (indeterminate) MTBDRs/ results**

We excluded indeterminate test results from the analyses for determination of sensitivity and specificity. We determined the percentage of indeterminate MTBDRs/ results among the primary studies for each target condition and summarized these findings separately for indirect and direct testing when available, according to culture-based reference standard.

### **Investigations of heterogeneity**

Within each stratum (for example, SLID resistance, indirect testing), we investigated heterogeneity through visual examination of forest plots of sensitivity and specificity. Then, if sufficient studies were available, we explored the possible influence of the following prespecified categorical covariates: reference standard (culture, sequencing, culture and sequencing, culture followed by sequencing); resistance to the following drugs: ofloxacin, moxifloxacin, levofloxacin, gatifloxacin, amikacin, kanamycin, and capreomycin (we excluded resistance to ciprofloxacin because this drug is infrequently used in DST); and drug concentration used for culture-

based DST (WHO-recommended critical concentration used or a different concentration used). In addition, for this updated review, we added an investigation of heterogeneity in relation to microscopy smear grade. We assessed the significance of the difference in test accuracy (for example, between studies using culture versus those using sequencing as the reference standard) by a likelihood ratio test comparing models with and without covariate terms.

We had planned to investigate the effect of HIV status, the condition of the specimen (fresh or frozen), sample volume, and population (patients thought to have MDR-TB or XDR-TB) on summary estimates of sensitivity and specificity in a meta-regression analysis by adding covariate terms to the bivariate model. However, there were insufficient data for these additional analyses.

### **Sensitivity analyses**

For our primary analyses using the culture-based DST reference standard, we performed sensitivity analyses for QUADAS-2 items to explore whether the accuracy estimates were robust with respect to the methodological quality of the studies. We included the following signalling questions.

1. Was a consecutive or random sample of patients/specimens enrolled?
2. Was a case-control design avoided?
3. Were the index test results interpreted without knowledge of the results of the reference standard?
4. Was the test applied in the manner recommended by the manufacturer (index test domain, low concern about applicability)?

### **Assessment of reporting bias**

We did not undertake a formal assessment of publication bias of data included in this review using methods such as funnel plots or regression tests because such techniques have been unhelpful for determining publication bias within diagnostic test accuracy studies (Macaskill 2010).

### **Other analyses**

We had intended to summarize data on intra- and inter-reader variability; however inter-reader variability was the only information described in the included studies. We had also intended to summarize two patient outcomes, time-to-diagnosis, and time-to-treatment initiation; however time-to-diagnosis was the only outcome described in the included studies.

### **Assessment of the quality of evidence (certainty of the evidence)**

We assessed the quality of evidence (also called certainty of the evidence or confidence in effect estimates) using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Balshem 2011; GRADE 2013), and GRADEpro Guideline Development Tool (GDT) software (GRADEpro GDT 2015). In the context of a systematic review, the ratings of the certainty of the evidence reflect the extent of our confidence that the estimates of the effect (including test accuracy and associations) are correct. As recommended, we rated the quality of evidence as either high (not downgraded), moderate (downgraded by one level), low (downgraded by two levels), or very low (downgraded by more than two levels) for five domains: risk of bias, indirectness, inconsistency, imprecision, and publication bias. For each outcome, the quality of evidence started as high when there were high quality observational studies (cross-sectional or cohort studies) that enrolled participants with diagnostic uncertainty. If we found a reason for downgrading, we used our judgement to classify the reason as either serious (downgraded by one level) or very serious (downgraded by two levels).

Three review authors (GT, JP, and KRS) discussed judgments and applied GRADE in the following way.

1. Risk of bias: we used QUADAS-2 to assess risk of bias.
2. Indirectness: we considered indirectness from the perspective of test accuracy. We used QUADAS-2 for concerns of applicability and looked for important differences between the populations studied (for example, in the spectrum of disease), the setting, and the review questions.
3. Inconsistency: GRADE recommends downgrading for unexplained inconsistency in sensitivity and specificity estimates. We carried out prespecified analyses to investigate potential sources of heterogeneity and did not downgrade the quality of the evidence when we felt we could explain inconsistency in the accuracy estimates.

4. Imprecision: we considered a precise estimate to be one that would allow a clinically meaningful decision. We considered the width of the CI, and asked ourselves, “Would we make a different decision if the lower or upper boundary of the CI represented the truth?” In addition, we worked out projected ranges for TP, FN, TN, and FP for a given TB prevalence and made judgements on imprecision from these calculations.

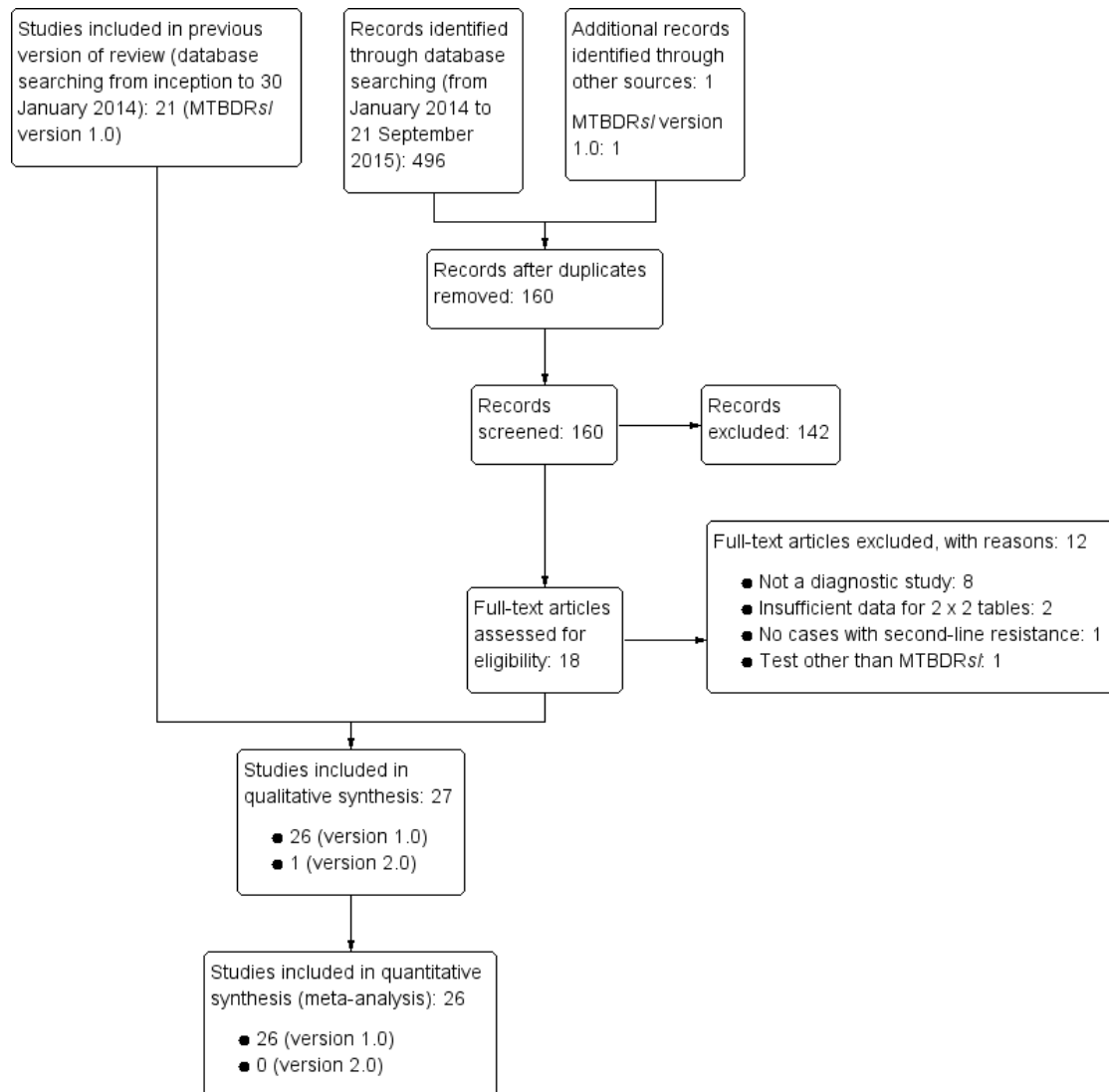
5. Publication bias: we rated publication bias as undetected (not serious) because of the comprehensiveness of the literature search and extensive outreach to TB researchers to identify studies.

## RESULTS

### Results of the search

We identified 27 unique studies that met the inclusion criteria of this review. All studies but two (Fan 2011, written in Chinese and Chikamatsu 2012, written in Japanese) were written in English. For MTBDRs/version 1.0, we included 26 studies: 21 studies from Theron 2014, the original Cochrane review (Ajbani 2012; Barnard 2012; Brossier 2010a; Chikamatsu 2012; Fan 2011; Ferro 2013; Hillemann 2009; Huang 2011; Ignatyeva 2012; Jin 2013; Kiet 2010; Kontsevaya 2011; Kontsevaya 2013; Lacoma 2012; Lopez-Roa 2012; Miotto 2012; Said 2012; Surcouf 2011; Tukvadze 2014; van Ingen 2010; Zivanovic 2012) and five new studies (Catanzaro 2015; Kambli 2015a; Kambli 2015b; Simons 2015; Tomasicchio 2016). For MTBDRs/version 2.0, we included one study (Tagliani 2015). Figure 3 shows the flow of studies in the review. We recorded the excluded studies and the reasons for their exclusion in the 'Characteristics of excluded studies' table.

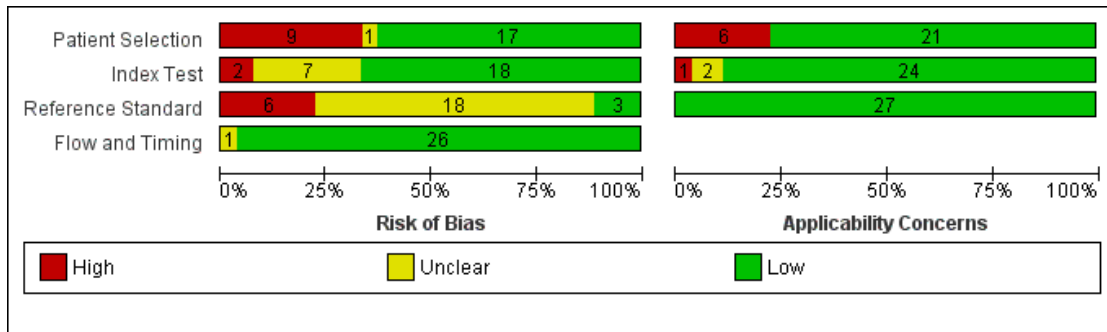
**Figure 3. Study flow diagram for searches run from January 2014 to 21 September 2015.**



### Methodological quality of included studies

Figure 4 and Figure 5 show risk of bias and applicability concerns for each of the 27 included studies. In the patient selection domain, we judged that 17 studies (63%) had low risk of bias (Ajbani 2012; Barnard 2012; Catanzaro 2015; Ferro 2013; Huang 2011 Jin 2013; Kambli 2015a; Kambli 2015b; Kontsevaya 2011; Kontsevaya 2013; Lacombe 2012; Said 2012; Simons 2015; Surcouf 2011; Tagliani 2015; Tukvadze 2014; Zivanovic 2012). We judged that one study (4%) had unclear risk of bias because the manner of patient selection was unclear (Chikamatsu 2012). We judged that nine studies had high risk of bias for the following reasons.

**Figure 4. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.**



**Figure 5. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study. For direct testing of smear-positive specimens, Miotto 2012 and Tomasicchio 2016 had low risk of bias in the patient selection domain.**

|                  | Risk of Bias      |            |                    |                 | Applicability Concerns |            |                    |
|------------------|-------------------|------------|--------------------|-----------------|------------------------|------------|--------------------|
|                  | Patient Selection | Index Test | Reference Standard | Flow and Timing | Patient Selection      | Index Test | Reference Standard |
| Ajbani 2012      | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Barnard 2012     | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Brossier 2010a   | -                 | ?          | -                  | +               | -                      | ?          | +                  |
| Catanzaro 2015   | +                 | ?          | ?                  | +               | +                      | -          | +                  |
| Chikamatsu 2012  | ?                 | -          | -                  | +               | +                      | +          | +                  |
| Fan 2011         | -                 | ?          | ?                  | +               | +                      | +          | +                  |
| Ferro 2013       | +                 | ?          | ?                  | ?               | +                      | +          | +                  |
| Hillemann 2009   | -                 | +          | ?                  | +               | -                      | +          | +                  |
| Huang 2011       | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Ignatyeva 2012   | -                 | +          | ?                  | +               | -                      | +          | +                  |
| Jin 2013         | +                 | +          | -                  | +               | +                      | +          | +                  |
| Kambli 2015a     | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Kambli 2015b     | +                 | +          | +                  | +               | +                      | +          | +                  |
| Kiet 2010        | -                 | -          | -                  | +               | -                      | +          | +                  |
| Kontsevaya 2011  | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Kontsevaya 2013  | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Lacoma 2012      | +                 | +          | -                  | +               | +                      | +          | +                  |
| Lopez-Roa 2012   | -                 | ?          | +                  | +               | +                      | +          | +                  |
| Miotto 2012      | -                 | +          | ?                  | +               | -                      | +          | +                  |
| Said 2012        | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Simons 2015      | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Surcouf 2011     | +                 | ?          | ?                  | +               | +                      | +          | +                  |
| Tagliani 2015    | +                 | +          | ?                  | +               | +                      | +          | +                  |
| Tomasicchio 2016 | -                 | +          | +                  | +               | +                      | +          | +                  |
| Tukvadze 2014    | +                 | ?          | ?                  | +               | +                      | ?          | +                  |
| van Ingen 2010   | -                 | +          | -                  | +               | -                      | +          | +                  |
| Zivanovic 2012   | +                 | +          | ?                  | +               | +                      | +          | +                  |

● High      ? Unclear      + Low

1. There was a case-control design (four studies: [Brossier 2010a](#); [Hillemann 2009](#); [Ignatyeva 2012](#); [Kiet 2010](#)).
2. There was a cross-sectional design for samples used in direct testing and case-control design for samples used in indirect testing (two studies: [Miotto 2012](#); [Tomasicchio 2016](#)).
3. Enrolment was by convenience (three studies: [Fan 2011](#); [Lopez-Roa 2012](#); [van Ingen 2010](#)).

Regarding applicability (patient characteristics and setting), we judged that 21 studies (78%) had low concern and six studies had high concern ([Brossier 2010a](#); [Hillemann 2009](#); [Ignatyeva 2012](#); [Kiet 2010](#); [Miotto 2012](#); [van Ingen 2010](#)). In the index test domain, we judged that 18 studies (67%) had low risk of bias ([Ajvani 2012](#); [Barnard 2012](#); [Hillemann 2009](#); [Huang 2011](#); [Ignatyeva 2012](#); [Jin 2013](#); [Kambli 2015a](#); [Kambli 2015b](#); [Kontsevaya 2011](#); [Kontsevaya 2013](#); [Lacoma 2012](#); [Miotto 2012](#); [Said 2012](#); [Simons 2015](#); [Tagliani 2015](#); [Tomasicchio 2016](#); [van Ingen 2010](#); [Zivanovic 2012](#)); seven (26%) studies had unclear risk of bias because information about blinding was not reported ([Brossier 2010a](#); [Catanzaro 2015](#); [Fan 2011](#); [Ferro 2013](#); [Lopez-Roa 2012](#); [Surcouf 2011](#); [Tukvadze 2014](#)); and two studies had high risk of bias because the index test results were not interpreted without knowledge of the reference standard results ([Chikamatsu 2012](#); [Kiet 2010](#)). Regarding applicability of the index test, we judged that 24 studies (89%) had low concern, one study (4%) had high concern ([Catanzaro 2015](#)), and two studies had unclear concern ([Brossier 2010a](#); [Tukvadze 2014](#)).

In the reference standard domain, we judged that only three studies (11%) had low risk of bias ([Kambli 2015b](#); [Lopez-Roa 2012](#); [Tomasicchio 2016](#)) because these studies used the WHO-recommended critical concentration for each drug included in the culture-based drug susceptibility testing (DST) reference standard; 18 studies (67%) had unclear risk of bias ([Ajvani 2012](#); [Barnard 2012](#); [Catanzaro 2015](#); [Hillemann 2009](#); [Huang 2011](#); [Ignatyeva 2012](#); [Fan 2011](#); [Ferro 2013](#); [Kambli 2015a](#); [Kontsevaya 2011](#); [Kontsevaya 2013](#); [Miotto 2012](#); [Said 2012](#); [Simons 2015](#); [Surcouf 2011](#); [Tagliani 2015](#); [Tukvadze 2014](#); [Zivanovic 2012](#)), because these studies used the World Health Organization (WHO)-recommended critical concentration for some, but not all of the drugs included in the culture-based DST reference standard; and six studies had high risk of bias ([Brossier 2010a](#); [Chikamatsu 2012](#); [Jin 2013](#); [Kiet 2010](#); [Lacoma 2012](#); [van Ingen 2010](#)) because these studies did not use WHO-recommended critical concentrations for any of the drugs included in the culture-based DST reference standard. Regarding applicability of the reference standard, we judged that all studies had low concern. In the flow and timing

domain, we judged that 26 studies (96%) had low risk of bias and one study had unclear risk of bias because we could not account for all patients in the analyses ([Ferro 2013](#)).

We noted industry involvement in eight studies (30%) and this included the following.

1. Donation of MTBDRs/ (five studies: [Ferro 2013](#); [Hillemann 2009](#); [Miotto 2012](#); [Surcouf 2011](#); [Tagliani 2015](#)).
2. Preferred pricing of MTBDRs/ (one study: [Barnard 2012](#)).
3. Financial support for non-test related study costs (one study: [Said 2012](#)).
4. Involvement in the design, analysis or manuscript production (one study: [Ajvani 2012](#)).

## Findings

We presented key characteristics of the 27 included studies in the 'Characteristics of included studies' table. Of 26 studies reporting specimen country origin, 15 studies (58%) evaluated patients from low- or middle-income countries. The median sample size (interquartile range) was 95 (44 to 176).

### MTBDRs/ version 1.0

[Table 2](#) (indirect testing) and [Table 3](#) (direct testing) show the number of studies that evaluated MTBDRs/ version 1.0, according to the reference standard and target condition. We did not identify any studies that evaluated the accuracy of MTBDRs/ for gatifloxacin resistance.

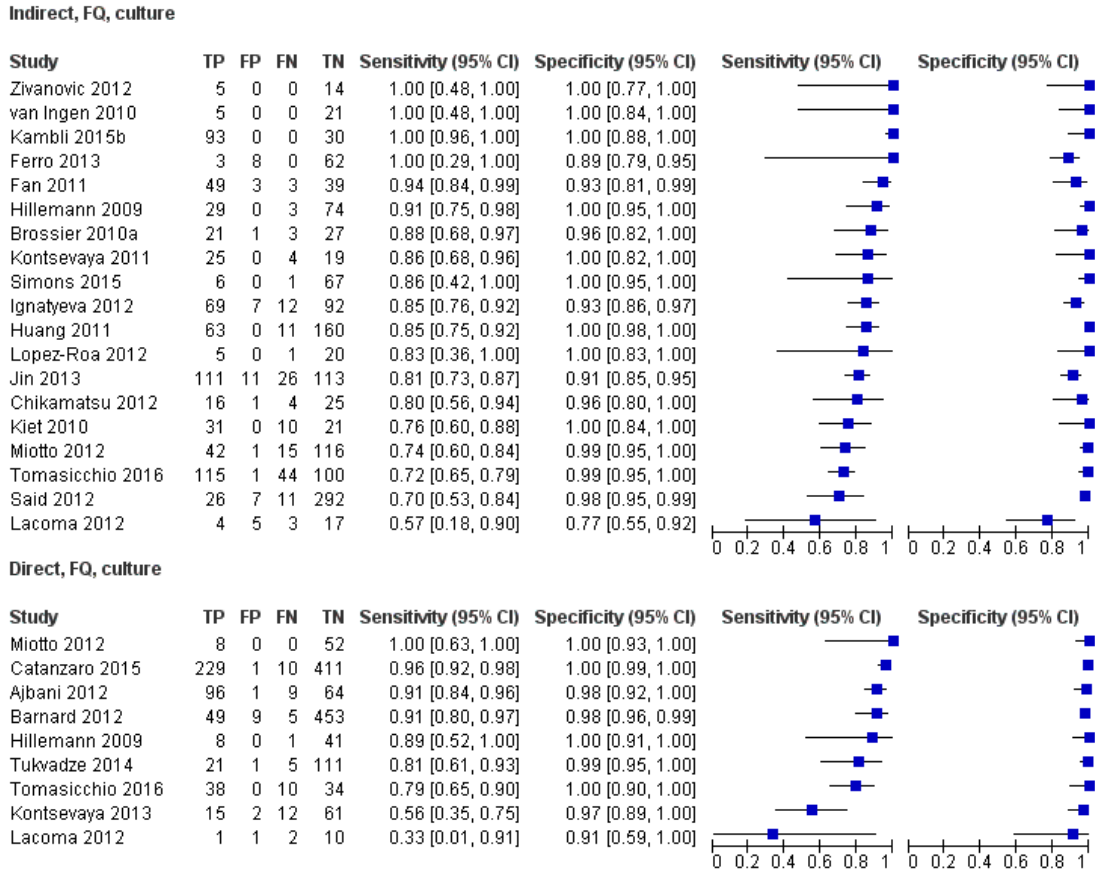
## I. Fluoroquinolone resistance detection

### A. Estimates of the diagnostic accuracy of MTBDRs/ using culture-based DST as a reference standard

#### 1. Indirect testing

Nineteen studies (2223 participants, 869 (39.1%) confirmed cases of fluoroquinolone (FQ)-resistant TB) evaluated MTBDRs/ by indirect testing for detection of FQ resistance ([Figure 6](#)). Sensitivity estimates ranged from 57% to 100% and specificity estimates ranged from 77% to 100%. The pooled sensitivity and specificity (95% CI) were 85.6% (79.2% to 90.4%) and 98.5% (95.7% to 99.5%) ([Table 4](#)).

**Figure 6. Forest plots of MTBDRsI sensitivity and specificity for fluoroquinolone (FQ) resistance, the test performed indirectly or directly against culture-based drug susceptibility (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**



## 2. Direct testing

Nine studies (1771 participants, 519 (29.3%) confirmed cases of FQ-resistant TB) evaluated MTBDRsI by direct testing for detection of FQ resistance, (Figure 6). Sensitivity estimates ranged from 33% to 100% and specificity estimates ranged from 91% to 100%. The pooled sensitivity and specificity (95% CI) were 86.2% (74.6% to 93.0%) and 98.6% (96.9% to 99.4%) (Table 4).

## 3. Comparison of indirect versus direct testing

### (a) Diagnostic accuracy

Based on analysis of all data, there was no evidence of a statistically significant difference in MTBDRsI version 1.0 accuracy for FQ resistance between indirect and direct testing (smear-positive specimen) when using culture-based DST as a reference standard (P values for differences in sensitivity and specificity of 0.932 and 0.333, respectively) (Table 4). Direct within-study comparisons were not possible because no studies performed MTBDRsI testing on specimens and isolates from the same patients.



### **(b) Indeterminate rates**

MTBDRs/l version 1.0: for indirect testing for culture-confirmed resistance to FQs, of 14 studies that reported indeterminate MTBDRs/l results, eight of 2065 results (0.4%) were indeterminate (seven culture-DST resistant and one culture-DST sensitive). For direct testing on a smear-positive specimen, of nine studies that reported indeterminate MTBDRs/l results, 147 of 2059 results (7.1%) were indeterminate (68 culture-DST resistant, 73 susceptible, and six whose culture phenotypic status was unknown). The indeterminate rates for direct testing for each smear-grade (smear-negative, scanty, 1+, 2+, 3+) were 61/190 (32.1%), 28/133 (21.1%), 35/272 (12.9%), 19/211 (9.0%), and 44/388 (11%), respectively.

## **B. Investigations of heterogeneity**

### **1. Indirect testing**

#### **(a) Individual drugs within the drug class**

We present accuracy estimates for MTBDRs/l by indirect testing for detection of resistance to ofloxacin, moxifloxacin, and levofloxacin in [Table 5](#) and [Appendix 4](#).

For detection of ofloxacin resistance by indirect testing, sensitivity estimates ranged from 70% to 100% and specificity estimates ranged from 91% to 100%. The pooled sensitivity and specificity (95% CI) were 85.2% (78.5% to 90.1%) and 98.5% (95.6% to 99.5%), (13 studies, 1927 participants).

For detection of moxifloxacin resistance by indirect testing, sensitivity estimates ranged from 57% to 100% and specificity estimates from 77% to 100%. The pooled sensitivity and specificity (95% CI) were 94.0% (82.2% to 98.1%) and 96.6% (85.2% to 99.3%), (six studies, 419 participants).

We identified two studies for detection of levofloxacin resistance by indirect testing. Sensitivity and specificity estimates (95% CI) were 80% (56% to 94%) and 96% (80% to 100%) for [Chikamatsu 2012](#), and 100% (96% to 100%) and 100% (88% to 100%) for [Kambli 2015b](#). We did not determine summary estimates because there were only two studies and the sensitivity was variable.

#### **(b) Drug concentration used in culture-based DST**

[Appendix 5](#) shows ofloxacin, levofloxacin, and moxifloxacin drug concentrations used in culture-based DST in relation to the WHO-recommended critical concentrations.

Ofloxacin: eight studies used the WHO-recommended critical concentration of ofloxacin ([Fan 2011](#); [Huang 2011](#); [Ignatyeva 2012](#); [Kambli 2015a](#); [Lopez-Roa 2012](#); [Miotto 2012](#); [Said 2012](#); [Tomasichio 2016](#)), whereas two did not ([Brossier 2010a](#); [Kiet](#)

[2010](#)). Two studies used two different types of culture medium but only used the WHO-recommended critical concentration of ofloxacin for one type of culture medium ([Hillemann 2009](#); [Zivanovic 2012](#)). [Jin 2013](#) used a non-WHO recommended concentration for one type of culture medium and no recommended concentration existed for the other culture type. There was no evidence of a statistically significant difference in MTBDRs/l version 1.0 accuracy for ofloxacin resistance between studies that did or did not use the WHO-recommended critical concentration (P values for differences in sensitivity and specificity of 0.960 and 0.904, respectively).

Moxifloxacin: [Ferro 2013](#) used the WHO-recommended critical concentration for low-level moxifloxacin resistance whereas [Lacoma 2012](#) used the concentration recommended for high-level resistance. Four studies did not use the recommended critical concentration of moxifloxacin ([Fan 2011](#); [Kambli 2015a](#); [Simons 2015](#); [van Ingen 2010](#)).

Levofloxacin: one study used the WHO-recommended critical concentration ([Kambli 2015b](#)), and one study ([Chikamatsu 2012](#)) used a culture media type for which a recommended concentration does not exist.

Comparisons between accuracy estimates for moxifloxacin and levofloxacin according to critical concentration were not possible given the small number of studies.

#### **(c) Type of reference standard**

We present MTBDRs/l accuracy estimates for detection of FQ resistance against different reference standards in [Table 6](#) and [Appendix 6](#).

##### **Reference standard is sequencing**

Using sequencing, MTBDRs/l version 1.0 sensitivity estimates ranged from 85% to 100% and specificity estimates ranged from 92% to 100%. Based on comparative studies, the pooled sensitivity and specificity (95% CI) were 99.3% (81.2% to 100.0%) and 99.3% (90.8% to 100.0%), (six studies, 873 participants). There was evidence of a significantly higher sensitivity using sequencing as the reference standard compared with culture-based DST (P value < 0.001), but not specificity (P value of 0.735).

##### **Reference standard is culture-based DST and sequencing (i.e. both investigations performed on all isolates)**

Using culture and sequencing, MTBDRs/l version 1.0 sensitivity estimates ranged from 74% to 91% and specificity estimates ranged from 99% to 100%. Based on comparative studies, the pooled sensitivity and specificity (95% CI) were 82.0% (77.7% to 85.6%) and 99.8% (98.5% to 100.0%), (seven studies, 1211 participants). There was no evidence of a statistically significant difference using both culture-based DST and sequencing as the reference standard compared with culture-based DST (P values for differences in sensitivity and specificity of 0.664 and 0.070, respectively).

Reference standard is culture-based DST followed by sequencing of discrepant index test-culture-based DST results

Using sequencing of discrepant results, MTBDRs/version 1.0 sensitivity estimates ranged from 73% to 100% and specificity estimates ranged from 94% to 100%. The pooled sensitivity and specificity (95% CI) were 83.7% (74.2% to 90.8%) and 99.7% (98.4% to 100.0%), (three studies, 427 participants). We did not perform within study comparisons between accuracy estimates using this reference standard and culture-based DST given the small number of studies in the former group.

## 2. Direct testing

### (a) Individual drugs within the drug class

We present accuracy estimates for MTBDRs/version 1.0 by direct testing for detection of resistance to ofloxacin and moxifloxacin in Table 5 and Appendix 7. We did not identify any studies that performed direct testing for detection for levofloxacin resistance. For detection of ofloxacin resistance by direct testing, MTBDRs/version 1.0 sensitivity estimates ranged from 79% to 100% and specificity estimates ranged from 98% to 100%, (seven studies, 1667 participants). The pooled sensitivity and specificity (95% CI) were 90.9% (84.7% to 94.7%) and 98.9% (97.8% to 99.4%). Based on all data, there was no evidence of a statistically significant difference between indirect and direct testing for ofloxacin resistance (P values for differences in sensitivity and specificity of 0.180 and 0.161, respectively).

For detection of moxifloxacin resistance by direct testing, Catanzaro 2015 reported MTBDRs/version 1.0 sensitivity of 96% and specificity of 99% and Ajbani 2012 reported a sensitivity of

92% and specificity of 98%. The pooled sensitivity and specificity (95% CI) were 95.0% (92.1% to 96.9%) and 99.0% (95% CI 97.5% to 99.6%), (two studies, 821 participants). Based on all data, there was no evidence of a statistically significant difference between indirect and direct testing for moxifloxacin resistance (P values for differences in sensitivity and specificity of 0.820 and 0.365, respectively).

### (b) Drug concentration used in culture-based DST

Appendix 5 shows ofloxacin and moxifloxacin drug concentrations used in culture-based DST in relation to the WHO-recommended critical concentrations.

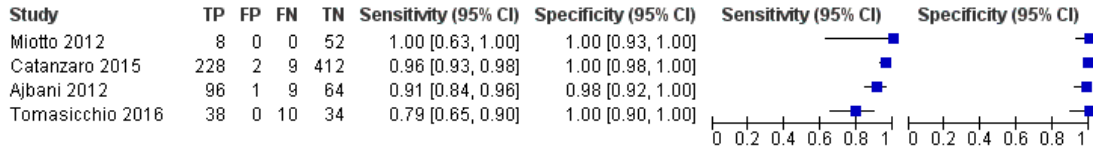
Ofloxacin: five studies used the WHO-recommended critical concentration of ofloxacin (Ajbani 2012; Barnard 2012; Catanzaro 2015; Miotto 2012; Tomasicchio 2016), whereas one study did not (Tukvadze 2014). Hillemann 2009, which used two types of culture medium, used the recommended concentration for one culture type and a non-recommended concentration for the other. Moxifloxacin: neither study used the WHO-recommended critical concentration of moxifloxacin (Ajbani 2012; Catanzaro 2015). Comparisons between accuracy estimates according to whether or not WHO-recommended critical drug concentrations were used for culture-based reference testing were not possible given the small number of studies.

### (c) Stratification by smear grade

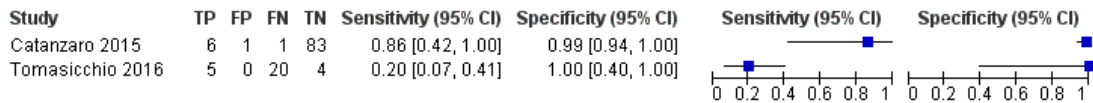
There were limited data on MTBDRs/version 1.0 accuracy for individual FQ drugs by smear grade. Figure 7 presents the forest plots for ofloxacin resistance by smear grade.

**Figure 7. Forest plots of MTBDRsl sensitivity and specificity for ofloxacin resistance by smear grade, using culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

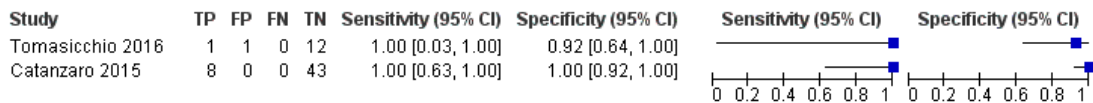
**Ofloxacin, smear positive**



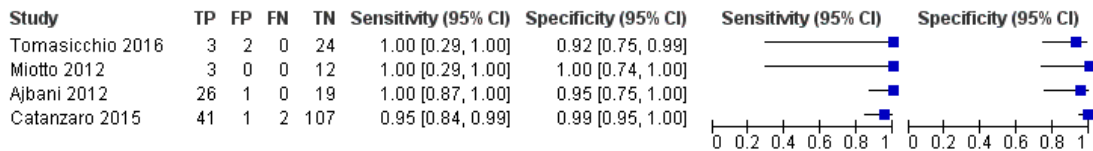
**Ofloxacin, smear negative**



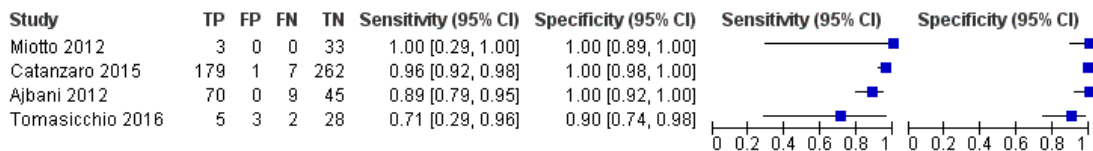
**Ofloxacin, smear grade = scanty**



**Ofloxacin, smear grade = 1+**



**Ofloxacin, smear grade ≥ 2+**



**(d) Type of reference standard**

Reference standard is sequencing

No studies performed direct MTBDRsl version 1.0 testing and used sequencing as a reference standard.

Reference standard is culture-based DST and sequencing (i.e. both investigations performed on all isolates)

No studies performed direct MTBDRsl version 1.0 testing and used both culture-based DST and sequencing (performed on all isolates) as a reference standard.

Reference standard is culture-based DST followed by sequencing of discrepant index test-culture-based DST results

Two studies (685 participants) reported MTBDRsl version 1.0 sensitivity and specificity when performed directly for the detection of resistance to FQs, with culture-based DST and sequencing performed only on discrepant results as a reference standard. The reported sensitivity and specificity (95% CI) were 91% (84% to 96%) and 98% (92% and 100%) for [Ajbani 2012](#), and 96% (88% to 100%) and 99% (98% to 100%) for [Barnard 2012](#).

**II. Second-line injectable drug resistance detection**

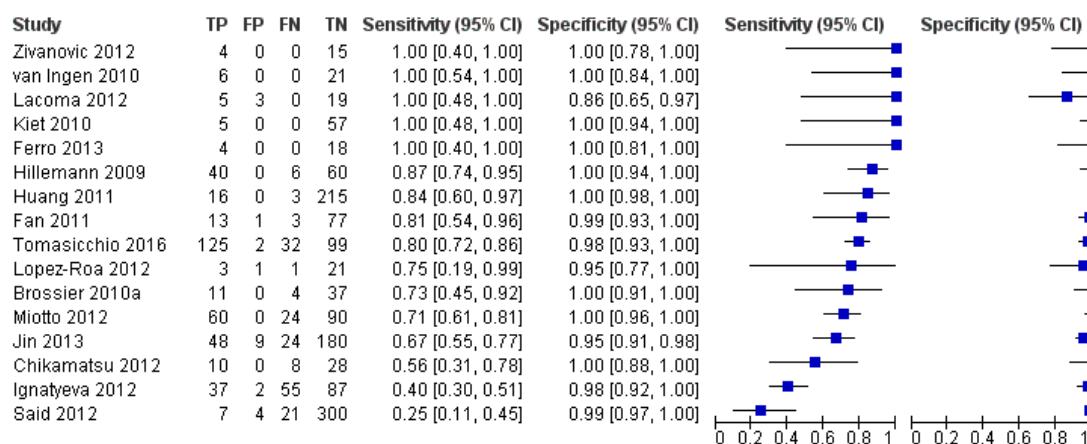
**A. Estimates of the diagnostic accuracy of MTBDRs<sub>l</sub> using culture-based DST as a reference standard**

**1. Indirect testing**

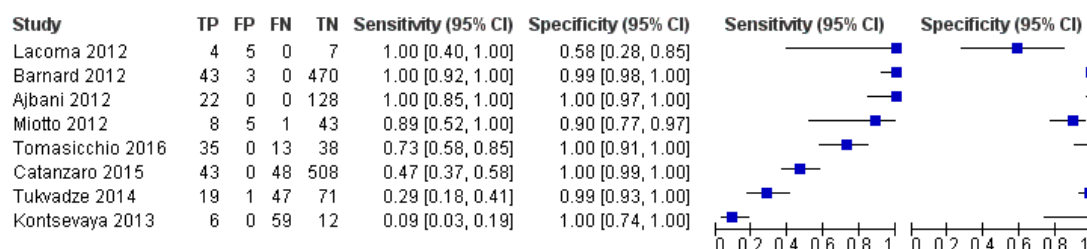
Sixteen studies (1921 participants, 575 (29.9%) confirmed cases of second-line injectable drug (SLID)-resistant TB) evaluated MTBDRs<sub>l</sub> version 1.0 by indirect testing for detection of SLID resistance (Figure 8). Sensitivity estimates ranged from 25% to 100% and specificity estimates ranged from 86% to 100%. The pooled sensitivity and specificity (95% CI) were 76.5% (63.3% to 86.0%) and 99.1% (97.3% to 99.7%) (Table 4).

**Figure 8. Forest plots of MTBDRs<sub>l</sub> sensitivity and specificity for SLID resistance, the test performed indirectly or directly against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

**Indirect, SLID, culture**



**Direct, SLID, culture**



## 2. Direct testing

Eight studies (1639 participants, 348 (21.2%) confirmed cases of FQ-resistant TB) evaluated MTBDRs<sub>1</sub>/version 1.0 by direct testing for detection of SLID resistance (Figure 8). For individual studies, sensitivity estimates ranged from 9% to 100% and specificity estimates ranged from 58% to 100%. The pooled sensitivity and specificity (95% CI) were 87.0% (38.1% to 98.6%) and 99.5% (93.6% to 100.0%) (Table 4).

## 3. Comparison of indirect versus direct testing

### (a) Diagnostic accuracy

Based on analysis of all data, there was no evidence of a statistically significant difference in MTBDRs<sub>1</sub>/version 1.0 accuracy for SLID resistance between indirect and direct testing when using culture-based DST as a reference standard (P values for differences in sensitivity or specificity of 0.547 and 0.664, respectively) (Table 4). Direct within-study comparisons were not possible because no studies performed MTBDRs<sub>1</sub> testing on specimens and isolates from the same patients.

### (b) Indeterminate rates

For indirect testing for culture-confirmed resistance to SLIDs, of 10 studies that reported indeterminate MTBDRs<sub>1</sub>/version 1.0 results, seven (0.5%) of 1316 results were indeterminate (two culture-DST resistant and five culture-DST sensitive). For direct testing on a smear-positive specimen, of four studies that reported indeterminate MTBDRs<sub>1</sub> results, 219 (13.5%) of 1627 results were indeterminate (34 were culture-DST resistant, 165 were culture-DST susceptible, and 20 whose culture phenotypic status was unknown). The indeterminate rates for direct testing for each smear-grade (smear-negative, scanty, 1+, 2+, 3+) were 76/180 (42.2%), 35/91 (38.5%), 47/213 (22.1%), 29/200 (14.5%), and 70/364 (19.2%), respectively.

## B. Investigations of heterogeneity

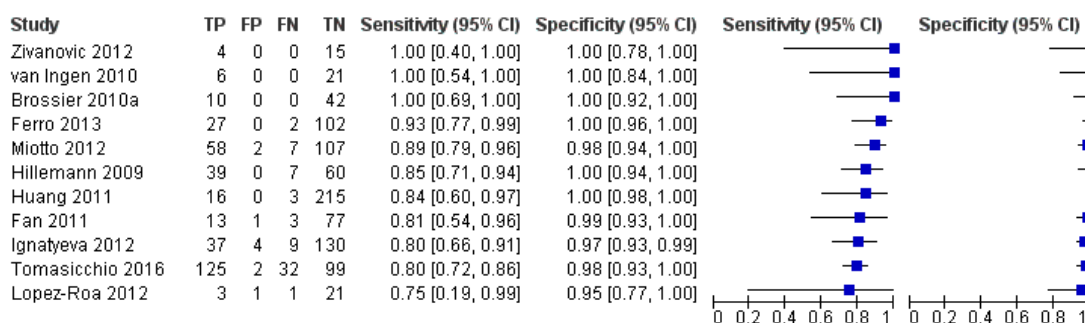
### 1. Indirect testing

#### (a) Individual drugs within the drug class

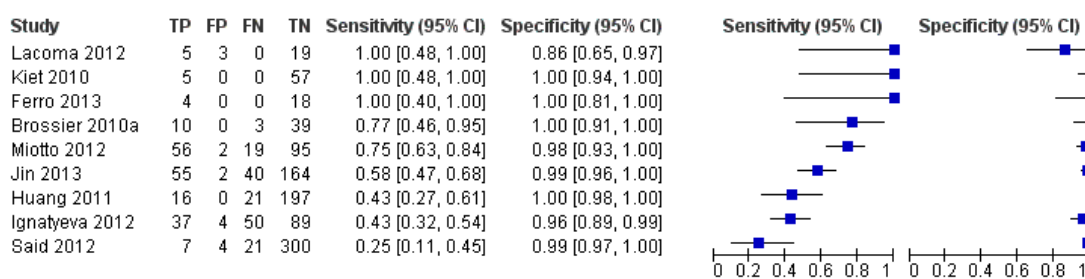
We present accuracy estimates for MTBDRs<sub>1</sub>/version 1.0 by indirect testing for detection of resistance to amikacin, kanamycin, and capreomycin in Table 5 and Figure 9.

**Figure 9. Forest plots of MTBDRsl sensitivity and specificity for the detection of resistance to amikacin, kanamycin, and capreomycin, the test performed indirectly against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

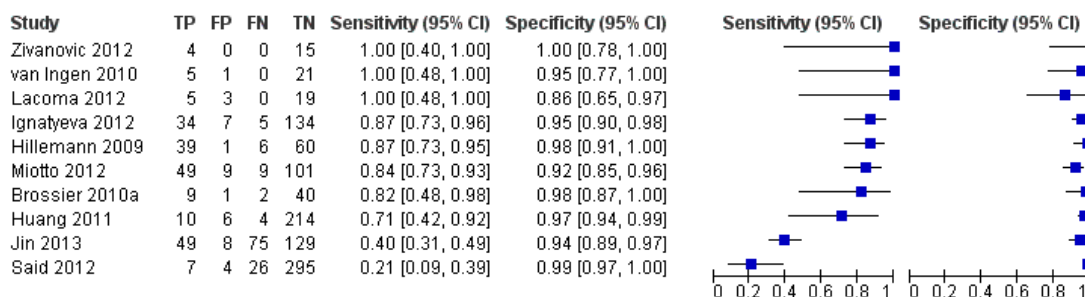
**Indirect, amikacin, culture**



**Indirect, kanamycin, culture**



**Indirect, capreomycin, culture**



For detection of amikacin resistance by indirect testing, MTBDR<sub>s</sub>/ version 1.0 sensitivity estimates ranged from 75% to 100% and specificity estimates ranged from 95% to 100%. The pooled sensitivity and specificity (95% CI) were 84.9% (79.2% to 89.1%) and 99.1% (97.6% to 99.6%), (11 studies, 1301 participants).

For detection of kanamycin resistance by indirect testing, MTBDR<sub>s</sub>/ version 1.0 sensitivity estimates ranged from 25% to 100% and specificity estimates ranged from 86% to 100%. The pooled sensitivity and specificity (95% CI) were 66.9% (44.1% to 83.8%) and 98.6% (96.1% to 99.5%), (nine studies, 1342 participants).

For detection of capreomycin resistance by indirect testing, MTBDR<sub>s</sub>/ version 1.0 sensitivity estimates ranged from 21% to 100% and specificity estimates from 86% to 100%. The pooled sensitivity and specificity (95% CI) were 79.5% (58.3% to 91.4%) and 95.8% (93.4% to 97.3%), (10 studies, 1406 participants).

### **(b) Drug concentration used in culture-based DST**

Appendix 5 shows amikacin, kanamycin, and capreomycin drug concentrations used in culture-based DST in relation to the WHO-recommended critical concentrations.

Amikacin: five studies used the WHO-recommended critical concentration of amikacin (Fan 2011; Ignatyeva 2012; Miotto 2012; Lopez-Roa 2012; Tomasicchio 2016), whereas three did not (Brossier 2010a; Ferro 2013; van Ingen 2010). Hillemann 2009 and Zivanovic 2012, which each used two types of culture media, used the WHO-recommended concentration for one culture type and a non-recommended concentration for the other type. Huang 2011 also used two types of culture media and used the WHO-recommended concentration for one culture type and for the other culture type, no recommended concentration exists. Between studies that used and did not use the WHO-recommended critical concentration for amikacin, there was evidence of a statistically significant difference in specificity (P value < 0.001), but not sensitivity (P value = 0.063).

Kanamycin: two studies used the WHO-recommended critical concentration of kanamycin (Ferro 2013; Huang 2011 for both types of culture media), whereas seven did not (Brossier 2010a; Ignatyeva 2012; Jin 2013; Kiet 2010; Lacoma 2012; Miotto 2012; Said 2012).

Capreomycin: four studies used the WHO-recommended critical concentration of capreomycin (Hillemann 2009 for both culture types; Ignatyeva 2012; Miotto 2012; Zivanovic 2012 for both culture types), and three did not (Brossier 2010a; Said 2012; van Ingen 2010). Huang 2011 used two types of culture media and used the WHO-recommended concentration for one culture type and for the other culture type, no recommended concentration exists. Lacoma 2012 used a culture type for which no recommended concentration exists and Jin 2013 did not report the critical con-

centration used. There was no evidence of a statistically significant difference in MTBDR<sub>s</sub>/ version 1.0 accuracy for capreomycin resistance between studies that used and did not use the WHO-recommended critical concentration (P values for differences in sensitivity and specificity of 0.161 and 0.625, respectively).

Comparisons between accuracy estimates for kanamycin according to critical concentration were not possible given the small number of studies.

### **(c) Type of reference standard**

We present MTBDR<sub>s</sub>/ version 1.0 accuracy estimates for detection of SLID resistance against different reference standards in Table 6 and Appendix 8.

#### Reference standard is sequencing

Using sequencing, MTBDR<sub>s</sub>/ version 1.0 sensitivity estimates ranged from 62% to 100% and specificity estimates ranged from 96% to 100%, (seven studies, 962 participants). We restricted the meta-analysis to comparative studies (six studies, 873 participants). The pooled sensitivity and specificity (95% CI) were 97.0% (77.0% to 99.7%) and 99.5% (94.5% to 100.0%). There was evidence of a significantly higher sensitivity using sequencing as the reference standard compared with culture-based DST (P value of 0.034), but not specificity (P value of 0.456).

#### Reference standard is culture-based DST and sequencing (i.e. both investigations performed on all isolates)

Using culture and sequencing, MTBDR<sub>s</sub>/ version 1.0 sensitivity estimates ranged from 30% to 85% and specificity estimates ranged from 99% to 100%, (seven studies, 1491 participants). We restricted the meta-analysis to comparative studies (six studies, 1159 participants). The pooled sensitivity and specificity (95% CI) were 61.3% (45.8% to 74.8%) and 99.9% (99.0% to 100.0%). There was no evidence of a statistically significant difference in accuracy using culture and sequencing as the reference standard compared with culture-based DST (P values for differences in sensitivity and specificity of 0.458 and 0.203, respectively).

#### Reference standard is culture-based DST followed by sequencing of discrepant index test-culture-based DST results

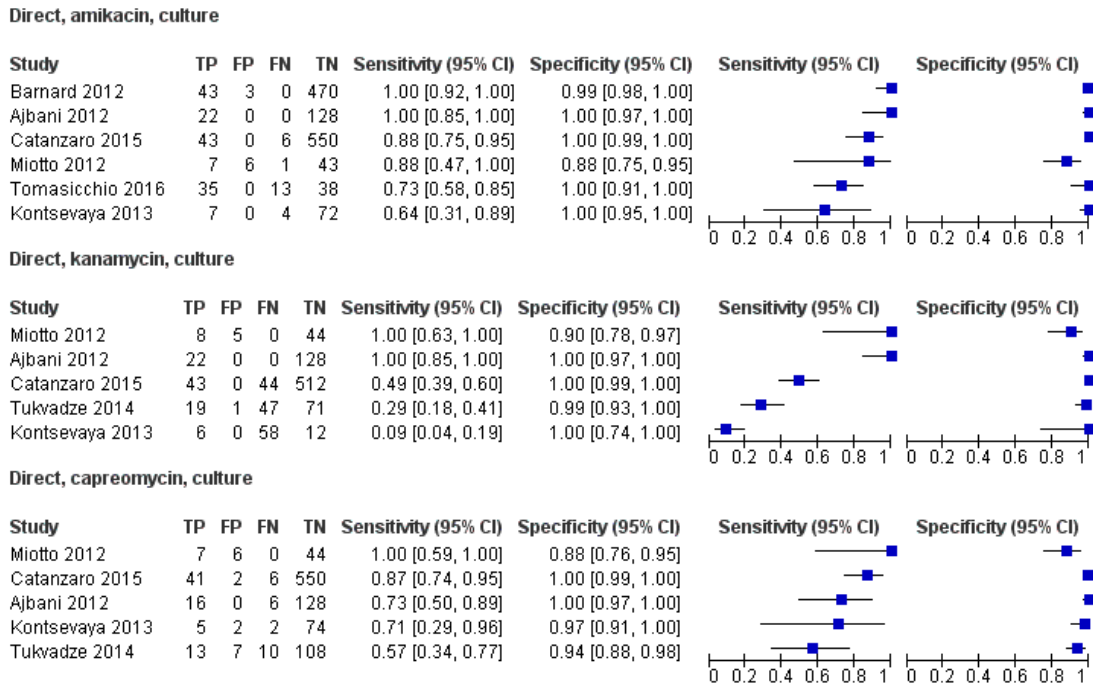
Using sequencing of discrepant results, MTBDR<sub>s</sub>/ version 1.0 sensitivity estimates ranged from 34% to 100% and specificity estimates ranged from 95% to 100%, (three studies, 619 participants). We did not determine summary estimates because there were only three studies and the sensitivity was variable.

## **2. Direct testing**

(a) Individual drugs within the drug class

We present MTBDRs/ accuracy estimates for detection of resistance to amikacin, kanamycin, and capreomycin by direct testing against a phenotypic culture-based reference standard in Figure 10.

**Figure 10. Forest plots of MTBDRs/ sensitivity and specificity for resistance to amikacin, kanamycin, and capreomycin, the test performed directly against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**



For detection of amikacin resistance by direct testing, MTBDRs/ version 1.0 sensitivity estimates ranged from 64% to 100% and specificity estimates ranged from 88% to 100%. The pooled sensitivity and specificity (95% CI) were 91.9% (71.5% to 98.1%) and 99.9% (95.2% to 100.0%), (six studies, 1491 participants). Based on all data, there was no evidence of a statistically significant difference in accuracy between indirect and direct testing for amikacin resistance (P values for differences in sensitivity of 0.338 and 0.213, respectively).

For detection of kanamycin resistance by direct testing, MTBDRs/ version 1.0 sensitivity estimates ranged from 9% to 100% and specificity estimates ranged from 90% to 100%. The pooled sensitivity and specificity (95% CI) were 78.7% (11.9% to 99.0%) and 99.7% (95% CI 93.8% to 100.0%), (five studies, 1020 participants). Based on all data, there was no evidence of a statistically significant difference in accuracy between indirect and direct testing for kanamycin resistance (P values for differences in sensitivity



of specificity 0.836 and 0.445, respectively).

For detection of capreomycin resistance by direct testing, MTBDR<sub>s</sub>/ version 1.0 sensitivity estimates ranged from 57% to 100% and specificity estimates ranged from 88% to 100%. The pooled sensitivity and specificity (95% CI) were 76.6% (61.1% to 87.3%) and 98.2% (92.5% to 99.6%), (five studies, 1027 participants). Based on all data, there was no evidence of a statistically significant difference in accuracy between indirect and direct testing for capreomycin resistance (P values for differences in sensitivity and specificity of 0.841 and 0.353, respectively).

***(b) Drug concentration used in culture-based***

Amikacin: all six studies in this category used the WHO-recommended critical concentration (Ajbani 2012; Barnard 2012; Catanzaro 2015; Kontsevaya 2013; Miotto 2012; Tomasicchio 2016).

Kanamycin: three studies used the WHO-recommended critical concentration for kanamycin (Ajbani 2012; Catanzaro 2015; Tukvadze 2014), whereas two did not (Kontsevaya 2013; Miotto 2012).

Capreomycin: all five studies used the WHO-recommended critical concentration (Ajbani 2012; Catanzaro 2015; Kontsevaya 2013; Miotto 2012; Tukvadze 2014).

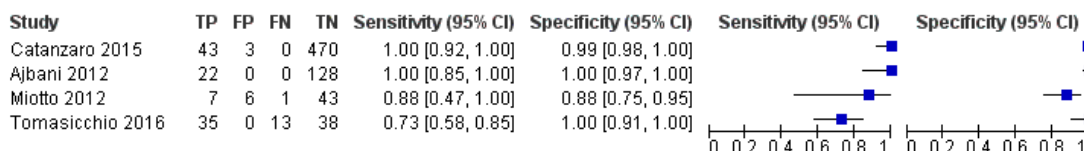
Comparisons between accuracy estimates according to whether or not WHO-recommended critical drug concentrations were used for culture-based reference testing were not possible.

***(c) Stratification by smear grade***

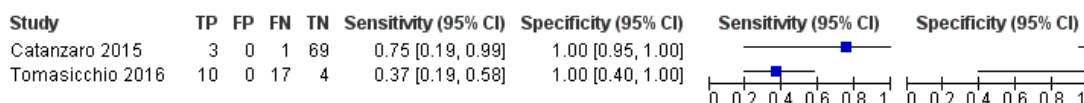
There were limited data on MTBDR<sub>s</sub>/ version 1.0 accuracy for individual SLID drugs by smear grade. Figure 11 presents the forest plots for amikacin resistance by smear grade.

**Figure 11. Forest plots of MTBDRsI sensitivity and specificity for detection of amikacin resistance by smear grade, using culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

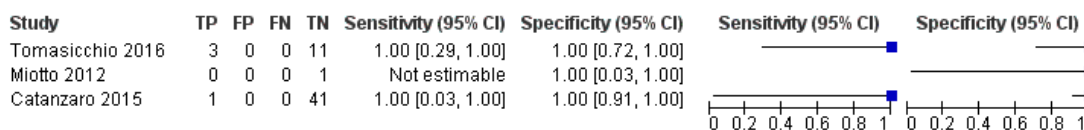
**Amikacin, smear positive**



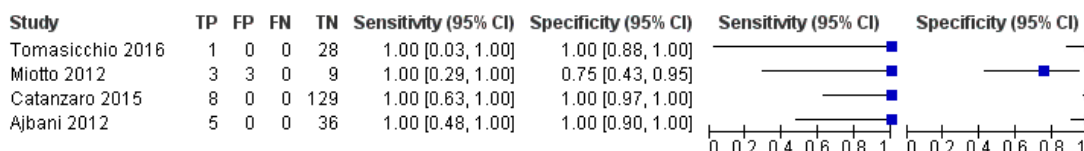
**Amikacin, smear negative**



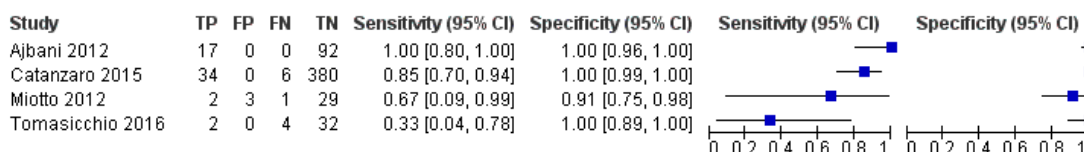
**Amikacin, smear grade = scanty**



**Amikacin, smear grade = 1+**



**Amikacin, smear grade ≥ 2+**



**(d) Type of reference standard**

Reference standard is sequencing

No studies performed direct MTBDRsI version 1.0 testing and used sequencing as a reference standard.

Reference standard is culture-based DST and sequencing (i.e. both investigations performed on all isolates)

No studies performed direct MTBDRsI version 1.0 testing and used both culture-based DST and sequencing (performed on all isolates) as a reference standard.

Reference standard is culture-based DST followed by sequencing

of discrepant index test culture-based DST results

We identified two studies, both of which reported perfect sensitivity and specificity (95% CI): 100% (85% to 100%) and 100% (97% to 100%) for [Ajbani 2012](#), and 100% (92% to 100%) and 100% (98% to 100%) for [Barnard 2012](#), respectively.

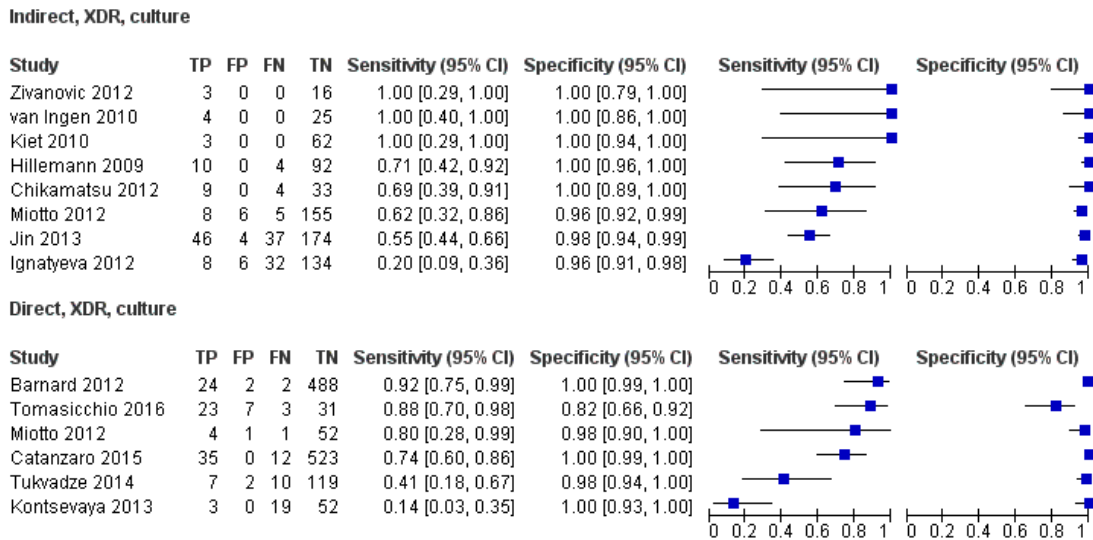
**III. XDR-TB detection**

**A. Estimates of the diagnostic accuracy of MTBDRsI using culture-based DST as a reference standard**

### 1. Indirect testing

Eight studies (880 participants, 173 (19.7%) confirmed cases of XDR-TB) evaluated MTBDR<sub>s</sub>/ version 1.0 by indirect testing for detection of XDR-TB, (Figure 12). Sensitivity estimates ranged from 20% to 100% and specificity estimates ranged from 96% to 100%. The pooled sensitivity and specificity (95% CI) were 70.9% (42.9% to 88.8%) and 98.8% (96.1% to 99.6%).

**Figure 12. Forest plots of MTBDR<sub>s</sub> sensitivity and specificity for the detection of XDR-TB, the test performed indirectly and directly against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**



### 2. Direct testing

Six studies (1420 participants, 143 (10.1%) confirmed cases of XDR-TB) evaluated MTBDR<sub>s</sub>/ version 1.0 by direct testing for detection of XDR-TB, (Figure 12). Sensitivity estimates ranged from 14% to 92% and specificity estimates ranged from 82% to 100%. The pooled sensitivity and specificity (95% CI) were 69.4% (38.8% to 89.0%) and 99.4% (95.0% to 99.3%).

### 3. Comparison of indirect versus direct testing

#### (a) Diagnostic accuracy

Based on analysis of all data, there was no evidence of a statistically significant difference in MTBDR<sub>s</sub>/ version 1.0 accuracy for XDR-TB between indirect and direct testing when using culture-based DST as a reference standard (P values for differences in sensitivity and specificity of 0.916 and 0.387, respectively) (Table 4). Direct within-study comparisons were not possible because no studies performed MTBDR<sub>s</sub>/version 1.0 testing on specimens and isolates from the same patients.

### **(b) Indeterminate rates**

For indirect testing for XDR-TB, of the six studies that reported version 1.0 results, one (0.2%) of 554 was indeterminate (one culture DST sensitive). For direct testing on a smear-positive specimen, of seven studies that reported indeterminate MTBDR<sub>s/l</sub> results, 224 (13.5%) of 1665 results were indeterminate (27 culture-DST resistant, 186 culture-DST susceptible, and 11 of unknown phenotypic culture status). The indeterminate rates for direct testing for each smear-grade (smear-negative, scanty, 1+, 2+, 3+) were 81/183 (44.2%), 39/186 (21.0%), 53/225 (23.5%), 33/301 (11.0%), and 82/177 (46.3%).

## **B. Investigations of heterogeneity**

### **1. Indirect testing**

#### **(a) Drugs used in the culture-based DST**

One of the eight studies that performed indirect testing for XDR-TB and used culture-based DST as a reference standard used ofloxacin and kanamycin (Kiet 2010). Two studies used ofloxacin, amikacin, and capreomycin (Hillemann 2009; Zivanovic 2012). One study used ofloxacin, amikacin, and kanamycin (Miotto 2012). One study used levofloxacin, amikacin, kanamycin, and capreomycin (Chikamatsu 2012). One study used ofloxacin, amikacin, kanamycin, and capreomycin (Ignatyeva 2012). One study used ofloxacin, kanamycin, and capreomycin (Jin 2013). One study used moxifloxacin, amikacin, and ofloxacin (van Ingen 2010).

As all but two studies used a different combination of drugs, we did not compare test performance according to drugs used in the culture-based DST.

#### **(b) Drug concentration used in culture-based DST**

Ofloxacin: two studies in this category used the WHO-recommended critical concentration for ofloxacin (Ignatyeva 2012;

Miotto 2012), and two did not (Jin 2013; Kiet 2010). Two studies used two different types of culture but only used the WHO-recommended critical concentration of ofloxacin for one type of culture (Hillemann 2009; Zivanovic 2012).

Moxifloxacin: van Ingen 2010 used moxifloxacin but did not use the WHO-recommended critical concentration.

Levofloxacin: for the study that used levofloxacin (Chikamatsu 2012), the WHO does not recommend a critical concentration for the type of culture used (Ogawa culture).

Amikacin: for the six studies that used amikacin, two used the WHO-recommended critical concentration (Ignatyeva 2012; Miotto 2012), one did not report the concentration used (Chikamatsu 2012), and one used a type of culture-based testing (Middlebrook 7H10 media) for which the WHO did not specify a recommended critical concentration (van Ingen 2010). Two studies used two types of culture medium, one of which was done at a recommended concentration and the other not (Hillemann 2009; Zivanovic 2012).

Kanamycin: of the five studies that used kanamycin, four did not use the WHO-recommended critical concentration (Ignatyeva 2012; Jin 2013; Kiet 2010; Miotto 2012), and one did not report the concentration used (Chikamatsu 2012).

Capreomycin: of the seven studies that used capreomycin, four used the WHO-recommended critical concentration (Hillemann 2009 for both culture types; Ignatyeva 2012; Miotto 2012; Zivanovic 2012 for both culture types), whereas one study did not (van Ingen 2010); and two studies did not report the concentration used (Chikamatsu 2012; Jin 2013).

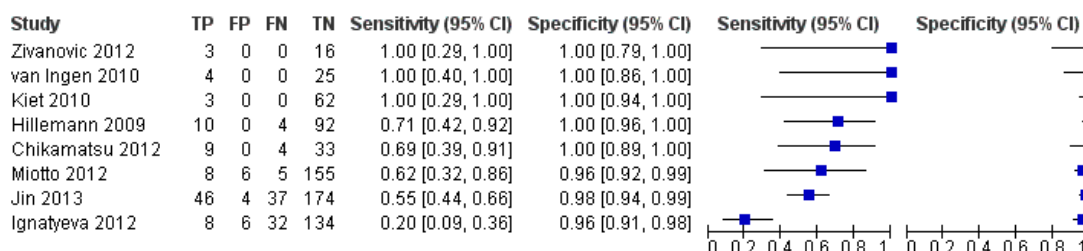
We did not compare MTBDR<sub>s/l</sub> version 1.0 accuracy according to drug concentrations used in the culture-based DST because there were few studies that used the same drugs and the same critical concentrations.

#### **(c) Type of reference standard**

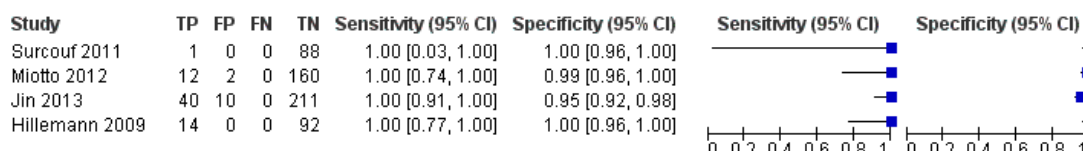
We present MTBDR<sub>s/l</sub> version 1.0 accuracy estimates for detection of XDR-TB against different reference standards in Table 6 and Figure 13.

**Figure 13. Forest plots of MTBDRsl sensitivity and specificity for extensively drug-resistant tuberculosis (XDR-TB), the test performed indirectly against three different reference standards. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

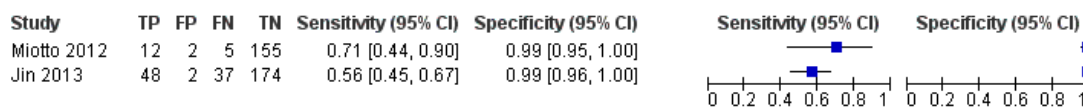
**Indirect, XDR, culture**



**Indirect, XDR, sequencing**



**Indirect, XDR, sequencing and culture**



Reference standard is sequencing

For individual studies (four studies, 630 participants), MTBDRsl version 1.0 sensitivity estimates in all four studies were 100% and specificity estimates ranged from 95% to 100%. We restricted the meta-analysis to comparative studies (three studies, 541 participants). Using a fixed-effect model, the pooled sensitivity and specificity (95% CI) were 100% (94.6% to 100.0%) and 97.9% (96.3% to 98.8%).

Reference standard is culture-based DST and sequencing (i.e. both investigations performed on all isolates)

We identified two studies with 435 participants. MTBDRsl version 1.0 sensitivity and specificity estimates (95% CI) were 56% (45% to 67%) and 99% (96% to 100%) for Jin 2013, and 71% (44% to 90%) and 99% (95% to 100%) for Miotto 2012. We did not perform a meta-analysis.

Reference standard is culture-based DST followed by sequencing of discrepant index test-culture-based DST results

No studies performed indirect MTBDRsl version 1.0 testing for XDR-TB and used culture-based DST and sequencing for discrepant results as a reference standard.

**2. Direct testing**

*(a) Drugs used in the culture-based DST*

Two of the six studies that performed direct testing for XDR-TB and used culture-based DST as a reference standard used ofloxacin and amikacin (Barnard 2012; Tomasicchio 2016). Two studies used ofloxacin, moxifloxacin, amikacin, kanamycin, and capreomycin (Catanzaro 2015; Kontsevaya 2013). One study used ofloxacin, amikacin, kanamycin, and capreomycin (Miotto 2012). One study used ofloxacin, kanamycin, and capreomycin (Tukvadze 2014).

As all but two studies used a different combination of drugs, we did not compare test performance according to drugs used in the culture-based DST.

### *(b) Drug concentration used in culture-based DST*

Ofloxacin: five studies in this category used the WHO-recommended critical concentration for ofloxacin (Barnard 2012; Catanzaro 2015; Kontsevaya 2013; Miotto 2012; Tomasicchio 2016), whereas one did not (Tukvadze 2014).

Moxifloxacin: neither study used the WHO-recommended critical concentration for moxifloxacin (Catanzaro 2015; Kontsevaya 2013).

Amikacin: four studies in this category used the WHO-recommended critical concentration for amikacin (Catanzaro 2015; Kontsevaya 2013; Miotto 2012; Tomasicchio 2016), whereas one study did not (Barnard 2012).

Kanamycin: two studies used the WHO-recommended critical concentration for kanamycin (Catanzaro 2015; Tukvadze 2014), whereas two studies did not (Kontsevaya 2013; Miotto 2012).

Capreomycin: all four studies in this category used the WHO-recommended critical concentration for capreomycin (Catanzaro 2015; Kontsevaya 2013; Miotto 2012; Tukvadze 2014).

We did not compare test accuracy according to drug concentrations used in the culture-based DST because there were few studies using the same drugs and the same critical concentrations.

### *(c) Type of reference standard*

Reference standard is sequencing

No studies performed direct MTBDRs/ version 1.0 testing for XDR-TB and used sequencing as a reference standard.

Reference standard is culture-based DST and sequencing (i.e. both investigations performed on all isolates)

No studies performed direct MTBDRs/ version 1.0 testing and used both culture-based DST and sequencing (performed on all isolates) as a reference standard.

Reference standard is culture-based DST followed by sequencing of discrepant index test-culture-based DST results

We identified two studies that used culture-based DST and performed sequencing only on discrepant results (Barnard 2012; Miotto 2012). These studies both reported sensitivities of 92% and specificities of 100%.

### **Sensitivity analyses**

We have presented the MTBDRs/ version 1.0 sensitivity analyses (using culture-based DST as the reference standard) for the FQs (Table 7) and the SLIDs (Table 8). The sensitivity analyses made no difference to any of the findings.

### **MTBDRs/ version 2.0**

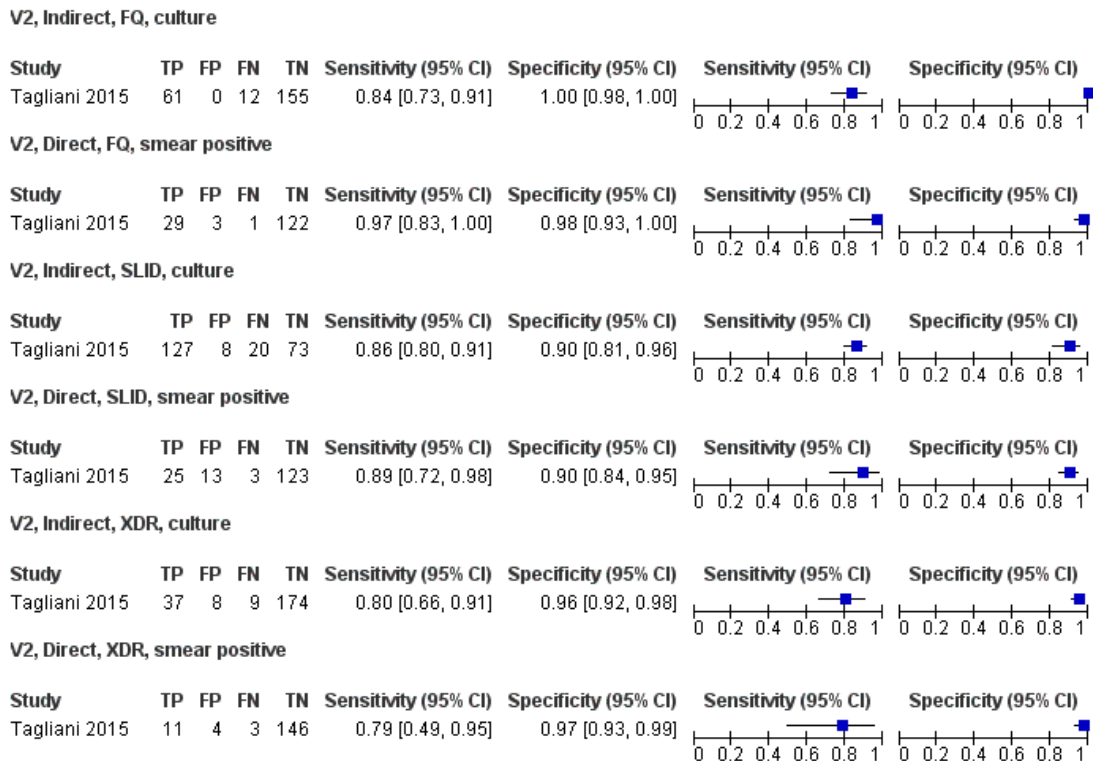
#### **I. Fluoroquinolone resistance detection**

##### **A. Estimates of the diagnostic accuracy of MTBDRs/ using culture-based DST as a reference standard**

###### **1. Indirect testing**

By indirect testing, MTBDRs/ version 2.0 sensitivity and specificity (95% CI) were 84% (73% to 91%) and 100% (98% to 100%) (Tagliani 2015; Figure 14).

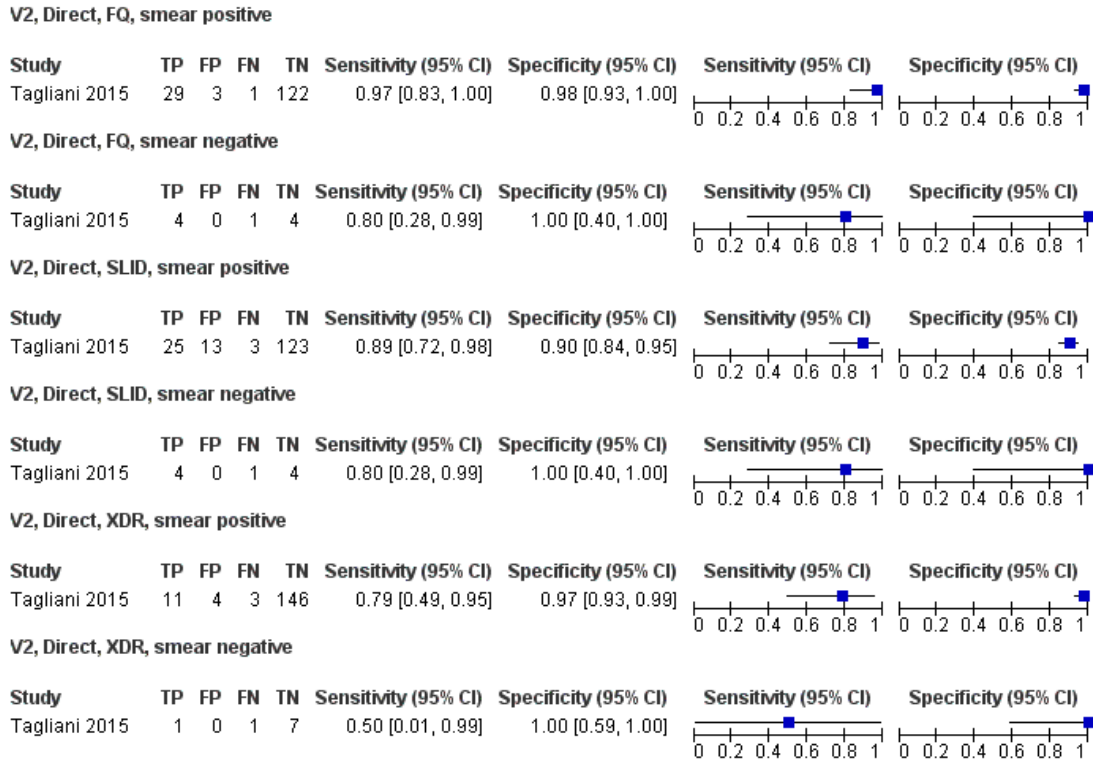
**Figure 14. Forest plots of MTBDRsl version 2.0 sensitivity and specificity for the detection of resistance to the fluoroquinolones (FQs) and second-line injectable drugs (SLIDs) and extensively drug-resistant tuberculosis (XDR-TB), the test performed indirectly and directly against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line).**



## 2. Direct testing

By direct testing, MTBDRsl version 2.0 sensitivity and specificity (95% CI) were 97% (83% to 100%) and 98% (93% to 100%) on a smear-positive specimen and 80% (28% to 99%) and 100% (40% to 100%) on a smear-negative specimen (Tagliani 2015; Figure 15).

**Figure 15. Forest plots of MTBDRsl version 2.0 sensitivity and specificity for the detection of resistance to the fluoroquinolones (FQs) and second-line injectable drugs (SLIDs) and extensively drug-resistant tuberculosis (XDR-TB), the test performed directly on smear-positive and smear-negative specimens against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line).**



### 3. Drug concentration used in culture-based DST

Ofloxacin: [Tagliani 2015](#) used the WHO-recommended critical concentration.

Moxifloxacin: [Tagliani 2015](#) used the WHO-recommended critical concentration for low-level moxifloxacin resistance.

Levofloxacin: [Tagliani 2015](#) used the WHO-recommended critical concentration.

### 4. Indeterminate rates

For both indirect and direct testing for culture-confirmed resistance to FQs, [Tagliani 2015](#) reported zero indeterminate results.

## II. SLID resistance detection

### A. Estimates of the diagnostic accuracy of MTBDRsl using culture-based DST as a reference standard

#### 1. Indirect testing

For MTBDRsl/ version 2.0 performed by indirect testing, sensitivity and specificity (95% CI) were 86% (80% to 91%) and 90% (81% to 96%) ([Figure 14](#); [Tagliani 2015](#)).



## 2. Direct testing

By direct testing, sensitivity and specificity (95% CI) were 89% (72% to 98%) and 90% (84% to 95%) on a smear-positive specimen and 80% (28% to 99%) and 100% (40% to 100%) on a smear-negative specimen (Figure 15; Tagliani 2015).

## 3. Drug concentration used in culture-based DST

Amikacin: Tagliani 2015 used the WHO-recommended critical concentration of amikacin.

Kanamycin: Tagliani 2015 used the WHO-recommended critical concentration of kanamycin.

Capreomycin: Tagliani 2015 used the WHO-recommended critical concentration of capreomycin.

## 4. Indeterminate rates

For both indirect and direct testing for culture-confirmed resistance to SLIDs, Tagliani 2015 reported zero indeterminate results.

# III. XDR-TB detection

## A. Estimates of the diagnostic accuracy of MTBDRs/ using culture-based DST as a reference standard

### 1. Indirect testing

For MTBDRs/ version 2.0 performed by indirect testing, sensitivity and specificity (95% CI) were 80% (66% to 91%) and 96% (82% to 98%) (Figure 14; Tagliani 2015).

## 2. Direct testing

By direct testing, sensitivity and specificity (95% CI) were 79% (49% to 95%) and 97% (93% to 99%) on a smear-positive specimen and 50% (1% to 99%) and 100% (59% to 100%) on a smear-negative specimen (Figure 15; Tagliani 2015).

## 3. Indeterminate rates

Tagliani 2015 reported zero indeterminate results for both indirect and direct testing for XDR-TB.

## Comparison of the accuracy of version 1.0 and 2.0

As we identified only one study that evaluated MTBDRs/ version 2.0, we could not compare the accuracy of MTBDRs/ version 1.0 and 2.0.

## Other analyses

We did not identify any reports on intra-reader variability. One study described inter-reader variability and reported high concordance > 95% (Ignatyeva 2012).

Regarding patient outcomes, only four studies described the effect of MTBDRs/ on time-to-diagnosis. Lopez-Roa 2012 reported the test to have a time-to-diagnosis of eight hours, compared to DST using the agar proportion method (21 days) or the MGIT 960 method (eight days). Said 2012 stated that MTBDRs/ had a median time-to-diagnosis of two days, compared to 11 days for the agar proportion method. Tukvadze 2014 noted a median time-to-diagnosis using MTBDRs/ of 10 days, versus 70 to 104 days for culture-based DST. Barnard 2012 reported MTBDRs/ to have a median turn-around-time of one day (after the diagnosis of first-line resistance), whereas the median turn-around-time for phenotypic culture-based DST was 31 days.

## ADDITIONAL SUMMARY OF FINDINGS *[Explanation]*

**Participants:** patients with rifampicin-resistant or MDR-TB  
**Prior testing:** patients who received MTBDR<sub>s</sub>/ testing may have first received smear microscopy, Xpert<sup>®</sup> MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB and Xpert<sup>®</sup> MTB/RIF, MTBDR<sub>s</sub>*plus* version 2.0 or an alternative line-probe assay to detect first-line drug resistance  
**Role:** The role of MTBDR<sub>s</sub>/ would be as the initial test, replacing culture-based drug susceptibility testing, for detecting second-line drug resistance  
**Settings:** intermediate or central level laboratories  
**Index (new) test:** MTBDR<sub>s</sub>/ version 1.0.\* The test was performed by direct testing on smear-positive specimens  
**Reference standard:** culture-based drug susceptibility testing  
**Studies:** cross-sectional studies  
**Limitations:** most included studies did not consistently use the World Health Organization (WHO)-recommended concentrations for drugs in the culture-based reference standard  
**Pooled sensitivity (95% CI):** 87.0% (38.1% to 98.6%)  
**Pooled specificity (95% CI):** 99.5% (93.6% to 100.0%)

| Test result   | Number of results per 1000 patients tested (95% CI) |                   |                   | Number of participants (studies) | Quality of the evidence (GRADE) |
|---|---|-------------------|-------------------|----------------------------------|---------------------------------|
|   | Prevalence of 5%                                    | Prevalence of 10% | Prevalence of 15% |                                  |                                 |
| <b>True positives</b><br>(patients correctly diagnosed with SLID resistance)              | 44 (19 to 49)                                       | 87 (38 to 99)     | 131 (57 to 148)   | 348 (8)                          | ⊕⊕○○ <sup>1,2,3,4</sup><br>low  |
| <b>False negatives</b><br>(patients incorrectly classified as not having SLID resistance) | 6 (1 to 31)   | 13 (1 to 62)      | 19 (2 to 93)      |                                  |                                 |
| <b>True negatives</b><br>(patients correctly classified as not having SLID resistance)    | 945 (889 to 950)                                    | 896 (842 to 900)  | 846 (796 to 850)  | 8 (1291)                         | ⊕⊕⊕○ <sup>1,2</sup><br>moderate |

|   |             |             |             |
|---|-------------|-------------|-------------|
| <b>False positives</b><br>(patients incorrectly classified as having SLID resistance) | 5 (0 to 61) | 4 (0 to 58) | 4 (0 to 54) |
|---|-------------|-------------|-------------|

Abbreviations: CI: confidence interval; DST: drug susceptibility testing; FQ: fluoroquinolone; GRADE: Grading of Recommendations, Assessment, Development and Evaluation; SLID: second-line injectable drug; TB: tuberculosis; XDR-TB: extensively drug-resistant TB.

By indirect testing, MTBDR<sub>s</sub>/sensitivity and specificity (95% CI) were 76.5% (63.3% to 86.0%) and 99.1% (97.3% to 99.7%).

\*This systematic review mainly evaluated MTBDR<sub>s</sub>/ version 1.0, which has recently been replaced with version 2.0. We considered the findings in this review to be applicable to the current version of the test.

<sup>1</sup>We used QUADAS-2 to assess risk of bias. All studies used consecutive or random sampling. In six studies, the reader of the index test was blinded to results of the reference standard in two studies information about blinding to the reference standard was not reported. Fifty per cent of the studies used critical concentrations for the culture-based DST reference standard that differed from the concentrations recommended by the WHO. We downgraded one level.

<sup>2</sup>We considered indirectness (applicability) from the perspective of diagnostic accuracy and had low concern. We did not downgrade.

<sup>3</sup>For individual studies, sensitivity estimates ranged from 9% to 100%. We thought heterogeneity could be explained in part by the use of different drugs, critical concentrations, and types of culture media in the reference standard and likely presence of *eis* mutations in patients in Eastern Europe. We did not downgrade for inconsistency and considered this in the context of other factors, in particular imprecision.

<sup>4</sup>The wide CI around true positives and false negatives may lead to different decisions depending on which confidence limits are assumed. We downgraded one level.

| <p><b>Participants:</b> patients with rifampicin-resistant or MDR-TB<br/> <b>Prior testing:</b> patients who received MTBDR<sub>s</sub>/ testing may have first received smear microscopy, Xpert<sup>®</sup> MTB/RIF or other nucleic acid amplification test, and culture to diagnose TB and Xpert<sup>®</sup> MTB/RIF, MTBDR<sub>s</sub><i>plus</i> version 2.0 or an alternative line-probe assay to detect first-line drug resistance<br/> <b>Role:</b> The role of MTBDR<sub>s</sub>/ would be as the initial test, replacing culture-based drug susceptibility testing, for detecting second-line drug resistance<br/> <b>Settings:</b> intermediate or central level laboratories<br/> <b>Index (new) test:</b> MTBDR<sub>s</sub>/ version 1.0.* The test was performed by direct testing on smear-positive specimens<br/> <b>Reference standard:</b> culture-based drug susceptibility testing<br/> <b>Studies:</b> cross-sectional studies<br/> <b>Limitations:</b> most included studies did not consistently use the World Health Organization (WHO)-recommended concentrations for drugs in the culture-based reference standard<br/> <b>Pooled sensitivity (95% CI):</b> 69.4% (38.8% to 89.0%)<br/> <b>Pooled specificity (95% CI):</b> 99.4% (95.0% to 99.3%)</p> |   |                  |                   |                                  |                                 |
|--|---|------------------|-------------------|----------------------------------|---------------------------------|
| Test result  | Number of results per 1000 patients tested (95% CI) |                  |                   | Number of participants (studies) | Quality of the evidence (GRADE) |
|  | Prevalence of 1%                                    | Prevalence of 5% | Prevalence of 10% |                                  |                                 |
| <b>True positives</b><br>(patients correctly diagnosed with XDR-TB)  | 7 (4 to 9)  | 35 (19 to 45)    | 69 (39 to 89)     | 143 (6)                          | ⊕⊕○○ <sup>1,2,3,4</sup><br>low  |
| <b>False negatives</b><br>(patients incorrectly classified as not having XDR-TB)   | 3 (1 to 6)  | 15 (5 to 31)     | 31 (11 to 61)     |                                  |                                 |
| <b>True negatives</b><br>(patients correctly classified as not having XDR-TB)  | 980 (941 to 983)                                    | 941 (903 to 943) | 891 (855 to 894)  | 1277 (6)                         | ⊕⊕⊕○ <sup>1,2</sup><br>moderate |
| <b>False positives</b><br>(patients incorrectly classified as having XDR-TB)   | 10 (7 to 49)  | 9 (7 to 47)      | 9 (6 to 45)       |                                  |                                 |

Abbreviations: CI: confidence interval; DST: drug susceptibility testing; FQ: fluoroquinolone; GRADE: Grading of Recommendations, Assessment, Development and Evaluation; SLID: second-line injectable drug; TB: tuberculosis; WHO: World Health Organization; XDR-TB: extensively drug-resistant TB.

By indirect testing, MTBDR<sub>s</sub>/sensitivity and specificity (95% CI) were 70.9% (42.9% to 88.7%) and 98.8% (96.1% to 99.6%).

\*This systematic review mainly evaluated MTBDR<sub>s</sub>/ version 1.0, which has recently been replaced with version 2.0. We considered the findings in this review to be applicable to the current version of the test.

<sup>1</sup>We used QUADAS-2 to assess risk of bias. All studies used consecutive sampling. In four studies, the reader of the test was blinded to results of the reference standard and in two studies information about blinding was not reported. Most studies used critical concentrations for the phenotypic culture-based DST reference standard that differed from the concentrations recommended by the WHO. We downgraded the evidence by one level.

<sup>2</sup>We considered indirectness (applicability) from the perspective of diagnostic accuracy and had low concern. We did not downgrade.

<sup>3</sup>For individual studies, sensitivity estimates ranged from 14% to 92%. We thought heterogeneity could be explained in part by the use of different drugs, critical concentrations, and types of culture media in the reference standard and likely presence of *eis* mutations in patients in Eastern Europe. We did not downgrade for inconsistency and considered this in the context of other factors, in particular imprecision.

<sup>4</sup>The wide CI for true positives and false negatives may lead to different decisions depending on which confidence limits are assumed. We downgraded one level.

## DISCUSSION

This updated systematic review summarizes the current literature and includes 27 studies and integrates six new studies: five new studies for MTBDRs<sub>1</sub> version 1.0 identified since the original Cochrane review (Theron 2014), and one study for MTBDRs<sub>1</sub> version 2.0. For MTBDRs<sub>1</sub> version 1.0, the findings in this updated review are consistent with those reported in the previous version of the review. We have presented the average sensitivities and specificities of MTBDRs<sub>1</sub> version 1.0 for detection of resistance to fluoroquinolones (FQs) and second-line injectable drugs (SLIDs) and for extensively drug-resistant tuberculosis (XDR-TB) in the 'Summary of findings' tables (Summary of findings 1; Summary of findings 2; Summary of findings 3) and Table 4 Table 5, and Table 6

We found that, when we compared MTBDRs<sub>1</sub> version 1.0 accuracy according to whether the test was performed directly or indirectly, the sensitivities were similar for FQ resistance and SLID resistance. When used indirectly on a culture isolate, MTBDRs<sub>1</sub> had higher pooled sensitivity for detection of FQ resistance (85.6%) than for detection of SLID resistance (76.5%). When used directly on a smear-positive specimen, MTBDRs<sub>1</sub> had similar pooled sensitivity for FQ resistance (86.2%) and SLID resistance (87.0%); however the pooled sensitivity for SLID resistance was imprecise (95% confidence interval (CI) 38.1% to 98.8%). When SLID resistance was analysed for individual drugs, the pooled sensitivity (direct testing) was highest for amikacin (92.0%). For detection of XDR-TB the pooled sensitivity of MTBDRs<sub>1</sub> was 70.9% by indirect testing and 69.4% by direct testing. For detection of resistance to FQs and SLIDs and XDR-TB by either indirect or direct testing, MTBDRs<sub>1</sub> pooled specificity was high (> 98%).

We compared the accuracy of MTBDRs<sub>1</sub> version 1.0 against different reference standards comprised of culture-based drug susceptibility testing (DST) (the traditional reference standard) or sequencing. We looked at MTBDRs<sub>1</sub> version 1.0 accuracy against each type of reference standard alone or in combination (where all specimens received both culture-based DST and sequencing). When used indirectly on a culture isolate for detection of FQ resistance, MTBDRs<sub>1</sub> version 1.0 had higher pooled sensitivity against sequencing than against culture-based DST (99.3% versus 82.4%). This suggests that MTBDRs<sub>1</sub> is sensitive for detecting FQ resistance caused by mutations in *gyrA* (the only gene that is targeted by MTBDRs<sub>1</sub> for detection of FQ resistance). However, against culture-based DST, MTBDRs<sub>1</sub> sensitivity for FQ resistance was only 82.4% suggesting that just less than one in five cases may be caused by mutations outside of *gyrA*, such as in *gyrB*, a gene which is not targeted by MTBDRs<sub>1</sub> version 1.0. An alternative explanation is that the proportion of bacilli harbouring the mutation is below the threshold of detection by MTBDRs<sub>1</sub> but not below the threshold of detection of the phenotypic test.

Similarly, we found higher pooled sensitivity for SLID resistance when MTBDRs<sub>1</sub> version 1.0 was evaluated against sequencing

rather than culture-based DST (97.0% versus 74.6%). In this case, both sequencing and MTBDRs<sub>1</sub> only target the *rrs* gene for resistance to SLIDs. This approach can potentially miss mutations outside of this region that are responsible for SLID resistance. Using culture-based DST (sensitivity 74.6%), it appears that around one in four cases of SLID-resistant TB may be caused by mutations outside of *rrs*. The prevalence of these non-*rrs* mutations, which can occur in regions such as *thyA*, *eis*, and *gidB* (Georghiou 2012), appears to be most pronounced for kanamycin given the reduced sensitivity (66.9%) of MTBDRs<sub>1</sub> for resistance to this drug compared to the other SLIDs (sensitivity of 84.9% and 79.5% for amikacin and capreomycin, respectively, MTBDRs<sub>1</sub> performed indirectly against culture-based DST). The sensitivity of MTBDRs<sub>1</sub> for SLID resistance, and in particular kanamycin resistance, is likely to vary according to the genetic background of *M. tuberculosis* strains, where some may have a greater frequency of resistance-causing mutations that fall outside of *rrs* and different levels of cross-resistance within the SLIDs. As in the case of the FQs, the mutation-harboring bacilli may fall below the limit of detection of MTBDRs<sub>1</sub>.

We are aware of two systematic reviews of MTBDRs<sub>1</sub> (Feng 2013; WHO 2013). Feng 2013 (11 published studies) determined MTBDRs<sub>1</sub> sensitivity and specificity for resistance to second-line anti-TB drugs using a fixed-effect meta-analysis model, rather than fitting the bivariate random-effects model currently recommended. Accuracy estimates for resistance to kanamycin and capreomycin resistance were substantially lower than the ones we found. As in our review, WHO 2013 (11 published and seven unpublished studies) used a random-effects meta-analysis model and arrived at similar summary estimates. Our review included additional studies not included in these previous reviews. Key questions remain regarding test accuracy and potential sources of heterogeneity, including risk of bias, type of testing (indirect versus direct testing), and reference standard (for example, culture-based DST versus genetic sequencing). We addressed several of these questions in this review. Although we intended to investigate whether the observed test accuracy varied between studies according to HIV infection status, specimen condition (frozen versus fresh), specimen type (induced sputum or extrapulmonary specimen), the drug concentration used in culture-based DST (studies that used the World Health Organization (WHO)-recommended concentrations versus those that did not) or population (patients suspected of having MDR-TB or XDR-TB), there were unfortunately insufficient data to perform these additional analyses for each target condition. We were also unable to examine sources of heterogeneity for detection of XDR-TB due to insufficient data. We also had limited data to investigate the influence of smear grade on accuracy estimates.

### Summary of main results

The main results are presented in the 'Summary of findings' tables.

1. When performed indirectly on a culture isolate, MTBDRs/ version 1.0 sensitivity for FQ resistance was 85.6% compared with 86.2% when performed directly on a smear-positive specimen; the specificities for indirect testing (98.5%) and direct testing (98.6%) were high.

2. When performed indirectly, MTBDRs/ version 1.0 sensitivity for SLID resistance was 76.5% compared with 87.0% when performed directly on a smear-positive specimen; the specificities for indirect testing (99.1%) and direct testing (99.5%) were high.

3. When performed indirectly, MTBDRs/ version 1.0 sensitivity for XDR-TB was 70.9% compared with 69.4% performed directly on a smear-positive specimen; the specificities for indirect testing (98.8%) and direct testing (98.0%) were high.

4. For MTBDRs/ version 1.0, we found no evidence of a statistically significant difference in accuracy between indirect testing and direct testing on a smear-positive specimen for FQ resistance, SLID resistance, or XDR-TB.

5. We had insufficient data to estimate summary diagnostic accuracy of MTBDRs/ version 2.0 (smear-positive or smear-negative specimens) or compare accuracy of the two versions.

### **Application of the meta-analysis to a hypothetical cohort**

In the 'Summary of findings' tables we have summarized the review findings for MTBDRs/ version 1.0 by applying the results to a hypothetical cohort of 1000 individuals with rifampicin-resistant TB or MDR-TB thought to have resistance to a FQ, or SLID, or both (Summary of findings 1; Summary of findings 2; Summary of findings 3). We have presented several scenarios, based on the prevalence of drug-resistant TB suggested by the WHO of 5%, 10%, and 15% for FQ and SLID resistance and 1%, 5%, and 10% for XDR-TB. We chose these thresholds based on the findings from global surveillance of second-line resistance among patients with rifampicin-resistance TB or MDR-TB from 75 countries (WHO 2015). The consequences of false positive (FP) results are likely patient anxiety, morbidity from additional testing, possible delays in further diagnostic evaluation, and prolonged and unnecessary treatment with drugs that may have lower bactericidal activity than second-line regimens and often have serious side effects. The consequences of false negative (FN) results are an increased risk of patient morbidity and mortality, and continued risk of community transmission of drug-resistant TB.

### **MTBDRs/ version 1.0, direct testing (smear-positive specimen) for fluoroquinolone resistance**

By direct testing (smear-positive specimen), the test detected 86% of people with FQ resistance and rarely gave a positive result for people without resistance. In a population of 1000 people, where 150 have FQ resistance, MTBDRs/ will correctly identify 129 people with FQ resistance and miss 21 people. In this same population

of 1000 people, where 850 people do not have FQ resistance, the test will correctly classify 838 people as not having FQ resistance and misclassify 12 people as having resistance.

### **MTBDRs/ version 1.0, direct testing (smear-positive specimen) for SLID resistance**

By direct testing (smear-positive specimen), the test detected 87% of people with SLID resistance and rarely gave a positive result for people without resistance. In a population of 1000 people, where 150 have SLID resistance, MTBDRs/ will correctly identify 131 people with SLID resistance and miss 19 people. In this same population of 1000 people, where 850 do not have SLID resistance, the test will correctly classify 846 people as not having SLID resistance and misclassify four people as having resistance.

### **MTBDRs/ version 1.0, direct testing (smear-positive specimen) for XDR-TB**

By direct testing (smear-positive specimen), the test detected 69% of people with XDR-TB and rarely gave a positive result for people without resistance. In a population of 1000 people, where 100 have XDR-TB, MTBDRs/ will correctly identify 69 people with XDR-TB and miss 31 people. In this same population of 1000 people, where 900 do not have XDR-TB, the test will correctly classify 891 people as not having SLID resistance and misclassify nine people as having resistance.

## **Strengths and weaknesses of the review**

The results of this review are based on strict and careful literature searches, study inclusion, and data extraction. The strength of this review is that it allows an assessment of different methods of testing (indirect versus direct) and different reference standards.

### **Completeness of evidence**

This is a reasonably complete data set. We included any non-English studies we found from which we could obtain accuracy data. However, we acknowledge that we may have missed some studies despite the comprehensive search.

### **Accuracy of the reference standards used**

For our primary analysis, we used culture-based DST. This was the most frequently deployed reference standard in the included studies. Although considered to be the best reference standard for drug-resistant TB, culture-based DST is not 100% accurate for detection of drug resistance, in particular with respect to detection of second-line drug resistance. We also determined MTBDRs/ accuracy using sequencing (gene sequencing of loci known to be associated with drug resistance) and both sequencing and culture-based

DST as reference standards. Many TB experts consider sequencing to be the best available reference standard, provided it encompasses all the possible resistance-determining regions. In addition, we determined the accuracy of MTBDRs/ against a fourth reference standard, where sequencing was only performed as part of an analysis for culture-DST-MTBDRs/ discrepant results. However, in most cases we were unable to determine summary estimates due to the small number of studies and therefore were unable to compare MTBDRs/ accuracy estimates using this reference standard with those obtained using culture-based DST, or sequencing, or both, as the reference standard. MTBDRs/ accuracy was generally greater when measured against a reference standard that included genetic testing. However, such genetic testing was only limited to the genes the MTBDRs/ targeted and did not detect mutations outside of these genes that may cause phenotypic drug resistance.

### Quality and quality of reporting of the included studies

We judged that greater than 50% of studies had low risk of bias for the patient selection, index test, and flow and timing domains. We judged that only three studies (11%) had low risk of bias for the reference standard domain because these studies used the WHO-recommended critical concentrations for every drug in the culture-based DST reference standard whereas the other studies did not. Regarding applicability, we had low concern for all QUADAS-2 domains. In general, studies were fairly well reported, though we corresponded with almost all study authors for additional data and missing information. However, accuracy data for individual drugs and smear grades was not well reported and blinding was not reported in a minority of studies. We strongly encourage future studies to follow the recommendations in the Standards for Reporting Diagnostic Accuracy (STARD) statement to improve the quality of reporting (Bossuyt 2015).

### Interpretability of subgroup analyses

We investigated potential sources of heterogeneity in the different reference standards used and the individual drugs in the FQ and SLID drug classes. We performed statistical testing and provided P values where appropriate. Where data were sufficient, we derived accuracy estimates from comparative studies of test accuracy in which the same set of studies was used for each test evaluation. Where data from comparative accuracy studies were limited, we used all relevant data. For some subgroups (for example, people living with HIV), there were insufficient data to perform an analysis.

### Completeness and relevance of the review

There are now several commercially-available tests in addition to MTBDRs/ for detection of resistance beyond MDR-TB. These include TB Resistance Module Fluoroquinolones/Ethambutol and

TB Resistance Module Kanamycin/Amikacin/Capreomycin/Streptomycin (Autoimmun Diagnostika GmbH (AID) Strassberg); MolecuTech REBA MTB-FQ<sup>®</sup>, MolecuTech REBA MTB-KM<sup>®</sup>, and MolecuTech REBA MTB-XDR<sup>®</sup> (YD diagnostics, Seoul); and NiPro LiPA FQ (NiPro Co, Osaka). Our review is the most complete analysis of the diagnostic accuracy of the MTBDRs/ test to date.

### Unpublished data

We did not include unpublished data.

### Applicability of findings to the review question

We had low concern about the applicability of the included studies to our review question as assessed by QUADAS-2. This review mainly evaluated MTBDRs/ version 1.0 which has recently been replaced with version 2.0. We reasoned that with the addition of new probes targeting more resistance-causing mutations, the sensitivity of MTBDRs/ version 2.0 would be expected to be the same or higher than that of MTBDRs/ version 1.0 and the specificity to remain the same or decrease slightly because of the small likelihood that at least one of the probes may misprime (resulting in an increase in false-positive results). Therefore the findings of this review should be considered applicable to the test. However, it is important to note that this review assessed sensitivity and specificity in research settings. Although the patient characteristics and settings matched our review question in most cases, as studies were carried out under research conditions, it is possible that the accuracy of MTBDRs/ may be lower in routine practice settings.

## AUTHORS' CONCLUSIONS

### Implications for practice

In people with rifampicin-resistant or multidrug-resistant tuberculosis, MTBDRs/ performed on a culture isolate or smear-positive specimen may be useful in detecting second-line drug resistance. MTBDRs/ performed on a smear-positive specimen correctly classified around six in seven people as having fluoroquinolone or SLID resistance, although the sensitivity estimates for SLID resistance varied. The test rarely gave a positive result for people without drug resistance. However, when second-line drug resistance is not detected (MTBDRs/ result is negative), conventional DST can still be used to evaluate patients for resistance to the fluoroquinolones or SLIDs.

An enhancement of MTBDRs/ version 2.0 in comparison to version 1.0 is that, according to the manufacturer, the test may be performed on a smear-negative specimen with high diagnostic accuracy, although we had insufficient data to investigate this.



## Implications for research

Future studies should evaluate MTBDRs<sub>l</sub> version 2.0 in particular on smear-negative specimens and in different laboratory settings and populations (for example, in HIV-positive people). Test accuracy should be determined and compared using strains from different geographical regions, as these are likely to have different frequencies of resistance-causing mutations that fall outside of the genes targeted by MTBDRs<sub>l</sub> (and therefore MTBDR<sub>s</sub><sub>l</sub> will likely have different sensitivities for each drug class in these strains). Such future research should include as a reference standard sequencing that targets all known resistance-determining mutations and not just those detectable using MTBDRs<sub>l</sub>. Future molecular tests for FQ and SLID resistance should have more genetic targets than *gyrA*, *gyrB*, *eis*, and *rrs*. Studies are needed to determine inter-reader variability of the test. Although we recognize that WHO-recommended critical concentrations for individual drugs may change over time, researchers should consider incorporating these critical concentrations into their culture-based reference standards. Studies are also needed to assess the effect of MTBDRs<sub>l</sub> implementation on time-to-treatment, patient health outcomes, and cost-effectiveness.

## ACKNOWLEDGEMENTS

Development of the systematic review was in part made possible with financial support from the United States Agency for International Development (USAID) administered by the World Health Organization (WHO) Global TB Programme and the Bill and Melinda Gates Foundation. The review authors thank Dr Jon Deeks for guidance while drafting the protocol (Theron 2013). The editorial base of the Cochrane Infectious Diseases Group is funded by UK aid from the UK Government for the benefit of developing countries (Grant: 5242). The views expressed in this review do not necessarily reflect UK government policy. GT is supported in part by the Wellcome Trust (WT099854MA) and a South African Medical Research Council Career Development Award (GT). MR is supported in part by the Effective Health Care Research Consortium, which is funded by UKaid from the UK Government Department for International Development (DFID). We are grateful to Vittoria Lutje, Information Specialist with the Cochrane Infectious Diseases Group (CIDG), for help with the search strategy. Sarah Donegan provided statistical support to the original review. We thank Katrina Ramsey, Oregon Health & Science University, Portland, USA, for technical assistance. We thank all authors of the included studies for answering our questions and providing additional data. We thank Dr Yang Wu and Dr Tomomi Kitamura for their assistance with translations and Emma L Doughty, Warwick Medical School, for help with the glossary of genetic terms.

## REFERENCES

### References to studies included in this review

#### Ajbani 2012 {published data only}

Ajbani K, Nikam C, Kazi M, Gray C, Boehme C, Balan K, et al. Evaluation of GenoType MTBDR<sub>s</sub><sub>l</sub> assay to detect drug resistance associated with fluoroquinolones, aminoglycosides and ethambutol on clinical sediments. *PLoS One* 2012;7(11):e49433.

#### Barnard 2012 {published data only}

Barnard M, Warren R, Gey Van Pittius N, van Helden P, Bosman M, Streicher E, et al. Genotype MTBDR<sub>s</sub><sub>l</sub> line probe assay shortens time to diagnosis of extensively drug-resistant tuberculosis in a high-throughput diagnostic laboratory. *American Journal of Respiratory and Critical Care Medicine* 2012;186(12):1298–305.

#### Brossier 2010a {published data only}

Brossier F, Veziris N, Aubry A, Jarlier V, Sougakoff W. Detection by GenoType MTBDR<sub>s</sub><sub>l</sub> test of complex mechanisms of resistance to second-line drugs and ethambutol in multidrug-resistant *Mycobacterium tuberculosis* complex isolates. *Journal of Clinical Microbiology* 2010;48(5):1683–9.

#### Catanzaro 2015 {published data only}

Catanzaro A, Rodwell TC, Catanzaro DG, Garfein RS, Jackson RL, Seifert M, et al. Performance comparison

of three rapid tests for the diagnosis of drug-resistant tuberculosis. *PLoS One* 2015;10(8):e0136861.

#### Chikamatsu 2012 {published data only}

Chikamatsu K, Aono A, Yamada H, Mitarai S. Evaluation of GenoType MTBDR<sub>s</sub><sub>l</sub> for testing resistance of *Mycobacterium tuberculosis* isolates to fluoroquinolone, aminoglycoside, and ethambutol. *Kekkaku [Tuberculosis]* 2012;87(10):641–7.

#### Fan 2011 {published data only}

Fan QW, Guo J, Zhang HZ, Wu XY, Hu XN, Qian XQ, et al. The characteristics of drug resistant relevant genes in multidrug-resistant and extensively drug-resistant tuberculosis by fast molecular assay. *Chinese Journal of Microbiology and Immunology* 2011;31(12):1133–7.

#### Ferro 2013 {published data only}

Ferro B, García P, Nieto LM, van Soolingen D. Predictive value of molecular drug resistance testing of *Mycobacterium tuberculosis* isolates in Valle del Cauca, Colombia. *Journal of Clinical Microbiology* 2013;51(7):2220–4.

#### Hillemann 2009 {published data only}

Hillemann D, Rüscher-Gerdes S, Richter E. Feasibility of the GenoType MTBDR<sub>s</sub><sub>l</sub> assay for fluoroquinolone, amikacin-capreomycin, and ethambutol resistance testing of *Mycobacterium tuberculosis* strains and clinical specimens. *Journal of Clinical Microbiology* 2009;47(6):1767–72.

- Huang 2011** *{published data only}*  
Huang WL, Chi TL, Wu MH, Jou R. Performance assessment of the GenoType MTBDRsl test and DNA sequencing for detection of second-line and ethambutol drug resistance among patients infected with multidrug-resistant *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology* 2011;**49**(7):2502-8.
- Ignatyeva 2012** *{published data only}*  
Ignatyeva O, Kontsevaya I, Kovalyov A, Balabanova Y, Nikolayevskyy V, Toit K, et al. Detection of resistance to second-line antituberculosis drugs by use of the genotype MTBDRsl assay: a multicenter evaluation and feasibility study. *Journal of Clinical Microbiology* 2012;**50**(5):1593-7.
- Jin 2013** *{published data only}*  
Jin J, Shen Y, Fan X, Diao N, Wang F, Wang S, et al. Underestimation of the resistance of *Mycobacterium tuberculosis* to second-line drugs by the new GenoType MTBDRsl test. *The Journal of Molecular Diagnostics* 2013;**15**(1):44-50.
- Kambli 2015a** *{published data only}*  
Kambli P, Ajbani K, Sadani M, Nikam C, Shetty A, Udwadia Z, et al. Correlating Minimum Inhibitory Concentrations of ofloxacin and moxifloxacin with gyrA mutations using the genotype MTBDRsl assay. *Tuberculosis (Edinb)* 2015;**95**(2):137-41.
- Kambli 2015b** *{published data only}*  
Kambli P, Ajbani K, Nikam C, Khillari A, Shetty A, Udwadia Z, et al. Determination of MICs of levofloxacin for *Mycobacterium tuberculosis* with gyrA mutations. *The International Journal of Tuberculosis and Lung Disease* 2015;**19**(10):1227-9.
- Kiet 2010** *{published data only}*  
Kiet VS, Lan NT, An DD, Dung NH, Hoa DV, van Vinh Chau N, et al. Evaluation of the MTBDRsl test for detection of second-line-drug resistance in *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology* 2010;**48**(8):2934-9.
- Kontsevaya 2011** *{published data only}*  
Kontsevaya I, Mironova S, Nikolayevskyy V, Balabanova Y, Mitchell S, Drobniewski F. Evaluation of two molecular assays for rapid detection of *Mycobacterium tuberculosis* resistance to fluoroquinolones in high-tuberculosis and -multidrug-resistance settings. *Journal of Clinical Microbiology* 2011;**49**(8):2832-7.
- Kontsevaya 2013** *{published data only}*  
Kontsevaya I, Ignatyeva O, Nikolayevskyy V, Balabanova Y, Kovalyov A, Kritsky A, et al. Diagnostic accuracy of the genotype MTBDRsl assay for rapid diagnosis of extensively drug-resistant tuberculosis in HIV-coinfected patients. *Journal of Clinical Microbiology* 2013;**51**(1):243-8.
- Lacoma 2012** *{published data only}*  
Lacoma A, García-Sierra N, Prat C, Maldonado J, Ruiz-Manzano J, Haba L, et al. GenoType MTBDRsl for molecular detection of second-line-drug and ethambutol resistance in *Mycobacterium tuberculosis* strains and clinical samples. *Journal of Clinical Microbiology* 2012;**50**(1):30-6.
- Lopez-Roa 2012** *{published data only}*  
López-Roa P, Ruiz-Serrano MJ, Alcalá L, García-Escribano Ráez N, García de Viedma D, Bouza E. Susceptibility testing to second-line drugs and ethambutol by GenoType MTBDRsl and Bactec MGIT 960 comparing with agar proportion method. *Tuberculosis* 2012;**92**(5):417-21.
- Miotto 2012** *{published data only}*  
Miotto P, Cabibbe AM, Mantegani P, Borroni E, Fattorini L, Tortoli E, et al. GenoType MTBDRsl performance on clinical samples with diverse genetic background. *European Respiratory Journal* 2012;**40**(3):690-8.
- Said 2012** *{published data only}*  
Said HM, Kock MM, Ismail NA, Baba K, Omar SV, Osman AG, et al. Evaluation of the GenoType® MTBDRsl assay for susceptibility testing of second-line anti-tuberculosis drugs. *International Journal of Tuberculosis and Lung Disease* 2012;**16**(1):104-9.
- Simons 2015** *{published data only}*  
Simons SO, van der Laan T, de Zwaan R, Kamst M, van Ingen J, Dekhuijzen PN, et al. Molecular drug susceptibility testing in the Netherlands: performance of the MTBDRplus and MTBDRsl assays. *The International Journal of Tuberculosis and Lung Disease* 2015;**19**(7):828-33.
- Surcouf 2011** *{published data only}*  
Surcouf C, Heng S, Pierre-Audigier C, Cadet-Daniel V, Namouchi A, Murray A, et al. Molecular detection of fluoroquinolone-resistance in multi-drug resistant tuberculosis in Cambodia suggests low association with XDR phenotypes. *BMC Infectious Diseases* 2011;**11**:255.
- Tagliani 2015** *{published data only}*  
Tagliani E, Cabibbe AM, Miotto P, Borroni E, Toro JC, Mansjö M, et al. Diagnostic performance of the new version (v2. 0) of GenoType MTBDRsl assay for detection of resistance to fluoroquinolones and second-line injectable drugs: a multicenter study. *Journal of Clinical Microbiology* 2015;**53**(9):2961-9.
- Tomasicchio 2016** *{published data only}*  
Tomasicchio M, Theron G, Pietersen E, Streicher E, Stanley-Josephs D, van Helden P, et al. The diagnostic accuracy of the MTBDRplus and MTBDRsl assays for drug-resistant TB detection when performed on sputum and culture isolates. *Scientific Reports* 2016;**6**:17850. DOI: 10.1038/srep17850
- Tukvadze 2014** *{published data only}*  
Tukvadze N, Bablishvili N, Apsindzelashvili R, Blumberg HM, Kempker RR. Performance of the MTBDRsl assay in Georgia. *The International Journal of Tuberculosis and Lung Disease* 2014;**18**(2):233-9.
- van Ingen 2010** *{published data only}*  
van Ingen J, Simons S, de Zwaan R, van der Laan T, Kamst-van Agterveld M, Boeree MJ, et al. Comparative study on genotypic and phenotypic second-line drug resistance testing of *Mycobacterium tuberculosis* complex isolates. *Journal of Clinical Microbiology* 2010;**48**(8):2749-53.
- Zivanovic 2012** *{published data only}*

Živanović I, Vuković D, Dakić I, Stefanović G, Savić B. Detection of drug-resistant *Mycobacterium tuberculosis* strains isolated in Serbia by the GenoType MTBDRsl assay. *Archives of Biological Sciences* 2012;64(4):1311–8.

## References to studies excluded from this review

### Aubry 2014 {published data only}

Aubry A, Sougakoff W, Bodzongo P, Delcroix G, Armand S, Millot G, et al. First evaluation of drug-resistant mycobacterium tuberculosis clinical isolates from Congo revealed misdetection of fluoroquinolone resistance by line probe assay due to a double substitution T80A-A90G in GyrA. *PLoS One* 2014;9(4):e95083.

### Babamahmoodi 2014 {published data only}

Babamahmoodi F, Mahdavi MR, Jalali H, Talebi B, Roshan P, Mahdavi M. Evaluation of gene mutations involved in drug resistance in *Mycobacterium tuberculosis* strains derived from tuberculosis patients in Mazandaran, Iran. *International Journal of Molecular and Cellular Medicine* 2013;3(3):190–5.

### Bantouna 2011 {published data only}

Bantouna V, Kontos F, Gitti Z, Spandidos DA. Evaluation of the new GenoType MTBDRsl (R) assay for rapid detection of resistance in *Mycobacterium tuberculosis* isolates at a low-incidence community. *International Journal of Molecular Medicine*. 2011; Vol. 28:576.

### Bergvala 2010 {published data only}

Bergvala L, Sengstakea S, Bachyiska E, Brankova N, Tankova K, Levterova V, et al. Evaluation of a new molecular test for the identification of drug resistance in *Mycobacterium tuberculosis* clinical isolates. *Problems of Infectious and Parasitic Diseases* 2010;38(2):27–30.

### Brossier 2010b {published data only}

Brossier F, Veziris N, Aubry A, Jarlier V, Sougakoff W. Detection by genotype MTBDRsl test of complex resistance mechanisms to second-line drugs and ethambutol in multidrug-resistant strains of *Mycobacterium tuberculosis*. *Clinical Microbiology and Infection*. 20th European Congress of Clinical Microbiology and Infectious Diseases (ECCMID); 2010 April 10-13; Vienna. 2010; Vol. 16: S53–4.

### Choi 2010 {published data only}

Choi JH, Lee KW, Kang HR, Hwang YI, Jang S, Kim DG, et al. Clinical efficacy of direct DNA sequencing analysis on sputum specimens for early detection of drug-resistant *Mycobacterium tuberculosis* in a clinical setting. *Chest* 2010; 137(2):393–400.

### Fallico 2012 {published data only}

Fallico L, Peracchi M, Khalil S, Rassu M, Pascarella M, Manganello R, et al. Molecular epidemiology and prevalence of mutations in *Mycobacterium tuberculosis* strains from an Italian northeastern area. *Clinical Microbiology and Infection*. 22nd European Congress of Clinical Microbiology and Infectious Diseases; 2012 Mar 31-Apr 3; London. 2012; Vol. 18:550.

### Felkel 2013 {published data only}

Felkel M, Exner R, Schleucher R, Lay H, Autenrieth I, Kempf VA, et al. Evaluation Of *Mycobacterium tuberculosis* drug susceptibility in clinical specimens from Nigeria using Genotype MTBDRplus and MTBDRsl assays. *European Journal of Microbiology and Immunology* 2013;3(4):252–7.

### Festoso 2011 {published data only}

Festoso I, Mantegani P, Miotto P, Cabibbe AM, Valente I, Borroni E, et al. Clinical evaluation of the GenoType MTBDRsl for the rapid diagnosis of second-line resistance in *Mycobacterium tuberculosis*. *Clinical Microbiology and Infection*. 21st ECCMID/27th ICC; 2011 May 7-10; Milan. 2011; Vol. 17:S592.

### Gkaravela 2012 {published data only}

Gkaravela L, Foka A, Athanassiou M, Lazarou N, Kontos F, Mollaki V, et al. The genetic diversity of *Mycobacterium tuberculosis* complex in Azerbaijan by 24 MIRU-VNTR loci genotyping in association with susceptibility testing by conventional and molecular methods. *Clinical Microbiology and Infection*. 22nd European Congress of Clinical Microbiology and Infectious Diseases; 2012 Mar 31-Apr 3; London. 2012; Vol. 18:829.

### Gomgnimbou 2015 {published data only}

Gomgnimbou M, Klotoe B, Molina B, Dominguez J, Refrégier G, Sola C. An “all-in-one” solution for simultaneous spoligotyping and drug resistance gene analysis of *Mycobacterium tuberculosis*: TB-SPRINT and TB-SPRINT<sub>plus</sub>. *International Journal of Mycobacteriology* 2015;4(Suppl 1):13–4.

### Iem 2013 {published data only}

Iem V, Somphavong S, Buisson Y, Steenkeste N, Breyse F, Chomarat M, et al. Resistance of *Mycobacterium tuberculosis* to antibiotics in Lao PDR: first multicentric study conducted in 3 hospitals. *BMC Infectious Diseases* 2013;13:275.

### Jang 2011 {published data only}

Jang E, Hwang S, Ryoo S. Practical value of the GenoType MTBDRsl assay for XDR testing of *Mycobacterium tuberculosis* from clinical isolates. *American Journal of Respiratory and Critical Care Medicine*. American Thoracic Society International Conference, ATS; 2011 May 13-18; Denver. 2011; Vol. 183:A1827.

### Karabela 2007 {published data only}

Karabela S, Nikolaou S, Kouseris I, Makrigiannis N, Kanavaki S. Extensively drug-resistant tuberculosis (XDR TB) in Athens. *International Journal of Antimicrobial Agents*. 17th ECCMID/25th ICC: 2007 March 31-Apr 3; Munich. 2007; Vol. 29 (Suppl 2):S574.

### Kaswa 2014 {published data only}

Kaswa MK, Aloni M, Nkuku L, Bakoko B, Lebeke R, Nzita A, et al. Pseudo-outbreak of pre-extensively drug-resistant (Pre-XDR) tuberculosis in Kinshasa: collateral damage caused by false detection of fluoroquinolone resistance by GenoType MTBDRsl. *Journal of Clinical Microbiology* 2014;52(8):2876–80.

- Kontos 2011** *{published data only}*  
Kontos F, Zande M, Mavromanolakis D, Bazigos S, Manidaki A, Gitri Z. Evaluation of the new GenoType MTBDRsl assay for rapid detection of resistance in *Mycobacterium tuberculosis* isolates in a low-incidence community. *Clinical Microbiology and Infection*. 21st ECCMID/27th ICC; 2011 May 7-10; Milan. 2011; Vol. 17:S587.
- Kontos 2012** *{published data only}*  
Kontos F, Bobola M, Mollaki V, Karonis T, Souli M, Dimopoulos G, et al. Tuberculosis: molecular susceptibility testing reveals mixed infections with three strains. *Clinical Microbiology and Infection*. 22nd European Congress of Clinical Microbiology and Infectious Diseases; 2012 Mar 31-Apr 3; London. 2012; Vol. 18:553-4.
- Lacoma 2015** *{published data only}*  
Lacoma A, Molina-Moya B, Prat C, Pimkina E, Diaz J, Dudnyk A, et al. Pyrosequencing for rapid detection of *Mycobacterium tuberculosis* second-line drugs and ethambutol resistance. *Diagnostic Microbiology and Infectious Disease* 2015;**83**(3):263-9.
- Lemus 2011** *{published data only}*  
Lemus D, Alvarez Y, Diaz R, Valdés I, Echemendía M, Martín A, et al. Molecular confirmation of resistance to first- and second-line antituberculosis drugs using GenoType MTBDRplus and GenoType MTBDRsl assays. *Clinical Microbiology and Infection*. 21st ECCMID/27th ICC; 2011 May 7-10; Milan. 2011; Vol. 17:S729.
- López-Roa 2010** *{published data only}*  
López Roa P, Ruiz-Serrano MJ, García-Escribano N, García de Viedma D, Bouza E. Rapid testing of susceptibility of multidrug-resistant *Mycobacterium tuberculosis* (MDRTB) to second-line drugs using GenoType MTBDRsl. Abstracts of the Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). 2010 Sept 12-15; Boston. 2010; Vol. 50.
- Mindru 2013** *{published data only}*  
Mindru R, Spînu V, Popa C, Botezatu E, Spînu R. [Conventional and molecular diagnosis on a group of patients with TB-DR]. *Pneumologia (Bucharest, Romania)* 2013;**63**(3):168-73.
- Molina-Moya 2015** *{published data only}*  
Molina-Moya B, Lacoma A, Prat C, Diaz J, Dudnyk A, Haba L, et al. AID TB resistance line probe assay for rapid detection of resistant *Mycobacterium tuberculosis* in clinical samples. *The Journal of Infection* 2015;**70**(4):400-8.
- Niehaus 2015** *{published data only}*  
Niehaus AJ, Mlisana K, Gandhi NR, Mathema B, Brust JC. High prevalence of inh A promoter mutations among patients with drug-resistant tuberculosis in KwaZulu-Natal, South Africa. *PLoS One* 2015;**10**(9):e0135003.
- Orikiriza 2015** *{published data only}*  
Orikiriza P, Tibenderana B, Siedner MJ, Mueller Y, Byarugaba F, Moore CC, et al. Low resistance to first and second line anti-tuberculosis drugs among treatment naive pulmonary tuberculosis patients in southwestern Uganda. *PLoS One* 2015;**10**(2):e0118191.
- Ouassa 2014** *{published data only}*  
Ouassa T, Loukou T, Dotia A, Faye-Kette H. Molecular characterization of the resistance of *Mycobacterium tuberculosis* to second line drugs in Côte d'Ivoire. *Tropical Journal of Pharmaceutical Research* 2014;**13**(5):727-30.
- Singh 2013** *{published data only}*  
Singh AK, Maurya AK, Kant S, Umrao J, Kushwaha RA, Nag VL, et al. Rapid detection of drug resistance and mutational patterns of extensively drug-resistant strains by a novel GenoType® MTBDRsl assay. *Journal of Postgraduate Medicine* 2013;**59**(3):179-85.
- Tessema 2012a** *{published data only}*  
Tessema B, Beer J, Emmrich F, Sack U, Rodloff AC. Analysis of gene mutations associated with isoniazid, rifampicin and ethambutol resistance among *Mycobacterium tuberculosis* isolates from Ethiopia. *BMC Infectious Diseases* 2012;**12**:37.
- Tessema 2012b** *{published data only}*  
Tessema B, Beer J, Emmrich F, Sack U, Rodloff AC. First- and second-line anti-tuberculosis drug resistance in Northwest Ethiopia. *The International Journal of Tuberculosis and Lung Disease* 2012;**16**(6):805-11.
- Totten 2011** *{published data only}*  
Totten S, May R, Heifets LB. Comparative study of two diagnostic tools for the rapid detection of *Mycobacterium tuberculosis* complex from smear positive respiratory specimens. American Journal of Respiratory and Critical Care Medicine. American Thoracic Society International Conference, ATS; 2011 May 13-18; Denver. 2011; Vol. 183:A4889.
- Walker 2015** *{published data only}*  
Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *The Lancet Infectious Diseases* 2015;**15**(10):1193-202.
- Wedajo 2014** *{published data only}*  
Wedajo W, Schön T, Bedru A, Kiros T, Hailu E, Mebrahtu T, et al. A 24-well plate assay for simultaneous testing of first and second line drugs against *Mycobacterium tuberculosis* in a high endemic setting. *BMC Research Notes* 2014;**7**:512.
- Zhang 2011** *{published data only}*  
Zhang L, Duo L, Lu X, Xie Y, Song X, Kang M, et al. Molecular epidemiology of multidrug-resistant (MDR) *Mycobacterium tuberculosis* revealed by GenoType MTBDR assays in southern China. *Clinical Microbiology and Infection*. 21st ECCMID/27th ICC; 2011 May 7-10; Milan. 2011; Vol. 17:S599-600.

## Additional references

- Balshem 2011**  
Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, et al. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 2011;**64**:401-6.

**Basu 2007**

Basu S, Andrews JR, Poolman EM, Gandhi NR, Shah NS, Moll A, et al. Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural South African district hospitals: an epidemiological modelling study. *The Lancet* 2007;**370**(9597):1500–7.

**Basu 2009**

Basu S, Friedland GH, Medlock J, Andrews JR, Shah NS, Gandhi NR, et al. Averting epidemics of extensively drug-resistant tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* 2009;**106**(18):7672–7.

**Bossuyt 2015**

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351**:h5527.

**Boyle 2015**

Boyle D, Pai M. *Tuberculosis Diagnostics Technology and Market Landscape*. 4th Edition. Geneva: World Health Organization (acting as the host organization for the Secretariat of UNITAID), 2015.

**CSLI 2011**

Woods GL, Brown-Elliott BA, Conville PS, Desmond EP, Hall GS, Lin G, et al. Susceptibility testing of mycobacteria, nocardia, and other aerobic actinomycetes: approved-standard. *Clinical and Laboratory Standards Institute (CLSI) document M24-A2*. 2nd Edition. Wayne: Clinical and Laboratory Standards Institute, 2011.

**Dowdy 2008**

Dowdy DW, Chaisson RE, Maartens G, Corbett EL, Dorman SE. Impact of enhanced tuberculosis diagnosis in South Africa: a mathematical model of expanded culture and drug susceptibility testing. *Proceedings of the National Academy of Sciences of the United States of America* 2008;**105**(32):11293–8.

**Feng 2013**

Feng Y, Liu S, Wang Q, Wang L, Tang S, Wang J, et al. Rapid diagnosis of drug resistance to fluoroquinolones, amikacin, capreomycin, kanamycin and ethambutol using genotype MTBDRsl assay: a meta-analysis. *PLoS One* 2013;**8**(2):e55292.

**Georghiou 2012**

Georghiou S, Magana M, Garfein RS, Catanzaro DG, Catanzaro A, Rodwell T. Evaluation of genetic mutations associated with *Mycobacterium tuberculosis* resistance to amikacin, kanamycin and capreomycin: a systematic review. *PLoS One* 2012;**7**(3):e33275.

**Gikalo 2012**

Gikalo MB, Nosova EY, Krylova LY, Moroz AM. The role of *eis* mutations in the development of kanamycin resistance in *Mycobacterium tuberculosis* isolates from the Moscow region. *Journal of Antimicrobial Chemotherapy* 2012;**67**(9):2107–9.

**GRADE 2013**

Schünemann H, Brozek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and

strength of recommendations. Updated October 2013. Available from [http://gdt.guidelinedevelopment.org/central\\_prod/\\_design/client/handbook/handbook.html](http://gdt.guidelinedevelopment.org/central_prod/_design/client/handbook/handbook.html). The GRADE Working Group, (accessed 21 April 2016).

**GRADEpro GDT 2015 [Computer program]**

McMaster University, (developed by Evidence Prime, Inc.). Available from [grade.pro](http://grade.pro). GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. McMaster University, (developed by Evidence Prime, Inc.). Available from [grade.pro](http://grade.pro), 2015.

**Hain Life Sciences 2015**

Hain Life Sciences. Rapid Diagnosis of Tuberculosis Brochure. [http://hain-lifescience.de/uploadfiles/file/produkte/mikrobiologie/mykobakterien/tb\\_eng.pdf](http://hain-lifescience.de/uploadfiles/file/produkte/mikrobiologie/mykobakterien/tb_eng.pdf) (accessed 23 December 2015).

**Hain Life Sciences 2015a**

Hain Life Sciences. GenoType MTBDRsl version 2.0. Instructions for use. 2015.

**Macaskill 2010**

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results In: Deeks JJ, Bossuyt PM, Gatsonis C (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0. The Cochrane Collaboration, 2010. Available from <http://srda.cochrane.org/>.

**Menzies 1997**

Menzies D. Effect of treatment on contagiousness of patients with active pulmonary tuberculosis. *Infection Control and Hospital Epidemiology* 1997;**18**(8):582–6.

**NIH 2015**

National Institutes of Health, National Human Genome Research Institute. Talking glossary of genetic terms. <http://www.genome.gov/glossary/> (accessed 16 November 2015).

**Orenstein 2009**

Orenstein EW, Basu S, Shah NS, Andrews JR, Friedland GH, Moll AP, et al. Treatment outcomes among patients with multidrug-resistant tuberculosis: systematic review and meta-analysis. *The Lancet Infectious Diseases* 2009;**9**(3):153–61.

**Pietersen 2014**

Pietersen E, Ignatius E, Streicher EM, Mastrapa B, Padanilam X, Pooran A, et al. Long-term outcomes of patients with extensively drug-resistant tuberculosis in South Africa: a cohort study. *The Lancet* 2014;**383**(9924):1230–9.

**Pooran 2013**

Pooran A, Pietersen E, Davids M, Theron G, Dheda K. What is the cost of diagnosis and management of drug resistant tuberculosis in South Africa?. *PLoS One* 2013;**8**(1):e54587.

**Reitsma 2005**

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in

- diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–90.
- RevMan 2014 [Computer program]**  
The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.
- Richter 2009**  
Richter E, Rüscher-Gerdes S, Hillemann D. Drug-susceptibility testing in TB: current status and future prospects. *Expert Review of Respiratory Medicine* 2009;**3**(5): 497–510.
- Rigouts 2016**  
Rigouts L, Coeck N, Gumusboga M, de Rijk WB, Aung KJ, Hossain MA, et al. Specific gyrA gene mutations predict poor treatment outcome in MDR-TB. *Journal of Antimicrobial Chemotherapy* 2016;**71**(2):314–23.
- Sirgel 2012a**  
Sirgel FA, Warren RM, Streicher EM, Victor TC, van Helden PD, Böttger EC. gyrA mutations and phenotypic susceptibility levels to ofloxacin and moxifloxacin in clinical isolates of *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy* 2012;**67**(5):1088–93.
- Sirgel 2012b**  
Sirgel FA, Tait M, Warren RM, Streicher EM, Böttger EC, van Helden PD, et al. Mutations in the rrs A1401G gene and phenotypic resistance to amikacin and capreomycin in *Mycobacterium tuberculosis*. *Microbial Drug Resistance* 2012; **18**(2):193–7.
- Stata 2015 [Computer program]**  
StataCorp. *Stata Statistical Software*. Version 14. College Station: StataCorp LP, 2015.
- Streicher 2011**  
Streicher EM, Müller B, Chihota V, Mlambo C, Tait M, Pillay M, et al. Emergence and treatment of multidrug resistant (MDR) and extensively drug-resistant (XDR) tuberculosis in South Africa. *Infection, Genetics and Evolution* 2011;**12**(4):686–94.
- Takwoingi 2013**  
Takwoingi Y, Leeftang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Annals of Internal Medicine* 2013;**158**(7):544–54.
- Takwoingi 2015**  
Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical Methods in Medical Research* 2015 June 26 Epub ahead of print]. DOI: 10.1177/0962280215592269
- Whiting 2011**  
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. *Annals of Internal Medicine* 2011;**155**(8):529–36.
- WHO 2011**  
World Health Organization. Rapid implementation of the Xpert MTB/RIF diagnostic test. Technical and operational 'How-to'. Practical considerations, 2011. [http://apps.who.int/iris/bitstream/10665/44593/1/9789241501569\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/44593/1/9789241501569_eng.pdf) (accessed 26 August 2016).
- WHO 2012**  
World Health Organization. Updated interim critical concentrations for first-line and second-line DST, 2012. [http://stoptb.org/wg/gli/assets/documents/Updated%20critical%20concentration%20table\\_1st%20and%202nd%20line%20drugs.pdf](http://stoptb.org/wg/gli/assets/documents/Updated%20critical%20concentration%20table_1st%20and%202nd%20line%20drugs.pdf). Geneva: World Health Organization, (accessed 27 June 2016).
- WHO 2013**  
World Health Organization. The use of molecular line probe assay for the detection of resistance to second-line anti-tuberculosis drugs. Expert group meeting report 2013. [http://apps.who.int/iris/bitstream/10665/78099/1/WHO\\_HTM\\_TB\\_2013.01\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/78099/1/WHO_HTM_TB_2013.01_eng.pdf). Geneva: World Health Organization, (accessed 27 June 2016).
- WHO 2014**  
World Health Organization. Definitions and reporting framework for tuberculosis - 2013 revision (updated December 2014). [http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345_eng.pdf). Geneva: World Health Organization, (accessed 27 June 2016).
- WHO 2015**  
World Health Organization. Global tuberculosis report 2015. [http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1). Geneva: World Health Organization, (accessed 27 June 2016).
- WHO 2016**  
World Health Organization. The use of molecular line probe assays for the detection of resistance to second-line anti-tuberculosis drugs, 2016. <http://www.who.int/tb/WHOPolicyStatementSLPA.pdf>. Geneva: World Health Organization, (accessed 27 June 2016).
- World Bank 2015**  
World Bank. World Bank list of economies (July 2015). <http://www.worldbank.org> (accessed 17 October 2015).
- Zhao 2012**  
Zhao Y, Xu SF, Wang LX, Chin DP, Wang SF, Jiang GL, et al. National survey of drug-resistant tuberculosis in China. *The New England Journal of Medicine* 2012;**366**:2161–70.
- Ängeby 2012**  
Ängeby K, Jureen P, Kahlmeter G, Hoffner S, Schön T. Challenging a dogma: antimicrobial susceptibility testing breakpoints for *Mycobacterium tuberculosis*. *Bulletin of the World Health Organization* 2012;**90**(9):693–8.
- References to other published versions of this review**
- Theron 2013**  
Theron G, Peter J, Barnard M, Donegan S, Warren R, Steingart KR, et al. The GenoType® MTBDRsl test for

resistance to second-line anti-tuberculosis drugs. *Cochrane Database of Systematic Reviews* 2013, Issue 8. DOI: 10.1002/14651858.CD010705

**Theron 2014**

Theron G, Peter J, Richardson M, Barnard M, Donegan S, Warren R, et al. The diagnostic accuracy of the GenoType® MTBDRsl assay for the detection of resistance to second-line anti-tuberculosis drugs. *Cochrane Database of Systematic Reviews* 2014, Issue 10. DOI: 10.1002/14651858.CD010705.pub2

\* Indicates the major publication for the study

## CHARACTERISTICS OF STUDIES

### Characteristics of included studies [ordered by study ID]

Ajbani 2012

| Study characteristics                                    |   |              |                        |
|--|---|--------------|------------------------|
| Patient sampling   | Cross-sectional design with consecutive enrolment of participants, prospective data collection  |              |                        |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>1. Country of origin: India.</li> <li>2. World Bank classification of country: middle.</li> <li>3. Type of lab: hospital.</li> <li>4. Type of patients: confirmed multidrug-resistant tuberculosis (MDR-TB) patients.</li> <li>5. Patients were smear-positive.</li> </ol>   |              |                        |
| Index tests  | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: yes, in design, analysis, or manuscript production.</li> <li>2. Type of testing: direct.</li> <li>3. Type of specimens: smear-positive.</li> <li>4. Specimen treatment: NALC-NaOH.</li> <li>5. Specimen condition: frozen.</li> <li>6. Duration of freezing: &lt; 1 year.</li> </ol>                      |              |                        |
| Target condition and reference standard(s)               | <ol style="list-style-type: none"> <li>1. Culture (liquid; MGIT 960) used for fluoroquinolone (FQ), second-line injectable drug (SLID).</li> <li>2. FQ drugs: ofloxacin (2 µg/mL) and moxifloxacin (0.25 µg/mL).</li> <li>3. SLIDs: amikacin (1 µg/mL), capreomycin (2.5 µg/mL), and kanamycin (2.5 µg/mL).</li> <li>4. Discrepant analysis: yes, with sequencing.</li> </ol> |              |                        |
| Flow and timing  | Uninterpretable results reported: yes   |              |                        |
| Comparative  |   |              |                        |
| Notes  | Also performed discrepant analysis; in this analysis, not all patients received both culture-based drug susceptibility testing (DST) and genetic sequencing reference standards   |              |                        |
| Methodological quality                                   |   |              |                        |
| Item   | Authors' judgement  | Risk of bias | Applicability concerns |
| DOMAIN 1: Patient Selection                              |   |              |                        |
| Was a consecutive or random sample of patients enrolled? | Yes   |              |                        |
| Was a case-control design avoided?                       | Yes   |              |                        |



**Ajbani 2012** (Continued)

|  |         |                |            |
|--|---------|----------------|------------|
| Did the study avoid inappropriate exclusions?  | Yes     |                |            |
|  |         | <b>Low</b>     | <b>Low</b> |
| <b>DOMAIN 2: Index Test All tests</b>  |         |                |            |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes     |                |            |
| If a threshold was used, was it pre-specified?   | Yes     |                |            |
|  |         | <b>Low</b>     | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |         |                |            |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |                |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |                |            |
|  |         | <b>Unclear</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |            |
| Did all patients receive the same reference standard?  | Yes     |                |            |
| Were all patients included in the analysis?  | Yes     |                |            |
|  |         | <b>Low</b>     |            |

**Barnard 2012**

| <b>Study characteristics</b>                             |  |                     |                               |
|--|--|---------------------|-------------------------------|
| Patient sampling   | Cross-sectional design with consecutive enrolment of participants, prospective data collection   |                     |                               |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>1. Country of origin: South Africa.</li> <li>2. World Bank classification of country: middle.</li> <li>3. Type of lab: reference.</li> <li>4. Type of patients: confirmed MDR-TB patients, confirmed rifampicin-mono-resistant patients, confirmed isoniazid-mono-resistant patients.</li> <li>5. Patients were smear-positive.</li> </ol>  |                     |                               |
| Index tests  | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: yes, reduced price.</li> <li>2. Type of testing: direct.</li> <li>3. Type of specimens: smear-positive.</li> <li>4. Specimen treatment: NALC-NaOH.</li> <li>5. Specimen condition: fresh.</li> <li>6. Tested after storage at room temperature or refrigerated within 48 hours of collection.</li> </ol>   |                     |                               |
| Target condition and reference standard(s)               | <ol style="list-style-type: none"> <li>1. Culture (solid; AP (agar proportion) method on 7H11) used for FQ, SLID.</li> <li>2. FQ drugs: ofloxacin (2 µg/mL).</li> <li>3. SLIDs: amikacin (4 µg/mL).</li> <li>4. Discrepant analysis: yes, with sequencing.</li> </ol>  |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes  |                     |                               |
| Comparative  |  |                     |                               |
| Notes  | <ol style="list-style-type: none"> <li>1. Reported performance on extrapulmonary tuberculosis specimens.</li> <li>2. Reported on the utility of the index test on specimens that were culture-contaminated (and hence could not receive a phenotypic DST).</li> <li>3. Reported on time-to-result.</li> <li>4. Also performed discrepant analysis; in this analysis, not all patients received both culture-based DST and genetic sequencing reference standards.</li> </ol> |                     |                               |
| <b>Methodological quality</b>                            |  |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>                       |  |                     |                               |
| Was a consecutive or random sample of patients enrolled? | Yes  |                     |                               |
| Was a case-control design avoided?                       | Yes  |                     |                               |
| Did the study avoid inappropriate exclusions?            | Yes  |                     |                               |

**Barnard 2012** (Continued)

|  |         |                |            |
|--|---------|----------------|------------|
|  |         | <b>Low</b>     | <b>Low</b> |
| <b>DOMAIN 2: Index Test All tests</b>  |         |                |            |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes     |                |            |
| If a threshold was used, was it pre-specified?   | Yes     |                |            |
|  |         | <b>Low</b>     | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |         |                |            |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |                |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |                |            |
|  |         | <b>Unclear</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |            |
| Did all patients receive the same reference standard?  | Yes     |                |            |
| Were all patients included in the analysis?  | Yes     |                |            |
|  |         | <b>Low</b>     |            |

**Brossier 2010a**

|                              |  |
|------------------------------|--|
| <b>Study characteristics</b> |  |
| Patient sampling             | Case-control design with unknown mechanism of enrolment of participants, prospective data collection |

**Brossier 2010a** (Continued)

|   |   |                     |                               |
|---|---|---------------------|-------------------------------|
| Patient characteristics and setting   | <ol style="list-style-type: none"> <li>Country of origin: France.</li> <li>World Bank classification of country: high.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed MDR-TB patients, confirmed extensively drug resistant tuberculosis (XDR-TB) patients, confirmed drug-susceptible TB patients.</li> </ol>   |                     |                               |
| Index tests   | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>   |                     |                               |
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>Culture (solid; agar proportion method on LJ (Löwenstein-Jensen) and sequencing used for FQ, SLID.</li> <li>FQ drugs: ofloxacin (2 µg/mL).</li> <li>SLIDs: amikacin (20 µg/mL), kanamycin (20 µg/mL), capreomycin (20 µg/mL).</li> <li>Genes sequenced for FQ: <i>gyrA</i> and <i>gyrB</i>.</li> <li>Genes sequenced for SLIDs: <i>rrs</i>.</li> <li>Discrepant analysis: no.</li> </ol> |                     |                               |
| Flow and timing   | Uninterpretable results reported: yes   |                     |                               |
| Comparative   |   |                     |                               |
| Notes   |   |                     |                               |
| <b>Methodological quality</b>   |   |                     |                               |
| <b>Item</b>   | <b>Authors' judgement</b>   | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>  |   |                     |                               |
| Was a consecutive or random sample of patients enrolled?  | Yes   |                     |                               |
| Was a case-control design avoided?  | No  |                     |                               |
| Did the study avoid inappropriate exclusions?   | No  |                     |                               |
|   |   | <b>High</b>         | <b>High</b>                   |
| <b>DOMAIN 2: Index Test All tests</b>   |   |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Unclear   |                     |                               |

**Brossier 2010a** (Continued)

|  |         |         |         |
|--|---------|---------|---------|
| If a threshold was used, was it pre-specified?   | Yes     |         |         |
|  |         | Unclear | Unclear |
| <b>DOMAIN 3: Reference Standard</b>  |         |         |         |
| Is the reference standards likely to correctly classify the target condition?                        | No      |         |         |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |         |         |
|  |         | High    | Low     |
| <b>DOMAIN 4: Flow and Timing</b>   |         |         |         |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |         |         |
| Did all patients receive the same reference standard?  | Yes     |         |         |
| Were all patients included in the analysis?  | Yes     |         |         |
|  |         | Low     |         |

**Catanzaro 2015**

|                                     |  |
|-------------------------------------|--|
| <b>Study characteristics</b>        |  |
| Patient sampling                    | Prospective, cross-sectional study that enrolled consecutively   |
| Patient characteristics and setting | <ol style="list-style-type: none"> <li>Country of origin: Manila, the Philippines; Mumbai, India; Chisinau, Moldova; and Port Elizabeth, South Africa.</li> <li>World Bank classification of country: low- and middle-income.</li> <li>Type of lab: unknown.</li> <li>Type of patients: patients failing TB therapy or possible MDR-TB.</li> </ol> |
| Index tests                         | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> <li>The test was applied to pooled sputum specimens.</li> </ol>  |

|   |  |                     |                               |
|---|--|---------------------|-------------------------------|
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>1. Culture (MGIT (Mycobacteria Growth Indicator Tube) 960).</li> <li>2. FQ drugs: ofloxacin (2 µg/mL) moxifloxacin (0.25 µg/mL).</li> <li>3. SLIDs: amikacin (1 µg/mL), kanamycin (2.5 µg/mL), capreomycin (2.5 µg/mL).</li> <li>4. Discrepant analysis: no.</li> </ol> |                     |                               |
| Flow and timing   | Uninterpretable results reported: yes  |                     |                               |
| Comparative   |  |                     |                               |
| Notes   | The test was applied to pooled sputum specimens  |                     |                               |
| <b>Methodological quality</b>   |  |                     |                               |
| <b>Item</b>   | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>  |  |                     |                               |
| Was a consecutive or random sample of patients enrolled?  | Yes  |                     |                               |
| Was a case-control design avoided?  | Yes  |                     |                               |
| Did the study avoid inappropriate exclusions?   | Yes  |                     |                               |
|   |  | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>   |  |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Unclear  |                     |                               |
| If a threshold was used, was it pre-specified?  | Yes  |                     |                               |
|   |  | <b>Unclear</b>      | <b>High</b>                   |
| <b>DOMAIN 3: Reference Standard</b>   |  |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                       | Unclear  |                     |                               |
| Were the reference standard results interpreted without knowledge                                   | Yes  |                     |                               |

Catanzaro 2015 (Continued)

|  |     |         |     |
|--|-----|---------|-----|
| of the results of the index tests?   |     |         |     |
|  |     | Unclear | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |     |         |     |
| Was there an appropriate interval between index test and reference standard? | Yes |         |     |
| Did all patients receive the same reference standard?                        | Yes |         |     |
| Were all patients included in the analysis?                                  | Yes |         |     |
|  |     | Low     |     |

Chikamatsu 2012

|  |  |
|--|--|
| <b>Study characteristics</b>               |  |
| Patient sampling                           | Cross-sectional design with unknown mechanism of enrolment of participants, unknown direction of data collection   |
| Patient characteristics and setting        | <ol style="list-style-type: none"> <li>Country of origin: Japan.</li> <li>World Bank classification of country: high.</li> <li>Type of lab: reference.</li> <li>Type of patients: unknown.</li> </ol>  |
| Index tests                                | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>  |
| Target condition and reference standard(s) | <ol style="list-style-type: none"> <li>Culture (solid; Ogawa solid culture for FQs, unclear for SLIDs) and sequencing used for FQ, SLID.</li> <li>FQ drugs: levofloxacin (1 µg/mL).</li> <li>SLIDs: amikacin (unknown concentration), kanamycin (unknown concentration), capreomycin (unknown concentration).</li> <li>Genes sequenced for FQ: <i>gyrA</i>.</li> <li>Genes sequenced for SLIDs: <i>rrs</i>.</li> <li>Discrepant analysis: no.</li> </ol> |
| Flow and timing                            | Uninterpretable results reported: yes  |
| Comparative                                |  |
| Notes                                      |  |

| <b>Methodological quality</b>  |                           |                     |                               |
|--|---------------------------|---------------------|-------------------------------|
| <b>Item</b>  | <b>Authors' judgement</b> | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>   |                           |                     |                               |
| Was a consecutive or random sample of patients enrolled?   | Unclear                   |                     |                               |
| Was a case-control design avoided?   | Yes                       |                     |                               |
| Did the study avoid inappropriate exclusions?  | Yes                       |                     |                               |
|  |                           | <b>Unclear</b>      | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>  |                           |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | No                        |                     |                               |
| If a threshold was used, was it pre-specified?   | Yes                       |                     |                               |
|  |                           | <b>High</b>         | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>  |                           |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                        | No                        |                     |                               |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes                       |                     |                               |
|  |                           | <b>High</b>         | <b>Low</b>                    |
| <b>DOMAIN 4: Flow and Timing</b>   |                           |                     |                               |
| Was there an appropriate interval between index test and reference standard?                         | Yes                       |                     |                               |



**Chikamatsu 2012** (Continued)

|   |     |            |  |
|---|-----|------------|--|
| Did all patients receive the same reference standard? | Yes |            |  |
| Were all patients included in the analysis?           | Yes |            |  |
|   |     | <b>Low</b> |  |

**Fan 2011**

| <b>Study characteristics</b>                             |  |                     |                               |
|--|--|---------------------|-------------------------------|
| Patient sampling   | Cross-sectional design with enrolment of participants by convenience, prospective data collection  |                     |                               |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>Country of origin: China.</li> <li>World Bank classification of country: middle.</li> <li>Type of lab: research.</li> <li>Type of patients: confirmed MDR-TB patients and confirmed XDR-TB patients.</li> </ol> |                     |                               |
| Index tests  | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>  |                     |                               |
| Target condition and reference standard(s)               | <ol style="list-style-type: none"> <li>Culture (liquid; MGIT 960).</li> <li>FQ drugs: ofloxacin (2 µg/mL), moxifloxacin (0.25 µg/mL).</li> <li>SLIDs: amikacin (1 µg/mL).</li> <li>Discrepant analysis: no.</li> </ol>                                 |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes  |                     |                               |
| Comparative  |  |                     |                               |
| Notes  |  |                     |                               |
| <b>Methodological quality</b>                            |  |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>                       |  |                     |                               |
| Was a consecutive or random sample of patients enrolled? | No   |                     |                               |
| Was a case-control design avoided?                       | Yes  |                     |                               |
| Did the study avoid inappropriate exclusions?            | Yes  |                     |                               |

|  |         | High    | Low |
|--|---------|---------|-----|
| <b>DOMAIN 2: Index Test All tests</b>  |         |         |     |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Unclear |         |     |
| If a threshold was used, was it pre-specified?   | Yes     |         |     |
|  |         | Unclear | Low |
| <b>DOMAIN 3: Reference Standard</b>  |         |         |     |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |         |     |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |         |     |
|  |         | Unclear | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |         |         |     |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |         |     |
| Did all patients receive the same reference standard?  | Yes     |         |     |
| Were all patients included in the analysis?  | Yes     |         |     |
|  |         | Low     |     |

**Ferro 2013**

| <b>Study characteristics</b> |   |
|------------------------------|---|
| Patient sampling             | Cross-sectional design with random enrolment of participants, prospective data collection |

|   |  |                     |                               |
|---|--|---------------------|-------------------------------|
| Patient characteristics and setting   | <ol style="list-style-type: none"> <li>Country of origin: Colombia.</li> <li>World Bank classification of country: middle.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed drug-susceptible TB patients, MDR-TB patients, MDR-TB patients with some known second-line resistance and XDR-TB patients.</li> </ol> |                     |                               |
| Index tests   | <ol style="list-style-type: none"> <li>Manufacturer involvement: yes, donation of test.</li> <li>Type of testing: indirect.</li> </ol>   |                     |                               |
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>Culture-based DST (solid, 7H10).</li> <li>FQ drugs: moxifloxacin 2 µg/mL.</li> <li>SLIDS: amikacin 5 µg/mL, kanamycin 5 µg/mL.</li> <li>No XDR-TB information reported.</li> <li>There was no discrepant analysis.</li> </ol>   |                     |                               |
| Flow and timing   | Not all participants were accounted for in the analyses. Uninterpretable results reported: yes   |                     |                               |
| Comparative   |  |                     |                               |
| Notes   |  |                     |                               |
| <b>Methodological quality</b>   |  |                     |                               |
| <b>Item</b>   | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>  |  |                     |                               |
| Was a consecutive or random sample of patients enrolled?  | Yes  |                     |                               |
| Was a case-control design avoided?  | Yes  |                     |                               |
| Did the study avoid inappropriate exclusions?   | Unclear  |                     |                               |
|   |  | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>   |  |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Unclear  |                     |                               |
| If a threshold was used, was it pre-specified?  | Yes  |                     |                               |
|   |  | <b>Unclear</b>      | <b>Low</b>                    |

| <b>DOMAIN 3: Reference Standard</b>  |         |                |            |
|--|---------|----------------|------------|
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |                |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |                |            |
|  |         | <b>Unclear</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |            |
| Did all patients receive the same reference standard?  | Yes     |                |            |
| Were all patients included in the analysis?  | Unclear |                |            |
|  |         | <b>Unclear</b> |            |

**Hillemann 2009**

| <b>Study characteristics</b>        |  |
|-------------------------------------|--|
| Patient sampling                    | Case-control design with the consecutive enrolment of participants, prospective data collection for clinical specimens, retrospective for culture isolates   |
| Patient characteristics and setting | <ol style="list-style-type: none"> <li>1. Country of origin: Germany.</li> <li>2. World Bank classification of country: high.</li> <li>3. Type of lab: reference.</li> <li>4. Type of patients: confirmed XDR-TB patients, confirmed drug-susceptible TB patients.</li> <li>5. The specimens tested were smear positive and smear negative.</li> </ol> |
| Index tests                         | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: yes, donation of tests.</li> <li>2. Type of testing: direct and indirect.</li> <li>3. Type of specimens: smear-positive.</li> <li>4. Specimen treatment: NALC-NaOH.</li> <li>5. Specimen condition: frozen.</li> <li>6. Duration of freezing: &gt; 1 year.</li> </ol>              |

|   |  |                     |                               |
|---|--|---------------------|-------------------------------|
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>1. Culture (liquid and solid; MGIT 960 and LJ) and sequencing used for FQ, SLID.</li> <li>2. FQ drugs: ofloxacin (2 µg/mL for MGIT and LJ).</li> <li>3. SLIDs: amikacin 1 µg/mL for MGIT and 40.0 µg/mL for LJ; capreomycin 2.5 µg/mL for MGIT and 40.0 µg/mL for LJ.</li> <li>4. Genes sequenced for FQ: <i>gyrA</i>.</li> <li>5. Genes sequenced for SLIDs: <i>rrs</i>.</li> <li>6. Discrepant analysis: no.</li> </ol> |                     |                               |
| Flow and timing   | Uninterpretable results reported: yes  |                     |                               |
| Comparative   |  |                     |                               |
| Notes   |  |                     |                               |
| <b>Methodological quality</b>   |  |                     |                               |
| <b>Item</b>   | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>  |  |                     |                               |
| Was a consecutive or random sample of patients enrolled?  | Yes  |                     |                               |
| Was a case-control design avoided?  | No   |                     |                               |
| Did the study avoid inappropriate exclusions?   | Yes  |                     |                               |
|   |  | <b>High</b>         | <b>High</b>                   |
| <b>DOMAIN 2: Index Test All tests</b>   |  |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Yes  |                     |                               |
| If a threshold was used, was it pre-specified?  | Yes  |                     |                               |
|   |  | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>   |  |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                       | Unclear  |                     |                               |

Hillemann 2009 (Continued)

|  |     |         |     |
|--|-----|---------|-----|
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes |         |     |
|  |     | Unclear | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |     |         |     |
| Was there an appropriate interval between index test and reference standard?                         | Yes |         |     |
| Did all patients receive the same reference standard?  | Yes |         |     |
| Were all patients included in the analysis?  | Yes |         |     |
|  |     | Low     |     |

Huang 2011

|  |  |
|--|--|
| <b>Study characteristics</b>               |  |
| Patient sampling                           | Cross-sectional design with consecutive enrolment of participants, prospective data collection   |
| Patient characteristics and setting        | <ol style="list-style-type: none"> <li>Country of origin: China.</li> <li>World Bank classification of country: middle.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed MDR-TB patients.</li> </ol>  |
| Index tests                                | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>  |
| Target condition and reference standard(s) | <ol style="list-style-type: none"> <li>Culture (solid; 7H11 and liquid; MGIT 960) and sequencing used for FQ, SLID.</li> <li>FQ drugs: ofloxacin (2 µg/mL).</li> <li>SLIDs (for 7H11): amikacin (6 µg/mL), kanamycin (6 µg/mL) and capreomycin (10 µg/mL).</li> <li>Genes sequenced for FQ: <i>gyrA</i> and <i>gyrB</i>.</li> <li>Genes sequenced for SLIDs: <i>rrs</i> and <i>eis</i>.</li> <li>Discrepant analysis: no.</li> </ol> |
| Flow and timing                            | Uninterpretable results reported: yes  |
| Comparative                                |  |
| Notes                                      |  |

| <b>Methodological quality</b>  |                           |                     |                               |
|--|---------------------------|---------------------|-------------------------------|
| <b>Item</b>  | <b>Authors' judgement</b> | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>   |                           |                     |                               |
| Was a consecutive or random sample of patients enrolled?   | Yes                       |                     |                               |
| Was a case-control design avoided?   | Yes                       |                     |                               |
| Did the study avoid inappropriate exclusions?  | Yes                       |                     |                               |
|  |                           | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>  |                           |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes                       |                     |                               |
| If a threshold was used, was it pre-specified?   | Yes                       |                     |                               |
|  |                           | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>  |                           |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear                   |                     |                               |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes                       |                     |                               |
|  |                           | <b>Unclear</b>      | <b>Low</b>                    |
| <b>DOMAIN 4: Flow and Timing</b>   |                           |                     |                               |
| Was there an appropriate interval between index test and reference standard?                         | Yes                       |                     |                               |

**Huang 2011** (Continued)

|   |     |            |  |
|---|-----|------------|--|
| Did all patients receive the same reference standard? | Yes |            |  |
| Were all patients included in the analysis?           | Yes |            |  |
|   |     | <b>Low</b> |  |

**Ignatyeva 2012**

| <b>Study characteristics</b>                             |   |                     |                               |
|--|---|---------------------|-------------------------------|
| Patient sampling   | Case-control design with consecutive enrolment of participants, prospective data collection   |                     |                               |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>Country of origin: Estonia.</li> <li>World Bank classification of country: high.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed MDR-TB patients, confirmed XDR-TB patients and confirmed drug-susceptible TB patients.</li> </ol> |                     |                               |
| Index tests  | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>   |                     |                               |
| Target condition and reference standard(s)               | <ol style="list-style-type: none"> <li>Culture (liquid; MGIT 960) used for FQ, SLID.</li> <li>FQ drugs: ofloxacin (2 µg/mL).</li> <li>SLIDs: amikacin (1 µg/mL), kanamycin (5 µg/mL), and capreomycin (2.5 µg/mL).</li> </ol>   |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes   |                     |                               |
| Comparative  |   |                     |                               |
| Notes  | Other findings: the interpretability of the GenoType® MTBDR <sub>sl</sub> assay was high, varying between 98.0% and 100% for the first reading and between 95.5% and 100% for the second reading (Table 3)  |                     |                               |
| <b>Methodological quality</b>                            |   |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>   | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>                       |   |                     |                               |
| Was a consecutive or random sample of patients enrolled? | Yes   |                     |                               |
| Was a case-control design avoided?                       | No  |                     |                               |



**Ignatyeva 2012** (Continued)

|  |         |                |             |
|--|---------|----------------|-------------|
| Did the study avoid inappropriate exclusions?  | Yes     |                |             |
|  |         | <b>High</b>    | <b>High</b> |
| <b>DOMAIN 2: Index Test All tests</b>  |         |                |             |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes     |                |             |
| If a threshold was used, was it pre-specified?   | Yes     |                |             |
|  |         | <b>Low</b>     | <b>Low</b>  |
| <b>DOMAIN 3: Reference Standard</b>  |         |                |             |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |                |             |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |                |             |
|  |         | <b>Unclear</b> | <b>Low</b>  |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |             |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |             |
| Did all patients receive the same reference standard?  | Yes     |                |             |
| Were all patients included in the analysis?  | Yes     |                |             |
|  |         | <b>Low</b>     |             |

| Study characteristics   |   |              |                        |
|---|---|--------------|------------------------|
| Patient sampling  | Cross-sectional design with consecutive enrolment of participants, prospective data collection  |              |                        |
| Patient characteristics and setting   | <ol style="list-style-type: none"> <li>1. Country of origin: China.</li> <li>2. World Bank classification of country: middle.</li> <li>3. Type of lab: reference.</li> <li>4. Type of patients: confirmed MDR-TB patients and confirmed XDR-TB patients.</li> </ol>   |              |                        |
| Index tests   | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: no.</li> <li>2. Type of testing: indirect.</li> </ol>   |              |                        |
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>1. Culture (solid; LJ) and BacT/ALERT 3D and sequencing used for FQ, SLID.</li> <li>2. FQ drugs: ofloxacin (5 and 50 µg/mL).</li> <li>3. SLIDs: kanamycin (10 µg/mL), capreomycin unknown.</li> <li>4. Genes sequenced for FQ: <i>gyrA</i>.</li> <li>5. Genes sequenced for SLIDs: <i>rrs</i>.</li> <li>6. Discrepant analysis: no.</li> </ol> |              |                        |
| Flow and timing   | Uninterpretable results reported: yes   |              |                        |
| Comparative   |   |              |                        |
| Notes   |   |              |                        |
| Methodological quality  |   |              |                        |
| Item  | Authors' judgement  | Risk of bias | Applicability concerns |
| DOMAIN 1: Patient Selection   |   |              |                        |
| Was a consecutive or random sample of patients enrolled?  | Yes   |              |                        |
| Was a case-control design avoided?  | Yes   |              |                        |
| Did the study avoid inappropriate exclusions?   | Yes   |              |                        |
|   |   | Low          | Low                    |
| DOMAIN 2: Index Test All tests  |   |              |                        |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Yes   |              |                        |

|  |     |             |            |
|--|-----|-------------|------------|
| If a threshold was used, was it pre-specified?   | Yes |             |            |
|  |     | <b>Low</b>  | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |     |             |            |
| Is the reference standards likely to correctly classify the target condition?                        | No  |             |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes |             |            |
|  |     | <b>High</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |     |             |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes |             |            |
| Did all patients receive the same reference standard?  | Yes |             |            |
| Were all patients included in the analysis?  | Yes |             |            |
|  |     | <b>Low</b>  |            |

**Kambli 2015a**

|  |   |
|--|---|
| <b>Study characteristics</b>               |   |
| Patient sampling                           | Cross-sectional study that enrolled consecutively   |
| Patient characteristics and setting        | <ol style="list-style-type: none"> <li>Country of origin: India.</li> <li>World Bank classification of country: middle.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed MDR-TB patients.</li> </ol> |
| Index tests                                | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>   |
| Target condition and reference standard(s) | MGIT 960<br>FQ drugs: ofloxacin (2 µg/mL), moxifloxacin (0.25 µg/mL)  |

**Kambli 2015a** (Continued)

|  |                                      |                     |                               |
|--|--------------------------------------|---------------------|-------------------------------|
| Flow and timing  | Uninterpretable results reported: no |                     |                               |
| Comparative  |                                      |                     |                               |
| Notes  |                                      |                     |                               |
| <b>Methodological quality</b>  |                                      |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>            | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>   |                                      |                     |                               |
| Was a consecutive or random sample of patients enrolled?   | Yes                                  |                     |                               |
| Was a case-control design avoided?   | Yes                                  |                     |                               |
| Did the study avoid inappropriate exclusions?  | Yes                                  |                     |                               |
|  |                                      | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>  |                                      |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes                                  |                     |                               |
| If a threshold was used, was it pre-specified?   | Yes                                  |                     |                               |
|  |                                      | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>  |                                      |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear                              |                     |                               |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes                                  |                     |                               |
|  |                                      | <b>Unclear</b>      | <b>Low</b>                    |
| <b>DOMAIN 4: Flow and Timing</b>   |                                      |                     |                               |

**Kambli 2015a** (Continued)

|  |     |            |  |
|--|-----|------------|--|
| Was there an appropriate interval between index test and reference standard? | Yes |            |  |
| Did all patients receive the same reference standard?                        | Yes |            |  |
| Were all patients included in the analysis?                                  | Yes |            |  |
|  |     | <b>Low</b> |  |

**Kambli 2015b**

| <b>Study characteristics</b>                             |  |                     |                               |
|--|--|---------------------|-------------------------------|
| Patient sampling   | Cross-sectional study that enrolled consecutively  |                     |                               |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>1. Country of origin: India.</li> <li>2. World Bank classification of country: middle.</li> <li>3. Type of lab: reference.</li> <li>4. Type of patients: confirmed MDR-TB cases.</li> </ol> |                     |                               |
| Index tests  | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: no.</li> <li>2. Type of testing: indirect.</li> </ol>  |                     |                               |
| Target condition and reference standard(s)               | MGIT 960<br>Levofloxacin (1.5 µg/mL)   |                     |                               |
| Flow and timing  | Uninterpretable results reported: no   |                     |                               |
| Comparative  |  |                     |                               |
| Notes  |  |                     |                               |
| <b>Methodological quality</b>                            |  |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>                       |  |                     |                               |
| Was a consecutive or random sample of patients enrolled? | Yes  |                     |                               |
| Was a case-control design avoided?                       | Yes  |                     |                               |

**Kambli 2015b** (Continued)

|  |     |            |            |
|--|-----|------------|------------|
| Did the study avoid inappropriate exclusions?  | Yes |            |            |
|  |     | <b>Low</b> | <b>Low</b> |
| <b>DOMAIN 2: Index Test All tests</b>  |     |            |            |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes |            |            |
| If a threshold was used, was it pre-specified?   | Yes |            |            |
|  |     | <b>Low</b> | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |     |            |            |
| Is the reference standards likely to correctly classify the target condition?                        | Yes |            |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes |            |            |
|  |     | <b>Low</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |     |            |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes |            |            |
| Did all patients receive the same reference standard?  | Yes |            |            |
| Were all patients included in the analysis?  | Yes |            |            |
|  |     | <b>Low</b> |            |

## Kiet 2010

| Study characteristics   |   |              |                        |
|---|---|--------------|------------------------|
| Patient sampling  | Case-control design with consecutive enrolment of participants, prospective data collection   |              |                        |
| Patient characteristics and setting   | <ol style="list-style-type: none"> <li>1. Country of origin: Vietnam.</li> <li>2. World Bank classification of country: middle.</li> <li>3. Type of lab: reference.</li> <li>4. Type of patients: confirmed MDR-TB patients with FQ resistance, confirmed FQ monoresistant patients.</li> </ol>                     |              |                        |
| Index tests   | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: no.</li> <li>2. Type of testing: indirect.</li> </ol>   |              |                        |
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>1. Culture (solid; LJ) used for FQ, SLID.</li> <li>2. FQ drugs: ofloxacin (2 µg/mL).</li> <li>3. SLIDs: kanamycin (20 µg/mL), not World Health Organization (WHO)-recommended critical concentrations for LJ solid culture.</li> <li>4. Discrepant analysis: yes.</li> </ol> |              |                        |
| Flow and timing   | Uninterpretable results reported: yes   |              |                        |
| Comparative   |   |              |                        |
| Notes   | Also performed discrepant analysis; in this analysis, not all patients received both culture-based DST and genetic sequencing reference standards   |              |                        |
| Methodological quality  |   |              |                        |
| Item  | Authors' judgement  | Risk of bias | Applicability concerns |
| DOMAIN 1: Patient Selection   |   |              |                        |
| Was a consecutive or random sample of patients enrolled?  | Yes   |              |                        |
| Was a case-control design avoided?  | No  |              |                        |
| Did the study avoid inappropriate exclusions?   | Yes   |              |                        |
|   |   | High         | High                   |
| DOMAIN 2: Index Test All tests  |   |              |                        |
| Were the index test results interpreted without knowledge of the results of the reference standard? | No  |              |                        |

**Kiet 2010** (Continued)

|  |     |             |            |
|--|-----|-------------|------------|
| If a threshold was used, was it pre-specified?   | Yes |             |            |
|  |     | <b>High</b> | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |     |             |            |
| Is the reference standards likely to correctly classify the target condition?                        | No  |             |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | No  |             |            |
|  |     | <b>High</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |     |             |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes |             |            |
| Did all patients receive the same reference standard?  | Yes |             |            |
| Were all patients included in the analysis?  | Yes |             |            |
|  |     | <b>Low</b>  |            |

**Kontsevaya 2011**

|                                     |  |
|-------------------------------------|--|
| <b>Study characteristics</b>        |  |
| Patient sampling                    | Cross-sectional design with consecutive enrolment of participants, prospective data collection   |
| Patient characteristics and setting | <ol style="list-style-type: none"> <li>Country of origin: UK.</li> <li>World Bank classification of country: high.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed MDR-TB patients.</li> </ol> |
| Index tests                         | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>  |



**Kontsevaya 2011** (Continued)

|  |   |                     |                               |
|--|---|---------------------|-------------------------------|
| Target condition and reference standard(s)   | 1. Culture (liquid; MGIT 960) used for FQ.<br>2. FQ drugs: ofloxacin (2 µg/mL), moxifloxacin (0.25 µg/mL).<br>3. Discrepant analysis: no. |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes   |                     |                               |
| Comparative  |   |                     |                               |
| Notes  |   |                     |                               |
| <b>Methodological quality</b>  |   |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>   | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>   |   |                     |                               |
| Was a consecutive or random sample of patients enrolled?   | Yes   |                     |                               |
| Was a case-control design avoided?   | Yes   |                     |                               |
| Did the study avoid inappropriate exclusions?  | Yes   |                     |                               |
|  |   | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>  |   |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes   |                     |                               |
| If a threshold was used, was it pre-specified?   | Yes   |                     |                               |
|  |   | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>  |   |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear   |                     |                               |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes   |                     |                               |

**Kontsevaya 2011** (Continued)

|  |     |         |     |
|--|-----|---------|-----|
|  |     | Unclear | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |     |         |     |
| Was there an appropriate interval between index test and reference standard? | Yes |         |     |
| Did all patients receive the same reference standard?                        | Yes |         |     |
| Were all patients included in the analysis?                                  | Yes |         |     |
|  |     | Low     |     |

**Kontsevaya 2013**

|  |  |
|--|--|
| <b>Study characteristics</b>               |  |
| Patient sampling                           | Cross-sectional design with consecutive enrolment of participants, prospective data collection   |
| Patient characteristics and setting        | <ol style="list-style-type: none"> <li>Country of origin: Russia.</li> <li>World Bank classification of country: high.</li> <li>Type of lab: unknown.</li> <li>Type of patients: confirmed MDR-TB cases.</li> <li>Median age: 35.</li> <li>All HIV-infected.</li> <li>Previous TB: 38/90.</li> <li>Male: 71/90.</li> </ol> |
| Index tests                                | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: direct.</li> <li>Type of specimens: smear-positive.</li> <li>Specimen treatment: unknown.</li> <li>Specimen condition: unknown.</li> <li>Duration of freezing: unknown.</li> </ol>  |
| Target condition and reference standard(s) | <ol style="list-style-type: none"> <li>Culture (liquid; MGIT 960) used for FQ, SLID.</li> <li>FQ drugs: ofloxacin (2 µg/mL) and moxifloxacin (0.25 µg/mL).</li> <li>SLIDs: kanamycin (5 µg/mL), amikacin (1 µg/mL), and capreomycin (2.5 µg/mL).</li> <li>Discrepant analysis: no.</li> </ol>                              |
| Flow and timing                            | Uninterpretable results reported: no   |
| Comparative                                |  |

|  |  |                     |                               |
|--|--|---------------------|-------------------------------|
| Notes  | Other findings: analysis of test performance stratified according to sputum smear positivity showed that the test readability for individual drugs and their drug groups ranged from 80.0% to 100.0%, with the lowest for specimens graded 1 (Table 5). Within this group of specimens, lower readability rates were observed for the AG/CP group of drugs (20.0% of tests failed), with higher readability rates for FQ and ethambutol. Similar trends were observed in specimens graded 2 and 3 (Figure 1). Total agreement between the molecular assay and phenotypic DST was the highest (84.1%) for FQs and lowest (23.5%) for the injectable drugs |                     |                               |
| <b>Methodological quality</b>  |  |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>   |  |                     |                               |
| Was a consecutive or random sample of patients enrolled?   | Yes  |                     |                               |
| Was a case-control design avoided?   | Yes  |                     |                               |
| Did the study avoid inappropriate exclusions?  | Yes  |                     |                               |
|  |  | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>  |  |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes  |                     |                               |
| If a threshold was used, was it pre-specified?   | Yes  |                     |                               |
|  |  | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>  |  |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear  |                     |                               |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes  |                     |                               |
|  |  | <b>Unclear</b>      | <b>Low</b>                    |

| DOMAIN 4: Flow and Timing  |     |            |  |
|--|-----|------------|--|
| Was there an appropriate interval between index test and reference standard? | Yes |            |  |
| Did all patients receive the same reference standard?                        | Yes |            |  |
| Were all patients included in the analysis?                                  | Yes |            |  |
|  |     | <b>Low</b> |  |

**Lacoma 2012**

| Study characteristics                      |  |
|--|--|
| Patient sampling                           | Cross-sectional design, all samples that received the reference standard were enrolled, prospective data collection  |
| Patient characteristics and setting        | <ol style="list-style-type: none"> <li>1. Country of origin: Spain.</li> <li>2. World Bank classification of country: high.</li> <li>3. Type of lab: hospital.</li> <li>4. Type of patients: confirmed MDR-TB cases.</li> <li>5. Smear-positive patients whose specimens were tested directly: 49/54.</li> </ol>                         |
| Index tests                                | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: no.</li> <li>2. Type of testing: direct and indirect.</li> <li>3. Type of specimens: smear-positive and smear negative.</li> <li>4. Specimen treatment: NALC-NaOH.</li> <li>5. Specimen condition: frozen.</li> <li>6. Duration of freezing: &gt; 1 year.</li> </ol> |
| Target condition and reference standard(s) | <ol style="list-style-type: none"> <li>1. Culture (liquid; BACTEC460TB) used for FQ, SLID.</li> <li>2. FQ drugs: moxifloxacin (0.5 µg/mL).</li> <li>3. SLIDs: kanamycin (5 µg/mL) and capreomycin (1.25 µg/mL).</li> <li>4. Discrepant analysis: yes (for indirect testing only).</li> </ol>   |
| Flow and timing                            | Uninterpretable results reported: yes  |
| Comparative                                |  |
| Notes                                      | Also performed discrepant analysis; in this analysis, not all patients received both culture-based DST and genetic sequencing reference standards  |
| Methodological quality                     |  |

| Item   | Authors' judgement | Risk of bias | Applicability concerns |
|--|--------------------|--------------|------------------------|
| <b>DOMAIN 1: Patient Selection</b>   |                    |              |                        |
| Was a consecutive or random sample of patients enrolled?   | Yes                |              |                        |
| Was a case-control design avoided?   | Yes                |              |                        |
| Did the study avoid inappropriate exclusions?  | Yes                |              |                        |
|  |                    | <b>Low</b>   | <b>Low</b>             |
| <b>DOMAIN 2: Index Test All tests</b>  |                    |              |                        |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes                |              |                        |
| If a threshold was used, was it pre-specified?   | Yes                |              |                        |
|  |                    | <b>Low</b>   | <b>Low</b>             |
| <b>DOMAIN 3: Reference Standard</b>  |                    |              |                        |
| Is the reference standards likely to correctly classify the target condition?                        | No                 |              |                        |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes                |              |                        |
|  |                    | <b>High</b>  | <b>Low</b>             |
| <b>DOMAIN 4: Flow and Timing</b>   |                    |              |                        |
| Was there an appropriate interval between index test and reference standard?                         | Yes                |              |                        |
| Did all patients receive the same reference standard?  | Yes                |              |                        |

**Lacoma 2012** (Continued)

|   |     |            |  |
|---|-----|------------|--|
| Were all patients included in the analysis? | Yes |            |  |
|   |     | <b>Low</b> |  |

**Lopez-Roa 2012**

|  |  |                     |                               |
|--|--|---------------------|-------------------------------|
| <b>Study characteristics</b>                             |  |                     |                               |
| Patient sampling   | Cross-sectional design with convenience-based enrolment of participants, prospective data collection   |                     |                               |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>Country of origin: Spain.</li> <li>World Bank classification of country: high.</li> <li>Type of lab: hospital.</li> <li>Type of patients: confirmed MDR-TB patients.</li> </ol>   |                     |                               |
| Index tests  | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>  |                     |                               |
| Target condition and reference standard(s)               | <ol style="list-style-type: none"> <li>Culture (solid, agar proportion): 7H11 (FQ) and 7H10 (SLID).</li> <li>FQ drugs: ofloxacin 2 µg/mL on 7H11.</li> <li>SLIDs: amikacin 4 µg/mL on 7H10.</li> <li>Discrepant analysis: yes.</li> </ol>  |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes  |                     |                               |
| Comparative  |  |                     |                               |
| Notes  | <p>For diagnostic accuracy, MGIT 960 was used, but MTBDR<sub>s/l</sub> data were presented (and thus extracted) using the agar proportion method in 7H10 or 7H11 as a reference standard.</p> <p>Other findings: the turnaround time for agar proportion, MGIT 960 and GenoType® MTBDR<sub>s/l</sub> were, respectively, 21 days, 8 days, and 8 hours. Also performed discrepant analysis; in this analysis, not all patients received both culture-based DST and genetic sequencing reference standards</p> |                     |                               |
| <b>Methodological quality</b>                            |  |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>                       |  |                     |                               |
| Was a consecutive or random sample of patients enrolled? | No   |                     |                               |
| Was a case-control design avoided?                       | Yes  |                     |                               |

**Lopez-Roa 2012** (Continued)

|  |         |                |            |
|--|---------|----------------|------------|
| Did the study avoid inappropriate exclusions?  | Yes     |                |            |
|  |         | <b>High</b>    | <b>Low</b> |
| <b>DOMAIN 2: Index Test All tests</b>  |         |                |            |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Unclear |                |            |
| If a threshold was used, was it pre-specified?   | Yes     |                |            |
|  |         | <b>Unclear</b> | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |         |                |            |
| Is the reference standards likely to correctly classify the target condition?                        | Yes     |                |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |                |            |
|  |         | <b>Low</b>     | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |            |
| Did all patients receive the same reference standard?  | Yes     |                |            |
| Were all patients included in the analysis?  | Yes     |                |            |
|  |         | <b>Low</b>     |            |

**Miotto 2012**

| <b>Study characteristics</b>                             |   |                     |                               |
|--|---|---------------------|-------------------------------|
| Patient sampling   | Isolates: case-control design with consecutive enrolment of participants, prospective data collection<br>Specimens: cross-sectional design with consecutive enrolment of participants, prospective data collection  |                     |                               |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>1. Country of origin: Unclear.</li> <li>2. World Bank classification of country: Unclear.</li> <li>3. Type of lab: hospital.</li> <li>4. Type of patients: confirmed MDR-TB cases, confirmed XDR-TB cases, confirmed MDR-TB patients with some known second-line resistance.</li> </ol>  |                     |                               |
| Index tests  | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: yes, donation of test.</li> <li>2. Type of testing: direct and indirect.</li> <li>3. Type of specimens: smear-positive.</li> <li>4. Specimen treatment: NALC-NaOH.</li> <li>5. Specimen condition: frozen.</li> <li>6. Duration of freezing: &gt; 1 year.</li> </ol>  |                     |                               |
| Target condition and reference standard(s)               | <ol style="list-style-type: none"> <li>1. Culture (liquid, MGIT 960) and sequencing used for FQ, SLID.</li> <li>2. FQ drugs: ofloxacin (2 µg/mL).</li> <li>3. SLIDs: amikacin (1.0 µg/mL), kanamycin (5.0 µg/mL), and capreomycin (2.5 µg/mL).</li> <li>4. Discrepant analysis: yes for XDR-TB, with sequencing.</li> <li>5. Genes for FQ: <i>gyrA</i>.</li> <li>6. Genes for SLIDs: <i>rrs</i>.</li> </ol> |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes   |                     |                               |
| Comparative  |   |                     |                               |
| Notes  | Other findings: negative predictive value for SLID is higher in Beijing strains   |                     |                               |
| <b>Methodological quality</b>                            |   |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>   | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>                       |   |                     |                               |
| Was a consecutive or random sample of patients enrolled? | Yes   |                     |                               |
| Was a case-control design avoided?                       | No  |                     |                               |
| Did the study avoid inappropriate exclusions?            | Yes   |                     |                               |
|  |   | <b>High</b>         | <b>High</b>                   |



**Miotto 2012** (Continued)

| <b>DOMAIN 2: Index Test All tests</b>  |         |                |            |
|--|---------|----------------|------------|
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes     |                |            |
| If a threshold was used, was it pre-specified?   | Yes     |                |            |
|  |         | <b>Low</b>     | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |         |                |            |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |                |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |                |            |
|  |         | <b>Unclear</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |            |
| Did all patients receive the same reference standard?  | Yes     |                |            |
| Were all patients included in the analysis?  | Yes     |                |            |
|  |         | <b>Low</b>     |            |

**Said 2012**

| <b>Study characteristics</b>        |   |
|-------------------------------------|---|
| Patient sampling                    | Cross-sectional design with consecutive-based enrolment of participants, prospective data collection  |
| Patient characteristics and setting | <ol style="list-style-type: none"> <li>1. Country of origin: South Africa.</li> <li>2. World Bank classification of country: middle.</li> <li>3. Type of lab: research.</li> <li>4. Type of patients: confirmed MDR-TB patients.</li> </ol> |

Said 2012 (Continued)

|  |   |
|--|---|
| Index tests                                | 1. Manufacturer involvement: yes, financial support.<br>2. Type of testing: indirect.   |
| Target condition and reference standard(s) | 1. Culture (solid; 7H11).<br>2. FQ drugs: ofloxacin (2 µg/mL).<br>3. SLIDs: kanamycin (5 µg/mL) and capreomycin (10 µg/mL). Not the WHO critical concentrations for SLIDs.<br>4. Discrepant analysis: no.   |
| Flow and timing                            | Uninterpretable results reported: yes   |
| Comparative                                |   |
| Notes                                      | Other findings: turnaround times for DST ranged from 6 to 21 days (median 11) for the agar proportion method and from 2 to 3 days (median 2) for the MTBDRs/ assay. DST results of the MTBDRs/ assay as compared to the agar proportion method are shown in Table 2 |

#### Methodological quality

| Item  | Authors' judgement | Risk of bias | Applicability concerns |
|---|--------------------|--------------|------------------------|
| <b>DOMAIN 1: Patient Selection</b>  |                    |              |                        |
| Was a consecutive or random sample of patients enrolled?  | Yes                |              |                        |
| Was a case-control design avoided?  | Yes                |              |                        |
| Did the study avoid inappropriate exclusions?   | Yes                |              |                        |
|   |                    | <b>Low</b>   | <b>Low</b>             |
| <b>DOMAIN 2: Index Test All tests</b>   |                    |              |                        |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Yes                |              |                        |
| If a threshold was used, was it pre-specified?  | Yes                |              |                        |
|   |                    | <b>Low</b>   | <b>Low</b>             |
| <b>DOMAIN 3: Reference Standard</b>   |                    |              |                        |

Said 2012 (Continued)

|  |         |         |     |
|--|---------|---------|-----|
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |         |     |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |         |     |
|  |         | Unclear | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |         |         |     |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |         |     |
| Did all patients receive the same reference standard?  | Yes     |         |     |
| Were all patients included in the analysis?  | Yes     |         |     |
|  |         | Low     |     |

Simons 2015

|  |   |
|--|---|
| <b>Study characteristics</b>               |   |
| Patient sampling                           | Retrospective, cross-sectional study with consecutive sampling  |
| Patient characteristics and setting        | <ol style="list-style-type: none"> <li>Country of origin: the Netherlands.</li> <li>World Bank classification of country: high.</li> <li>Type of lab: research.</li> <li>Type of patients: confirmed MDR-TB patients.</li> </ol>  |
| Index tests                                | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>   |
| Target condition and reference standard(s) | <ol style="list-style-type: none"> <li>Culture (liquid; MGIT 960 and 7H10).</li> <li>FQ drugs: moxifloxacin 0.5 µg/mL for MGIT and 1 µg/mL for 7H10.</li> <li>SLIDs: amikacin 1.0 µg/mL for MGIT and 5 µg/mL for 7H10; and capreomycin 2.5 µg/mL for MGIT and 10 µg/mL for 7H10.</li> </ol> |
| Flow and timing                            | Uninterpretable results reported: yes   |
| Comparative                                |   |

|  |                           |                     |                               |
|--|---------------------------|---------------------|-------------------------------|
| Notes  |                           |                     |                               |
| <b>Methodological quality</b>  |                           |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b> | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>   |                           |                     |                               |
| Was a consecutive or random sample of patients enrolled?   | Yes                       |                     |                               |
| Was a case-control design avoided?   | Yes                       |                     |                               |
| Did the study avoid inappropriate exclusions?  | Yes                       |                     |                               |
|  |                           | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>  |                           |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Unclear                   |                     |                               |
| If a threshold was used, was it pre-specified?   | Yes                       |                     |                               |
|  |                           | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>  |                           |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear                   |                     |                               |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes                       |                     |                               |
|  |                           | <b>Unclear</b>      | <b>Low</b>                    |
| <b>DOMAIN 4: Flow and Timing</b>   |                           |                     |                               |
| Was there an appropriate interval between index test and reference standard?                         | Yes                       |                     |                               |

**Simons 2015** (Continued)

|   |     |            |  |
|---|-----|------------|--|
| Did all patients receive the same reference standard? | Yes |            |  |
| Were all patients included in the analysis?           | Yes |            |  |
|   |     | <b>Low</b> |  |

**Surcouf 2011**

| <b>Study characteristics</b>                             |   |                     |                               |
|--|---|---------------------|-------------------------------|
| Patient sampling   | Cross-sectional design with consecutive-based enrolment of participants, prospective data collection  |                     |                               |
| Patient characteristics and setting                      | <ol style="list-style-type: none"> <li>Country of origin: Cambodia.</li> <li>World Bank classification of country: low.</li> <li>Type of lab: unknown.</li> <li>Type of patients: confirmed MDR-TB cases.</li> </ol>  |                     |                               |
| Index tests  | <ol style="list-style-type: none"> <li>Manufacturer involvement: yes, donation of tests.</li> <li>Type of testing: indirect.</li> </ol>   |                     |                               |
| Target condition and reference standard(s)               | <ol style="list-style-type: none"> <li>Sequencing used for reference standard.</li> <li>FQ genes: <i>gyrA</i>.</li> <li>SLID genes: <i>rrs</i>.</li> <li>Discrepant analysis: no.</li> </ol>  |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes   |                     |                               |
| Comparative  |   |                     |                               |
| Notes  | Other findings: spoligotyping results showed that most MDR strains belonged to the Beijing family (57/101, 56%) or were Beijing like (2/101, 2%). This percentage is higher in MDR FQ-R strains (10/14, 71%). This confirms that Beijing strains are more prone to accumulate antibiotic resistance |                     |                               |
| <b>Methodological quality</b>                            |   |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>   | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>                       |   |                     |                               |
| Was a consecutive or random sample of patients enrolled? | Yes   |                     |                               |
| Was a case-control design avoided?                       | Yes   |                     |                               |

Surcouf 2011 (Continued)

|  |         |         |     |
|--|---------|---------|-----|
| Did the study avoid inappropriate exclusions?  | Yes     |         |     |
|  |         | Low     | Low |
| <b>DOMAIN 2: Index Test All tests</b>  |         |         |     |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Unclear |         |     |
| If a threshold was used, was it pre-specified?   | Yes     |         |     |
|  |         | Unclear | Low |
| <b>DOMAIN 3: Reference Standard</b>  |         |         |     |
| Is the reference standards likely to correctly classify the target condition?                        | Yes     |         |     |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |         |     |
|  |         | Unclear | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |         |         |     |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |         |     |
| Did all patients receive the same reference standard?  | Yes     |         |     |
| Were all patients included in the analysis?  | Yes     |         |     |
|  |         | Low     |     |

| Study characteristics   |   |              |                        |
|---|---|--------------|------------------------|
| Patient sampling  | Cross-sectional design with consecutive-based enrolment of participants, prospective data collection  |              |                        |
| Patient characteristics and setting   | <ol style="list-style-type: none"> <li>1. Country of origin: Italy, Sweden, Germany, and Moldova.</li> <li>2. World Bank classification of country: high and middle.</li> <li>3. Type of lab: supranational reference laboratory.</li> <li>4. Type of patients: confirmed MDR-TB patients.</li> </ol>   |              |                        |
| Index tests   | <ol style="list-style-type: none"> <li>1. MTBDRs/ version 2.0</li> <li>2. Manufacturer involvement: yes, donation of tests.</li> <li>3. Type of testing: direct and indirect.</li> </ol>  |              |                        |
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>1. Culture (liquid MGIT 960, solid: LJ proportion).</li> <li>2. FQ drugs: MGIT: ofloxacin 2 µg/mL, levofloxacin 1.5 µg/mL, moxifloxacin 0.5 µg/mL; LJ: ofloxacin 4 µg/mL.</li> <li>3. SLIDs: amikacin 1 µg/mL, kanamycin 2.5 µg/mL, and capreomycin 2.5 µg/mL; LJ: amikacin 30 µg/mL, kanamycin 30 µg/mL, and capreomycin 40 µg/mL.</li> <li>4. Discrepant sequencing on false positive samples only.</li> </ol> |              |                        |
| Flow and timing   | Uninterpretable results reported: yes   |              |                        |
| Comparative   |   |              |                        |
| Notes   |   |              |                        |
| Methodological quality  |   |              |                        |
| Item  | Authors' judgement  | Risk of bias | Applicability concerns |
| <b>DOMAIN 1: Patient Selection</b>  |   |              |                        |
| Was a consecutive or random sample of patients enrolled?  | Yes   |              |                        |
| Was a case-control design avoided?  | Yes   |              |                        |
| Did the study avoid inappropriate exclusions?   | Yes   |              |                        |
|   |   | <b>Low</b>   | <b>Low</b>             |
| <b>DOMAIN 2: Index Test All tests</b>   |   |              |                        |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Yes   |              |                        |

**Tagliani 2015** (Continued)

|  |         |                |            |
|--|---------|----------------|------------|
| If a threshold was used, was it pre-specified?   | Yes     |                |            |
|  |         | <b>Low</b>     | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |         |                |            |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |                |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes     |                |            |
|  |         | <b>Unclear</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |            |
| Did all patients receive the same reference standard?  | No      |                |            |
| Were all patients included in the analysis?  | Yes     |                |            |
|  |         | <b>Low</b>     |            |

**Tomasicchio 2016**

|                                     |   |
|-------------------------------------|---|
| <b>Study characteristics</b>        |   |
| Patient sampling                    | Isolates: case-control design with consecutive enrolment of participants<br>Specimens: cross-sectional design with consecutive enrolment of participants  |
| Patient characteristics and setting | <ol style="list-style-type: none"> <li>Country of origin: South Africa.</li> <li>World Bank classification of country: middle.</li> <li>Type of lab: reference laboratory.</li> <li>Type of patients: confirmed MDR-TB patients and patients with confirmed second-line drug resistance.</li> </ol> |
| Index tests                         | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: direct and indirect.</li> </ol>  |



|  |   |                     |                               |
|--|---|---------------------|-------------------------------|
| Target condition and reference standard(s)   | 1. Culture (liquid MGIT 960).<br>2. FQ drugs: ofloxacin 2 µg/mL.<br>3. SLIDs: amikacin 1 µg/mL. |                     |                               |
| Flow and timing  | Uninterpretable results reported: yes   |                     |                               |
| Comparative  |   |                     |                               |
| Notes  | A case-control study design was used to evaluate culture isolates                               |                     |                               |
| <b>Methodological quality</b>  |   |                     |                               |
| <b>Item</b>  | <b>Authors' judgement</b>   | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>   |   |                     |                               |
| Was a consecutive or random sample of patients enrolled?   | Yes   |                     |                               |
| Was a case-control design avoided?   | No  |                     |                               |
| Did the study avoid inappropriate exclusions?  | Yes   |                     |                               |
|  |   | <b>High</b>         | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>  |   |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Yes   |                     |                               |
| If a threshold was used, was it pre-specified?   | Yes   |                     |                               |
|  |   | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>  |   |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                        | Yes   |                     |                               |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes   |                     |                               |

|  |     |     |     |
|--|-----|-----|-----|
|  |     | Low | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |     |     |     |
| Was there an appropriate interval between index test and reference standard? | Yes |     |     |
| Did all patients receive the same reference standard?                        | Yes |     |     |
| Were all patients included in the analysis?                                  | Yes |     |     |
|  |     | Low |     |

**Tukvadze 2014**

|  |  |                     |                               |
|--|--|---------------------|-------------------------------|
| <b>Study characteristics</b>               |  |                     |                               |
| Patient sampling                           | Cross-sectional design with consecutive enrolment of participants, prospective data collection   |                     |                               |
| Patient characteristics and setting        | <ol style="list-style-type: none"> <li>Country of origin: Georgia.</li> <li>World Bank classification of country: middle.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed MDR-TB patients.</li> </ol>                            |                     |                               |
| Index tests                                | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: direct testing.</li> </ol>  |                     |                               |
| Target condition and reference standard(s) | <ol style="list-style-type: none"> <li>Culture based DST, LJ.</li> <li>FQ: ofloxacin 2 µg/mL.</li> <li>SLIDS: capreomycin 40 µg/mL; kanamycin 30 µg/mL.</li> <li>There was no discrepant analysis.</li> <li>All reported XDR-TB resistance.</li> </ol> |                     |                               |
| Flow and timing                            | Uninterpretable results reported: yes  |                     |                               |
| Comparative                                |  |                     |                               |
| Notes                                      |  |                     |                               |
| <b>Methodological quality</b>              |  |                     |                               |
| <b>Item</b>                                | <b>Authors' judgement</b>  | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>         |  |                     |                               |

|  |         |                |                |
|--|---------|----------------|----------------|
| Was a consecutive or random sample of patients enrolled?   | Yes     |                |                |
| Was a case-control design avoided?   | Yes     |                |                |
| Did the study avoid inappropriate exclusions?  | Yes     |                |                |
|  |         | <b>Low</b>     | <b>Low</b>     |
| <b>DOMAIN 2: Index Test All tests</b>  |         |                |                |
| Were the index test results interpreted without knowledge of the results of the reference standard?  | Unclear |                |                |
| If a threshold was used, was it pre-specified?   | Yes     |                |                |
|  |         | <b>Unclear</b> | <b>Unclear</b> |
| <b>DOMAIN 3: Reference Standard</b>  |         |                |                |
| Is the reference standards likely to correctly classify the target condition?                        | Unclear |                |                |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |                |                |
|  |         | <b>Unclear</b> | <b>Low</b>     |
| <b>DOMAIN 4: Flow and Timing</b>   |         |                |                |
| Was there an appropriate interval between index test and reference standard?                         | Yes     |                |                |
| Did all patients receive the same reference standard?  | Yes     |                |                |
| Were all patients included in the analysis?  | Yes     |                |                |
|  |         | <b>Low</b>     |                |

| Study characteristics   |  |              |                        |
|---|--|--------------|------------------------|
| Patient sampling  | Cross-sectional design with enrolment of participants by convenience, retrospective data collection  |              |                        |
| Patient characteristics and setting   | <ol style="list-style-type: none"> <li>1. Country of origin: Netherlands</li> <li>2. World Bank classification of country: high</li> <li>3. Type of lab: reference</li> <li>4. Type of patients: confirmed MDR-TB patients, confirmed MDR-TB patients with some known second-line resistance.</li> </ol> |              |                        |
| Index tests   | <ol style="list-style-type: none"> <li>1. Manufacturer involvement: no.</li> <li>2. Type of testing: indirect.</li> </ol>  |              |                        |
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>1. Culture (solid; 7H10).</li> <li>2. FQ drugs: moxifloxacin (1 µg/mL).</li> <li>3. SLIDs: amikacin (5 µg/mL) and capreomycin (10 µg/mL). WHO critical concentrations not used for 7H10 solid culture.</li> <li>4. Discrepant analysis: no.</li> </ol>            |              |                        |
| Flow and timing   | Uninterpretable results reported: yes  |              |                        |
| Comparative   |  |              |                        |
| Notes   | Relevant clinical information? unclear   |              |                        |
| Methodological quality  |  |              |                        |
| Item  | Authors' judgement   | Risk of bias | Applicability concerns |
| DOMAIN 1: Patient Selection   |  |              |                        |
| Was a consecutive or random sample of patients enrolled?  | No   |              |                        |
| Was a case-control design avoided?  | Yes  |              |                        |
| Did the study avoid inappropriate exclusions?   | Yes  |              |                        |
|   |  | <b>High</b>  | <b>High</b>            |
| DOMAIN 2: Index Test All tests  |  |              |                        |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Yes  |              |                        |

|  |     |             |            |
|--|-----|-------------|------------|
| If a threshold was used, was it pre-specified?   | Yes |             |            |
|  |     | <b>Low</b>  | <b>Low</b> |
| <b>DOMAIN 3: Reference Standard</b>  |     |             |            |
| Is the reference standards likely to correctly classify the target condition?                        | No  |             |            |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes |             |            |
|  |     | <b>High</b> | <b>Low</b> |
| <b>DOMAIN 4: Flow and Timing</b>   |     |             |            |
| Was there an appropriate interval between index test and reference standard?                         | Yes |             |            |
| Did all patients receive the same reference standard?  | Yes |             |            |
| Were all patients included in the analysis?  | Yes |             |            |
|  |     | <b>Low</b>  |            |

**Zivanovic 2012**

|                                     |  |
|-------------------------------------|--|
| <b>Study characteristics</b>        |  |
| Patient sampling                    | Cross-sectional design with consecutive enrolment of participants, prospective data collection   |
| Patient characteristics and setting | <ol style="list-style-type: none"> <li>Country of origin: Serbia.</li> <li>World Bank classification of country: middle.</li> <li>Type of lab: reference.</li> <li>Type of patients: confirmed MDR-TB patients.</li> </ol> |
| Index tests                         | <ol style="list-style-type: none"> <li>Manufacturer involvement: no.</li> <li>Type of testing: indirect.</li> </ol>  |

|   |   |                     |                               |
|---|---|---------------------|-------------------------------|
| Target condition and reference standard(s)  | <ol style="list-style-type: none"> <li>1. Culture (solid and liquid; LJ and MGIT 960).</li> <li>2. FQ drugs: ofloxacin (2 µg/mL).</li> <li>3. SLIDs: amikacin (1 µg/mL for MGIT and 40 µg/mL for LJ; capreomycin 2.5 µg/mL for MGIT and 40.0 for LJ).</li> <li>4. Discrepant analysis: no.</li> </ol> |                     |                               |
| Flow and timing   | Uninterpretable results reported: yes   |                     |                               |
| Comparative   |   |                     |                               |
| Notes   |   |                     |                               |
| <b>Methodological quality</b>   |   |                     |                               |
| <b>Item</b>   | <b>Authors' judgement</b>   | <b>Risk of bias</b> | <b>Applicability concerns</b> |
| <b>DOMAIN 1: Patient Selection</b>  |   |                     |                               |
| Was a consecutive or random sample of patients enrolled?  | Yes   |                     |                               |
| Was a case-control design avoided?  | Yes   |                     |                               |
| Did the study avoid inappropriate exclusions?   | Yes   |                     |                               |
|   |   | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 2: Index Test All tests</b>   |   |                     |                               |
| Were the index test results interpreted without knowledge of the results of the reference standard? | Yes   |                     |                               |
| If a threshold was used, was it pre-specified?  | Yes   |                     |                               |
|   |   | <b>Low</b>          | <b>Low</b>                    |
| <b>DOMAIN 3: Reference Standard</b>   |   |                     |                               |
| Is the reference standards likely to correctly classify the target condition?                       | Unclear   |                     |                               |

|  |     |         |     |
|--|-----|---------|-----|
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Yes |         |     |
|  |     | Unclear | Low |
| <b>DOMAIN 4: Flow and Timing</b>   |     |         |     |
| Was there an appropriate interval between index test and reference standard?                         | Yes |         |     |
| Did all patients receive the same reference standard?  | Yes |         |     |
| Were all patients included in the analysis?  | Yes |         |     |
|  |     | Low     |     |

Abbreviations: AP: agar proportion; DST: drug susceptibility testing; FQ: fluoroquinolone; LJ: Löwenstein-Jensen; MGIT: Mycobacteria Growth Indicator Tube; SLID: second-line injectable drug; TB: tuberculosis; XDR-TB: extensively drug-resistant TB.

### Characteristics of excluded studies [ordered by study ID]

| Study                             | Reason for exclusion  |
|-----------------------------------|---|
| <a href="#">Aubry 2014</a>        | No cases with second-line resistance.   |
| <a href="#">Babamahmoodi 2014</a> | Not a diagnostic accuracy study.  |
| <a href="#">Bantouna 2011</a>     | Conference abstract.  |
| <a href="#">Bergvala 2010</a>     | Technical article. Not a diagnostic accuracy study.   |
| <a href="#">Brossier 2010b</a>    | Conference abstract.  |
| <a href="#">Choi 2010</a>         | Technical article. Not a diagnostic accuracy study.   |
| <a href="#">Fallico 2012</a>      | Conference abstract.  |
| <a href="#">Felkel 2013</a>       | Technical article. No diagnostic data for fluoroquinolones (FQs), second-line injectable drugs (SLIDs), or extensively drug-resistant tuberculosis (XDR-TB) |

(Continued)

|                                  |  |
|----------------------------------|--|
| <a href="#">Festoso 2011</a>     | Conference abstract.   |
| <a href="#">Gkaravela 2012</a>   | Conference abstract.   |
| <a href="#">Gomgnimbou 2015</a>  | Not a diagnostic accuracy study.   |
| <a href="#">Iem 2013</a>         | Technical article. Only one case of second-line resistance.  |
| <a href="#">Jang 2011</a>        | Conference abstract.   |
| <a href="#">Karabela 2007</a>    | Conference abstract.   |
| <a href="#">Kaswa 2014</a>       | Not a diagnostic accuracy study.   |
| <a href="#">Kontos 2011</a>      | Conference abstract.   |
| <a href="#">Kontos 2012</a>      | Conference abstract.   |
| <a href="#">Lacoma 2015</a>      | Not a diagnostic accuracy study.   |
| <a href="#">Lemus 2011</a>       | Conference abstract.   |
| <a href="#">López-Roa 2010</a>   | Conference abstract.   |
| <a href="#">Mindru 2013</a>      | Data insufficient for 2 x 2 tables.  |
| <a href="#">Molina-Moya 2015</a> | Test other than MTBDR <sub>s/l</sub> .   |
| <a href="#">Niehaus 2015</a>     | Not a diagnostic accuracy study.   |
| <a href="#">Orikiriza 2015</a>   | Data insufficient for 2 x 2 tables.  |
| <a href="#">Ouassa 2014</a>      | Not a diagnostic accuracy study.   |
| <a href="#">Singh 2013</a>       | Technical article. No information on resistance to the pre-specified FQs and no cases susceptible to the SLIDs |
| <a href="#">Tessema 2012a</a>    | Technical article. No information on resistance to FQs, SLIDs, or XDR-TB                                       |
| <a href="#">Tessema 2012b</a>    | Technical article. No information on resistance to FQs, SLIDs, or XDR-TB                                       |
| <a href="#">Totten 2011</a>      | Conference abstract.   |
| <a href="#">Walker 2015</a>      | Not a diagnostic accuracy study.   |
| <a href="#">Wedajo 2014</a>      | Not a diagnostic accuracy study.   |
| <a href="#">Zhang 2011</a>       | Conference abstract.   |



Abbreviations: FQ: fluoroquinolones; SLIDs: second-line injectable drugs; tuberculosis: TB; XDR-TB: extensively drug-resistant tuberculosis.

## DATA

Presented below are all the data for all of the tests entered into the review.

### Tests. Data tables by test

| Test  | No. of studies | No. of participants |
|---|----------------|---------------------|
| 1 Indirect, FQ, culture   | 19             | 2223                |
| 2 Indirect, ofloxacin, culture                                  | 13             | 1927                |
| 3 Indirect, moxifloxacin, culture                               | 6              | 419                 |
| 4 Indirect, levofloxacin, culture                               | 2              | 169                 |
| 5 Indirect, ofloxacin, WHO critical concentration used          | 8              | 1427                |
| 6 Indirect, ofloxacin, WHO critical concentration not used      | 4              | 481                 |
| 7 Indirect, SLID, culture                                       | 16             | 1921                |
| 8 Indirect, amikacin, culture                                   | 11             | 1301                |
| 9 Indirect, kanamycin, culture                                  | 9              | 1342                |
| 10 Indirect, capreomycin, culture                               | 10             | 1406                |
| 11 Indirect, amikacin, WHO critical concentration used          | 4              | 706                 |
| 12 Indirect, capreomycin, WHO critical concentration used       | 4              | 473                 |
| 13 Indirect, amikacin, WHO critical concentration not used      | 7              | 595                 |
| 14 Indirect, capreomycin, WHO critical concentration not used   | 6              | 933                 |
| 15 Indirect, XDR, culture                                       | 8              | 880                 |
| 16 Indirect, FQ, sequencing                                     | 7              | 974                 |
| 17 Indirect, SLID, sequencing                                   | 7              | 962                 |
| 18 Indirect, XDR, sequencing                                    | 4              | 630                 |
| 19 Indirect, FQ, sequencing and culture                         | 7              | 1211                |
| 20 Indirect, SLID, sequencing and culture                       | 7              | 1491                |
| 21 Indirect, XDR, sequencing and culture                        | 2              | 435                 |
| 22 Indirect, FQ, culture followed by sequencing of discrepant   | 3              | 427                 |
| 23 Indirect, SLID, culture followed by sequencing of discrepant | 3              | 619                 |
| 24 Direct, FQ, culture  | 9              | 1771                |
| 25 Direct, ofloxacin, culture                                   | 7              | 1667                |
| 26 Direct, moxifloxacin, culture                                | 2              | 821                 |
| 27 Ofloxacin, smear positive                                    | 4              | 963                 |
| 28 Ofloxacin, smear negative                                    | 2              | 120                 |
| 29 Ofloxacin, smear grade = scanty                              | 2              | 65                  |

|   |   |      |
|---|---|------|
| 30 Ofloxacin, smear grade = 1+                                | 4 | 241  |
| 31 Ofloxacin, smear grade $\geq$ 2+                           | 4 | 647  |
| 32 Moxifloxacin, smear positive                               | 2 | 821  |
| 33 Moxifloxacin, smear negative                               | 1 | 91   |
| 34 Moxifloxacin, smear grade = scanty                         | 1 | 51   |
| 35 Moxifloxacin, smear grade = 1+                             | 2 | 197  |
| 36 Moxifloxacin, smear grade $\geq$ 2+                        | 2 | 593  |
| 37 Direct, SLID, culture                                      | 8 | 1639 |
| 38 Direct, amikacin, culture                                  | 6 | 1491 |
| 39 Direct, kanamycin, culture                                 | 5 | 1020 |
| 40 Direct, capreomycin, culture                               | 5 | 1027 |
| 41 Amikacin, smear positive                                   | 4 | 809  |
| 42 Amikacin, smear negative                                   | 2 | 104  |
| 43 Amikacin, smear grade = scanty                             | 3 | 57   |
| 44 Amikacin, smear grade= 1+                                  | 4 | 222  |
| 45 Amikacin, smear grade $\geq$ 2+                            | 4 | 602  |
| 46 Kanamycin, smear positive                                  | 3 | 806  |
| 47 Kanamycin, smear negative                                  | 1 | 73   |
| 48 Kanamycin, smear grade = scanty                            | 2 | 43   |
| 49 Kanamycin, smear grade = +1                                | 3 | 193  |
| 50 Kanamycin, smear grade $\geq$ 2+                           | 3 | 564  |
| 51 Capreomycin, smear positive                                | 3 | 806  |
| 52 Capreomycin, smear negative                                | 1 | 73   |
| 53 Capreomycin, smear grade = 1+                              | 3 | 193  |
| 54 Capreomycin, smear grade = scanty                          | 2 | 43   |
| 55 Capreomycin, smear grade $\geq$ 2+                         | 3 | 564  |
| 56 Direct, XDR, culture                                       | 6 | 1420 |
| 57 Direct, FQ, culture followed by sequencing of discrepant   | 2 | 685  |
| 58 Direct, SLID, culture followed by sequencing of discrepant | 2 | 666  |
| 59 Direct, XDR, culture followed by sequencing of discrepant  | 2 | 570  |
| 60 V2, Indirect, FQ, culture                                  | 1 | 228  |
| 61 V2, Direct, FQ, smear positive                             | 1 | 155  |
| 62 V2, Direct, FQ, smear negative                             | 1 | 9    |
| 63 V2, Indirect, ofloxacin, culture                           | 1 | 226  |
| 64 V2, Indirect, moxifloxacin, culture                        | 1 | 97   |
| 65 V2, Indirect, SLID, culture                                | 1 | 228  |
| 66 V2, Direct, SLID, smear positive                           | 1 | 164  |
| 67 V2, Direct, SLID, smear negative                           | 1 | 9    |
| 68 V2, Indirect, amikacin, culture                            | 1 | 226  |

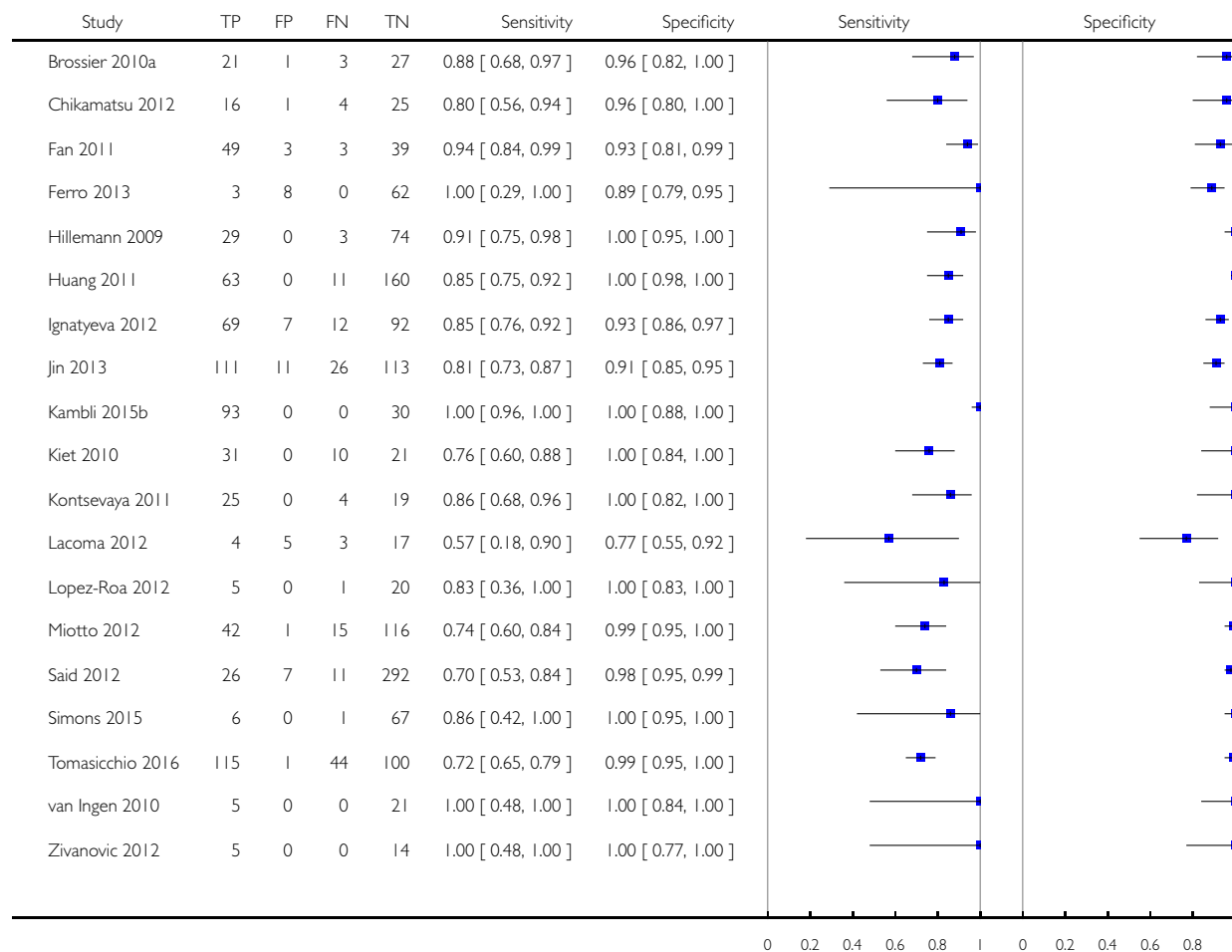
|  |   |     |
|--|---|-----|
| 69 V2, Indirect, kanamycin, culture        | 1 | 224 |
| 70 V2, Indirect, capreomycin, culture      | 1 | 218 |
| 71 V2, Ofloxacin, smear positive           | 1 | 153 |
| 72 V2, Ofloxacin, smear negative           | 1 | 9   |
| 73 V2, Ofloxacin, smear grade = scanty     | 1 | 38  |
| 74 V2, Ofloxacin, smear grade = 1+         | 1 | 56  |
| 75 V2, Ofloxacin, smear grade $\geq$ 2+    | 1 | 49  |
| 76 V2, Moxifloxacin, smear positive        | 1 | 22  |
| 77 V2, Moxifloxacin, smear grade $\geq$ 2+ | 1 | 8   |
| 78 V2, Levofloxacin, smear positive        | 1 | 53  |
| 79 V2, Levofloxacin, smear grade = scanty  | 1 | 25  |
| 80 V2, Levofloxacin, smear grade = 1+      | 1 | 22  |
| 81 V2, Levofloxacin, smear grade $\geq$ 2+ | 1 | 6   |
| 82 V2, Amikacin, smear positive            | 1 | 155 |
| 83 V2, Amikacin, smear negative            | 1 | 9   |
| 84 V2, Amikacin, smear grade = scanty      | 1 | 40  |
| 85 V2, Amikacin, smear grade = 1+          | 1 | 57  |
| 86 V2, Amikacin, smear grade $\geq$ 2+     | 1 | 49  |
| 87 V2, Kanamycin, smear positive           | 1 | 155 |
| 88 V2, Kanamycin, smear negative           | 1 | 7   |
| 89 V2, Kanamycin, smear grade = scanty     | 1 | 37  |
| 90 V2, Kanamycin, smear grade = 1+         | 1 | 55  |
| 91 V2, Kanamycin, smear grade $\geq$ 2+    | 1 | 45  |
| 92 V2, Capreomycin, smear positive         | 1 | 164 |
| 93 V2, Capreomycin, smear negative         | 1 | 9   |
| 94 V2, Capreomycin, smear grade = scanty   | 1 | 40  |
| 95 V2, Capreomycin, smear grade = 1+       | 1 | 57  |
| 96 V2, Capreomycin, smear grade $\geq$ 2+  | 1 | 49  |
| 97 V2, Indirect, XDR, culture              | 1 | 228 |

|                                    |   |     |
|------------------------------------|---|-----|
| 98 V2, Direct, XDR, smear positive | 1 | 164 |
| 99 V2, Direct, XDR, smear negative | 1 | 9   |

**Test 1. Indirect, FQ, culture.**

Review: GenoType MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs

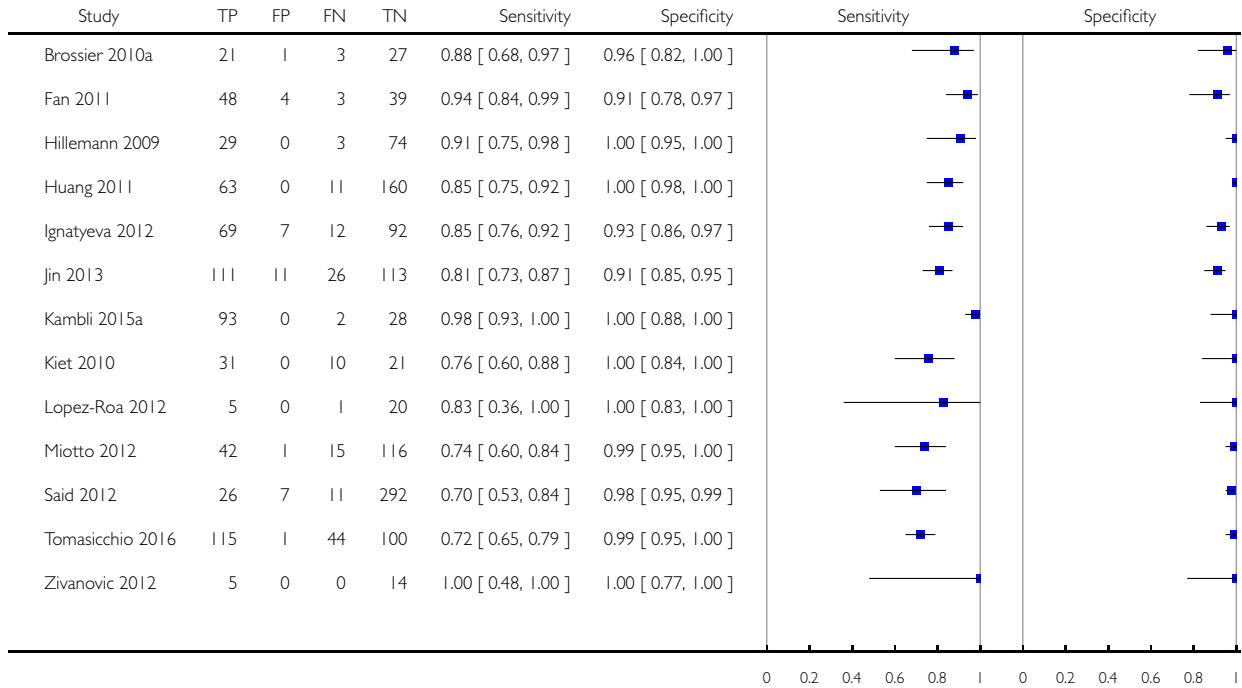
Test: 1 Indirect, FQ, culture



## Test 2. Indirect, ofloxacin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

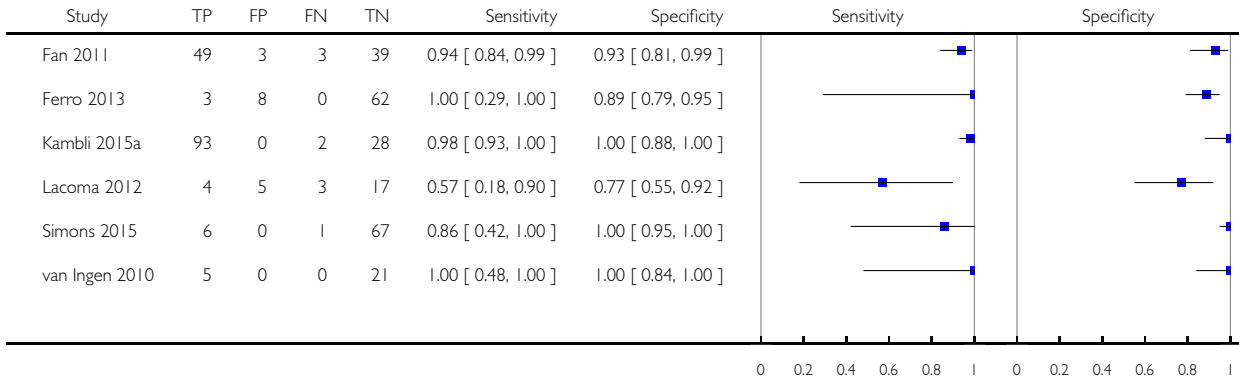
Test: 2 Indirect, ofloxacin, culture



### Test 3. Indirect, moxifloxacin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

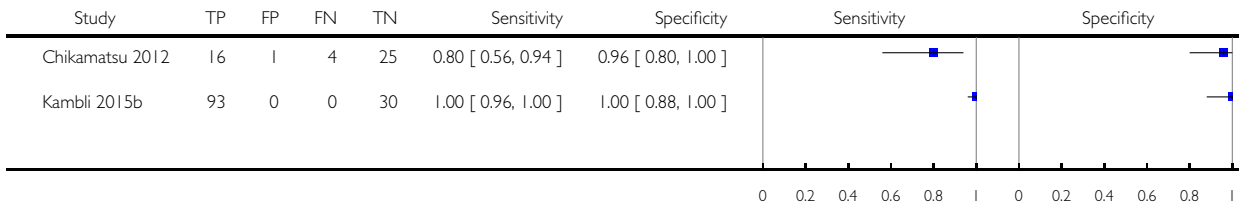
Test: 3 Indirect, moxifloxacin, culture



### Test 4. Indirect, levofloxacin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

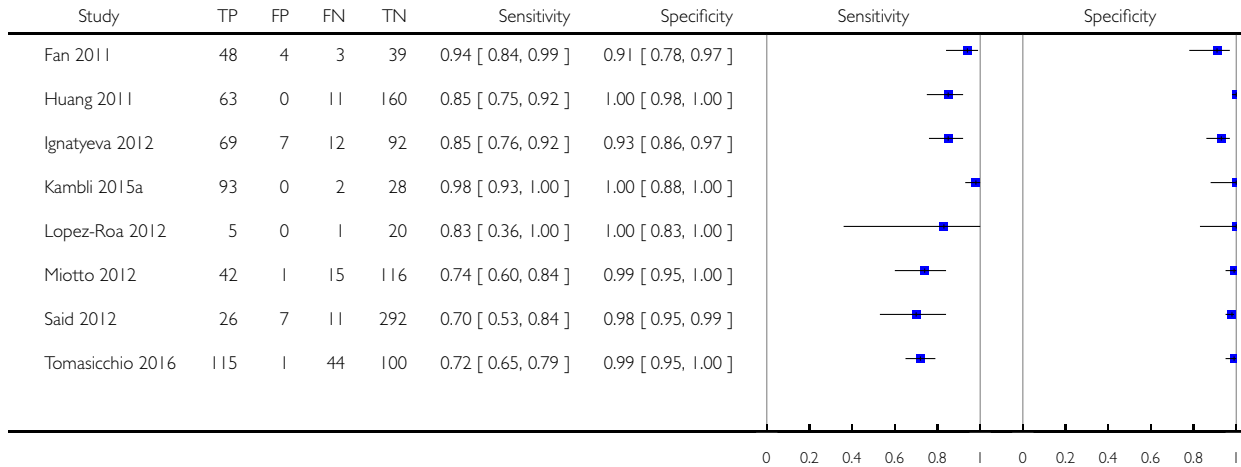
Test: 4 Indirect, levofloxacin, culture



### Test 5. Indirect, ofloxacin, WHO critical concentration used.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

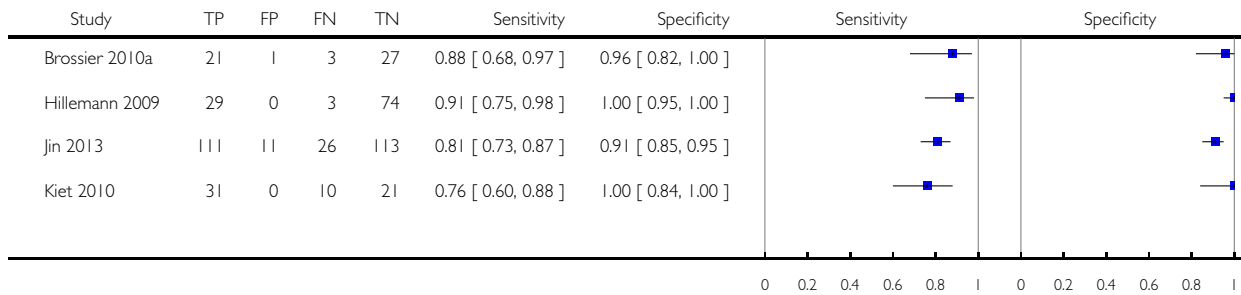
Test: 5 Indirect, ofloxacin, WHO critical concentration used



### Test 6. Indirect, ofloxacin, WHO critical concentration not used.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

Test: 6 Indirect, ofloxacin, WHO critical concentration not used

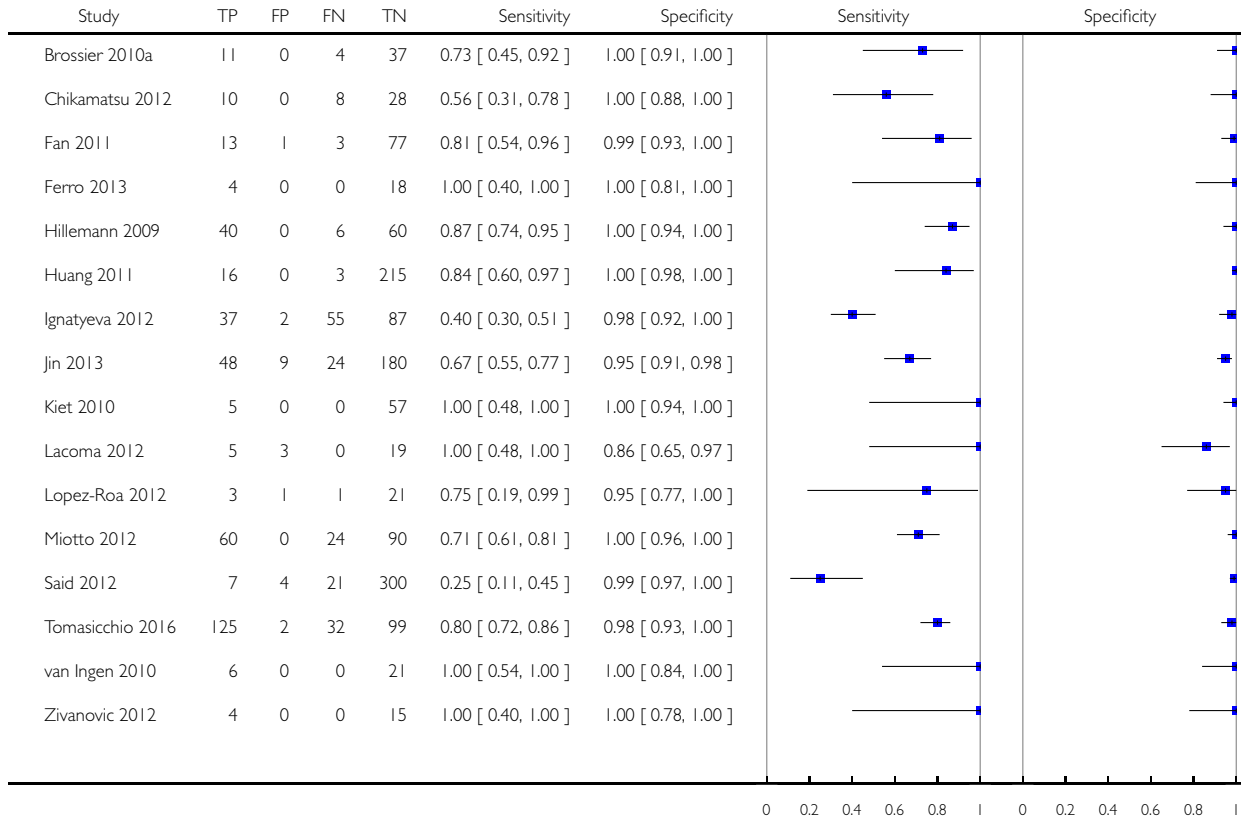




### Test 7. Indirect, SLID, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

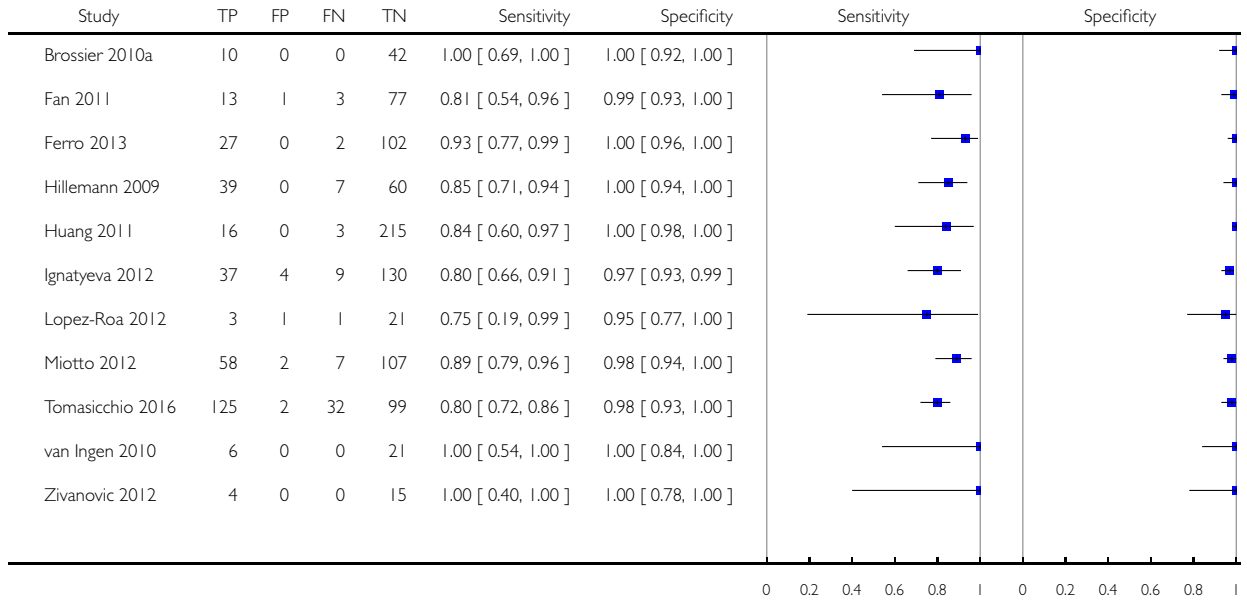
Test: 7 Indirect, SLID, culture



### Test 8. Indirect, amikacin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

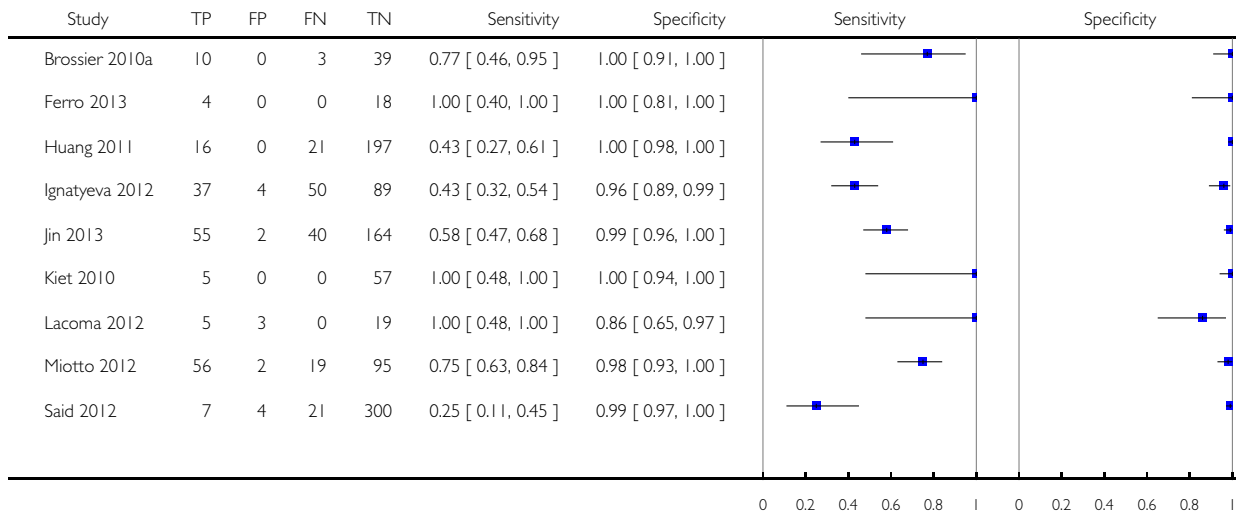
Test: 8 Indirect, amikacin, culture



### Test 9. Indirect, kanamycin, culture.

Review: GenoType MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs

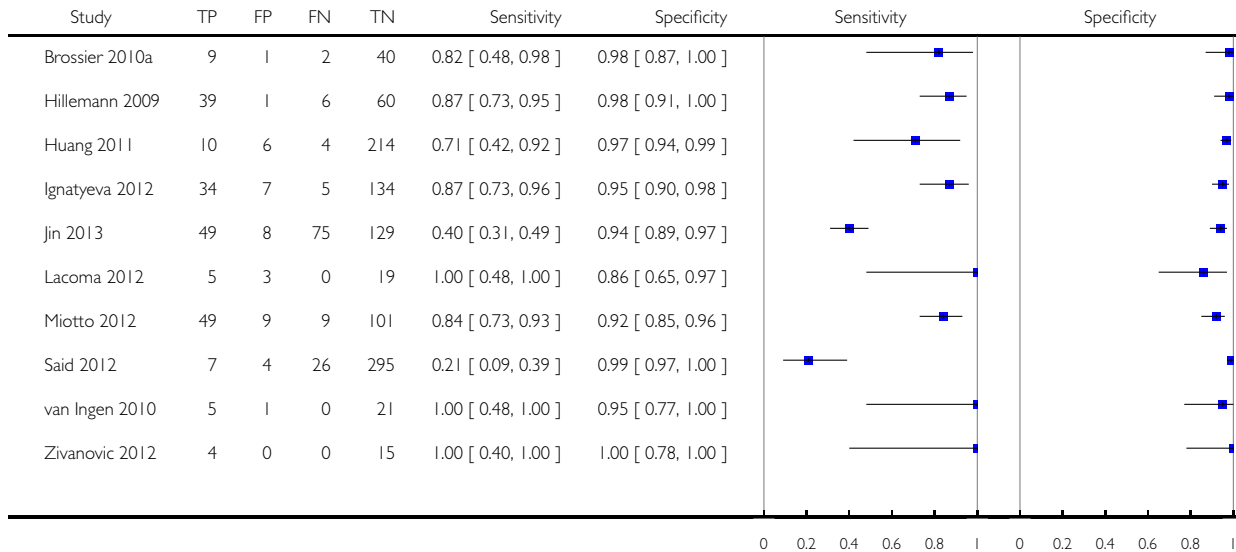
Test: 9 Indirect, kanamycin, culture



### Test 10. Indirect, capreomycin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

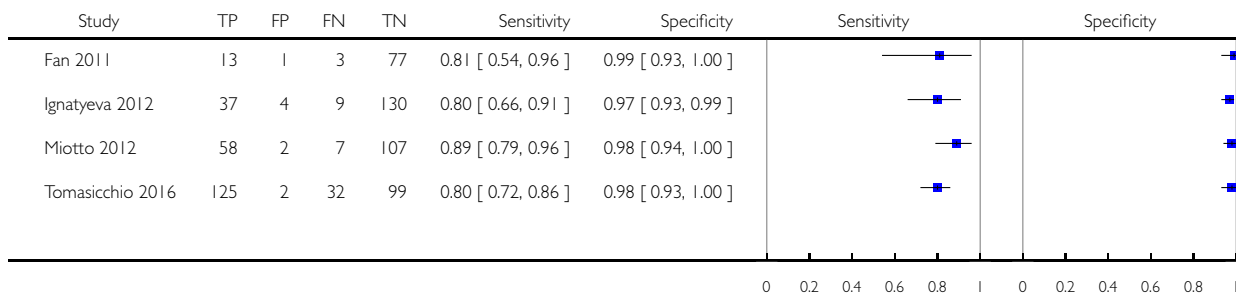
Test: 10 Indirect, capreomycin, culture



### Test 11. Indirect, amikacin, WHO critical concentration used.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

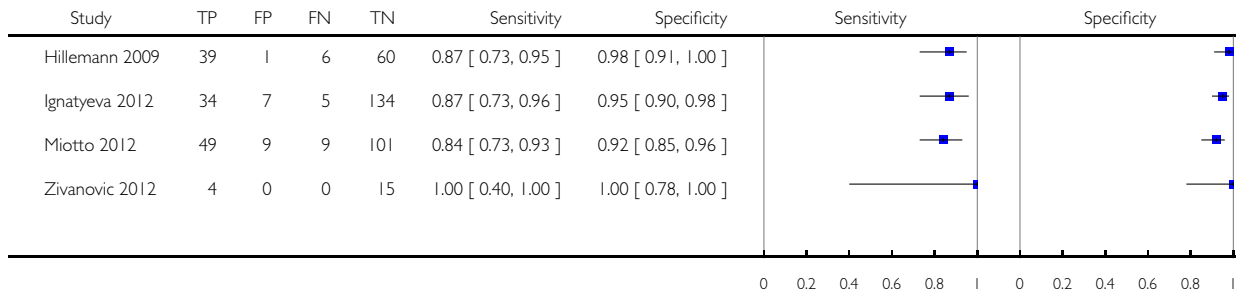
Test: 11 Indirect, amikacin, WHO critical concentration used



### Test 12. Indirect, capreomycin, WHO critical concentration used.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

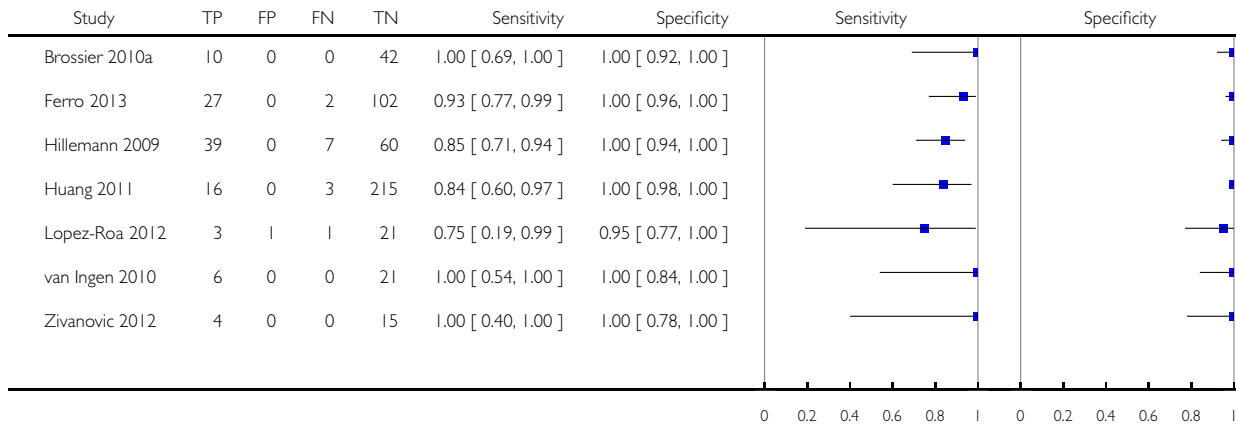
Test: 12 Indirect, capreomycin, WHO critical concentration used



### Test 13. Indirect, amikacin, WHO critical concentration not used.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

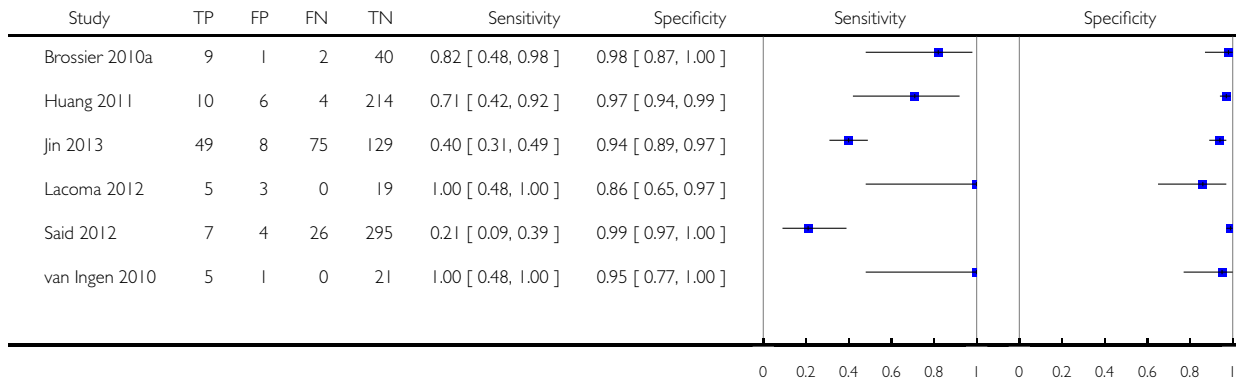
Test: 13 Indirect, amikacin, WHO critical concentration not used



### Test 14. Indirect, capreomycin, WHO critical concentration not used.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

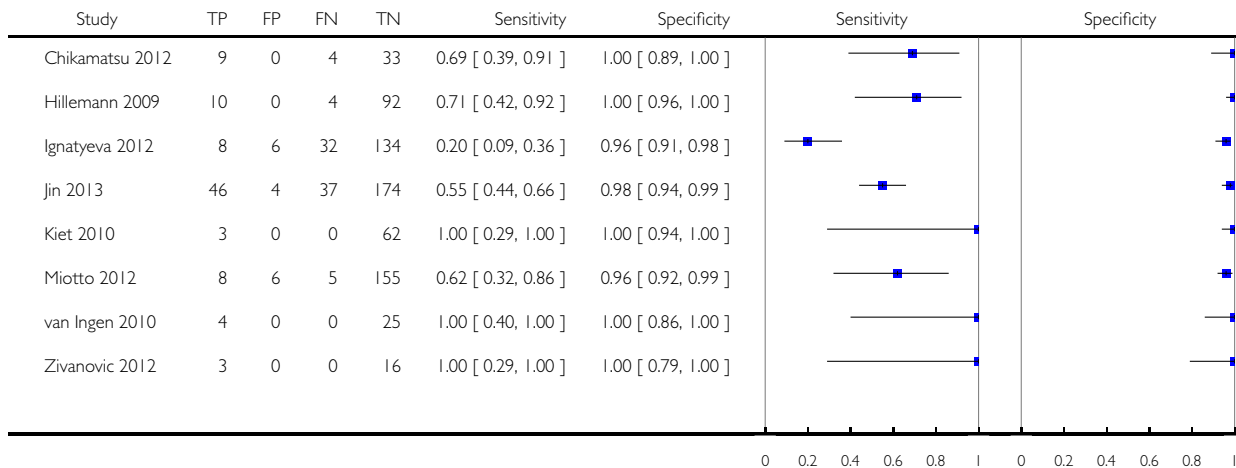
Test: 14 Indirect, capreomycin, WHO critical concentration not used



### Test 15. Indirect, XDR, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

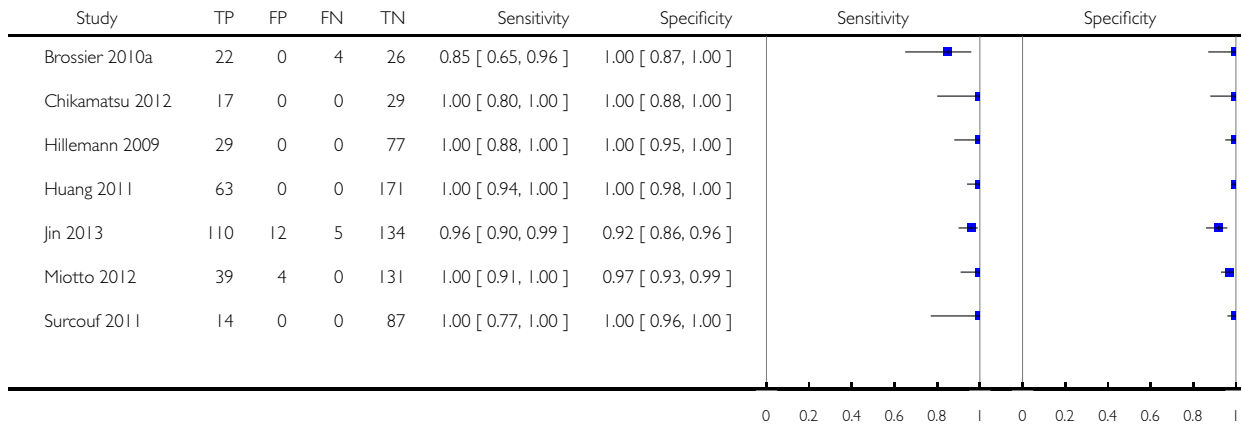
Test: 15 Indirect, XDR, culture



### Test 16. Indirect, FQ, sequencing.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

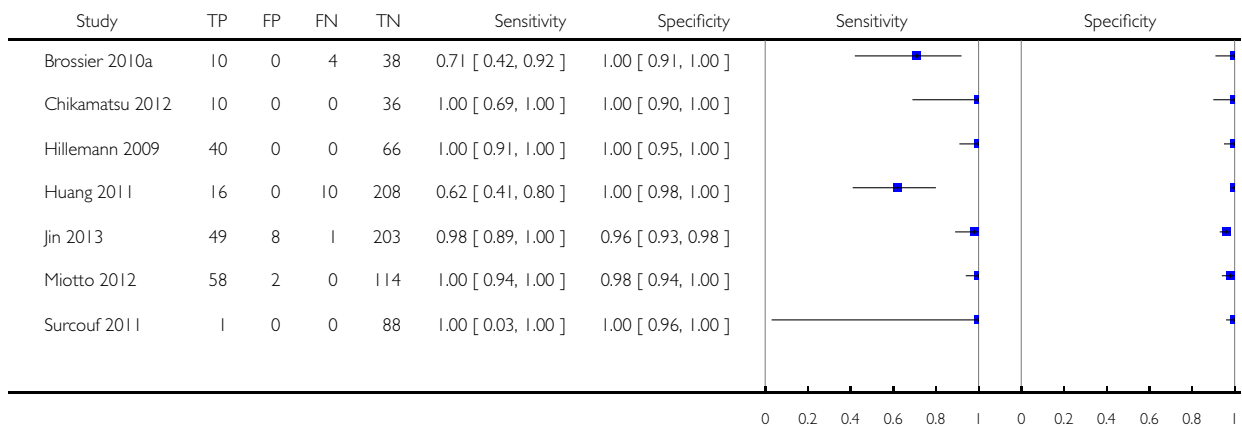
Test: 16 Indirect, FQ, sequencing



### Test 17. Indirect, SLID, sequencing.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

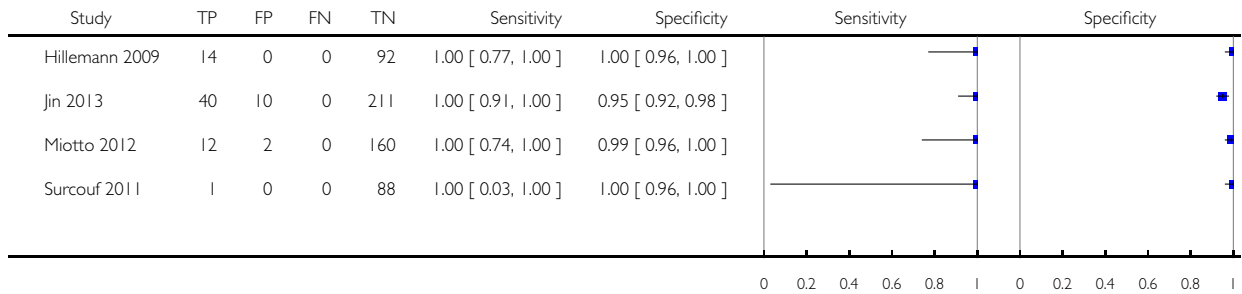
Test: 17 Indirect, SLID, sequencing



### Test 18. Indirect, XDR, sequencing.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

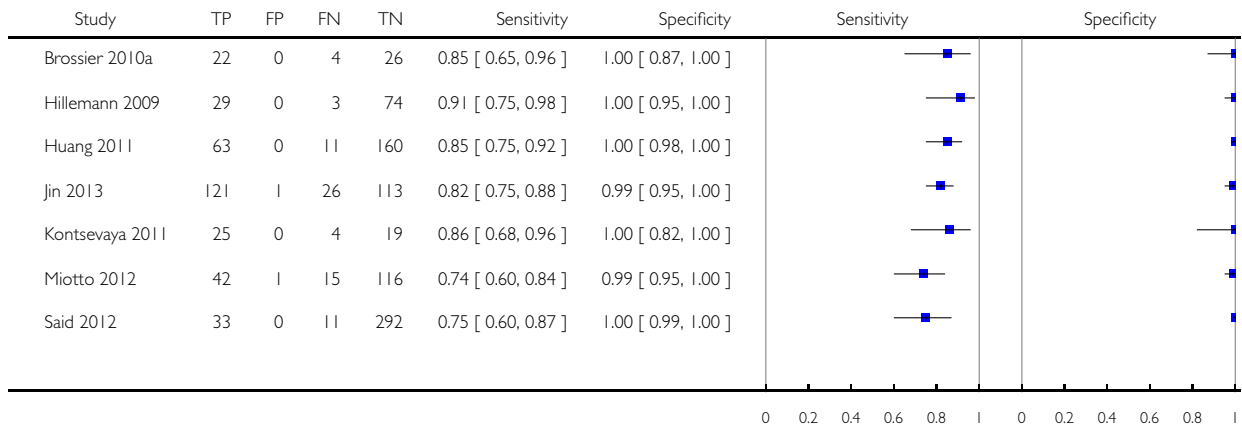
Test: 18 Indirect, XDR, sequencing



### Test 19. Indirect, FQ, sequencing and culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

Test: 19 Indirect, FQ, sequencing and culture

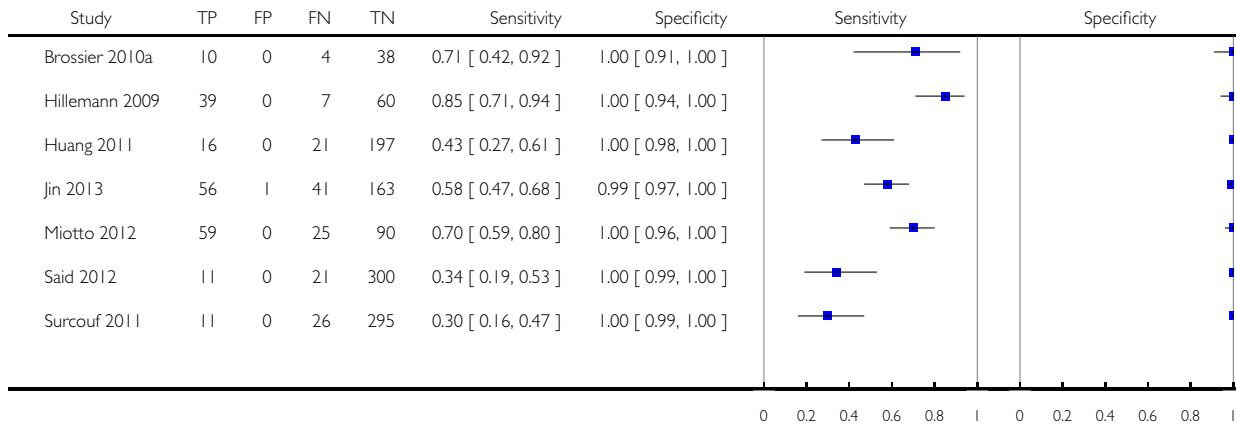




### Test 20. Indirect, SLID, sequencing and culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

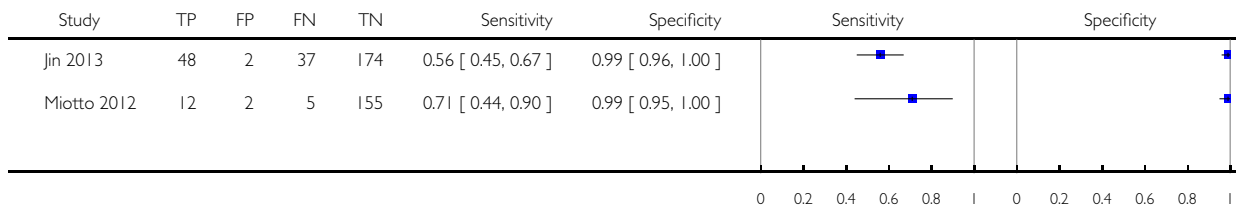
Test: 20 Indirect, SLID, sequencing and culture



### Test 21. Indirect, XDR, sequencing and culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

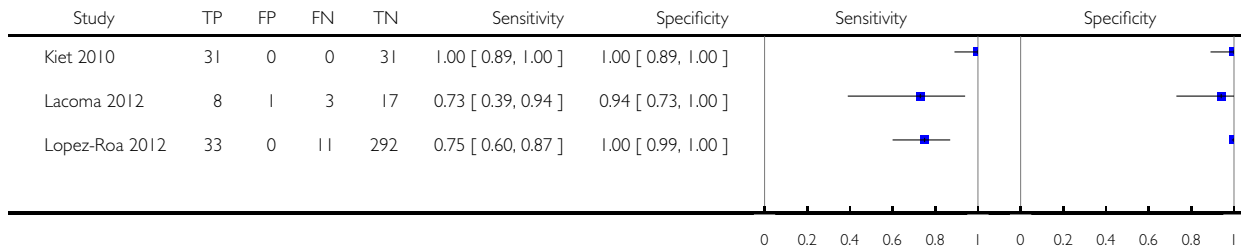
Test: 21 Indirect, XDR, sequencing and culture



### Test 22. Indirect, FQ, culture followed by sequencing of discrepant.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

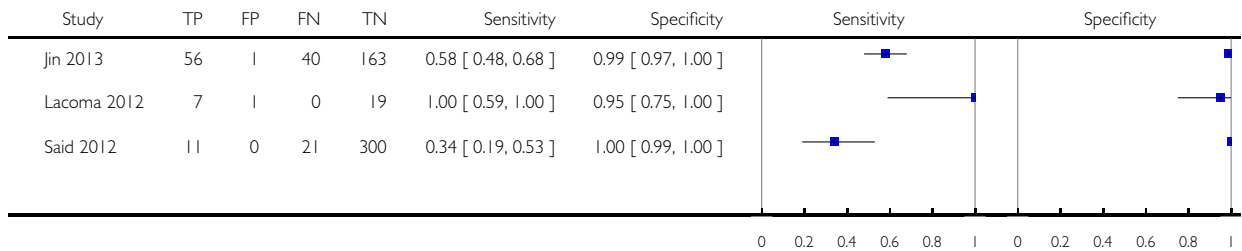
Test: 22 Indirect, FQ, culture followed by sequencing of discrepant



### Test 23. Indirect, SLID, culture followed by sequencing of discrepant.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

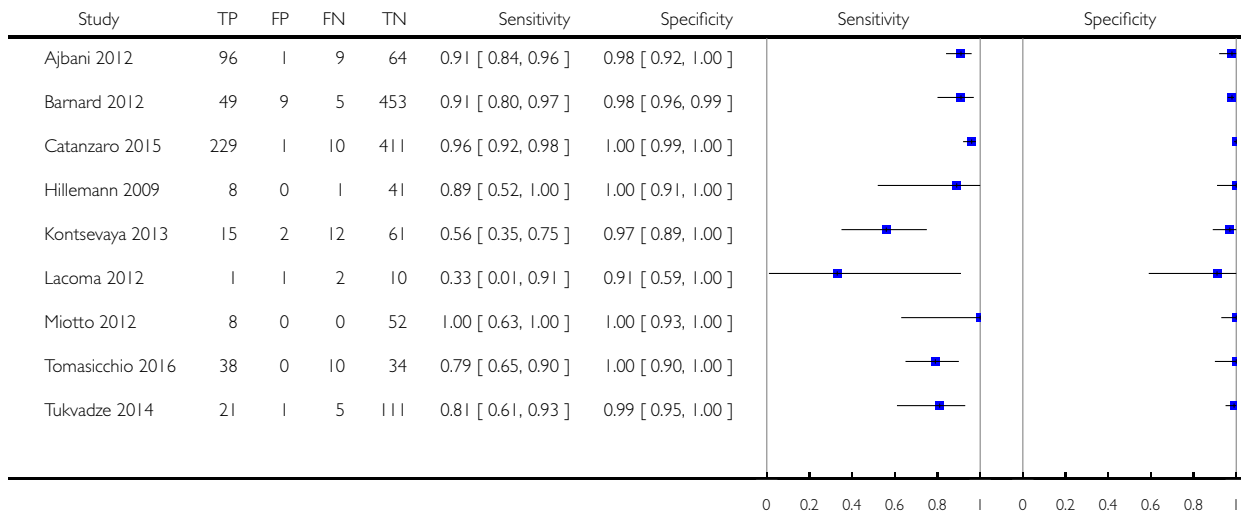
Test: 23 Indirect, SLID, culture followed by sequencing of discrepant



### Test 24. Direct, FQ, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

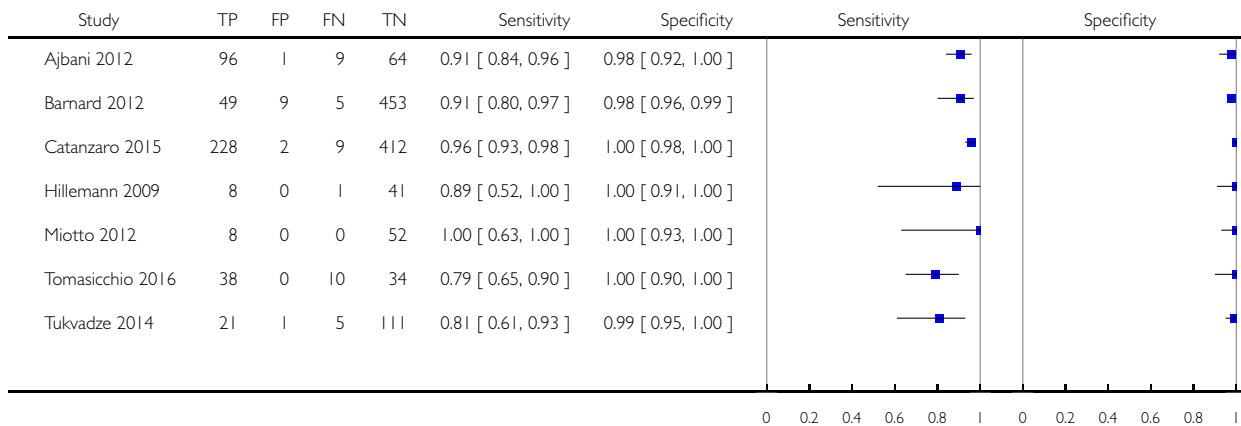
Test: 24 Direct, FQ, culture



### Test 25. Direct, ofloxacin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

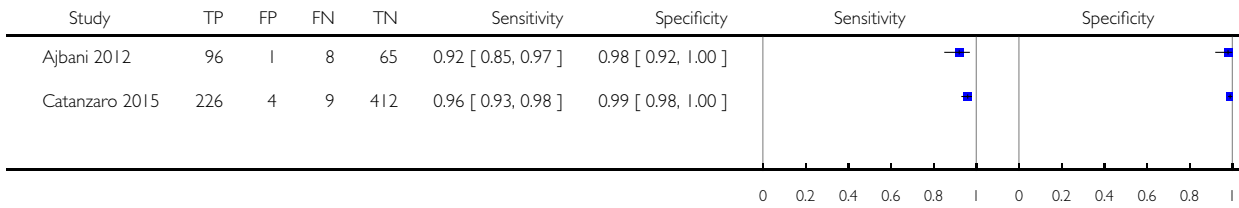
Test: 25 Direct, ofloxacin, culture



### Test 26. Direct, moxifloxacin, culture.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

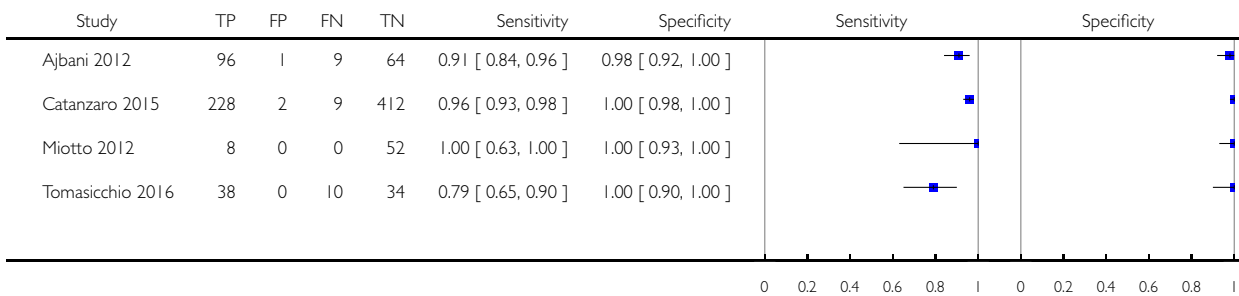
Test: 26 Direct, moxifloxacin, culture



### Test 27. Ofloxacin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

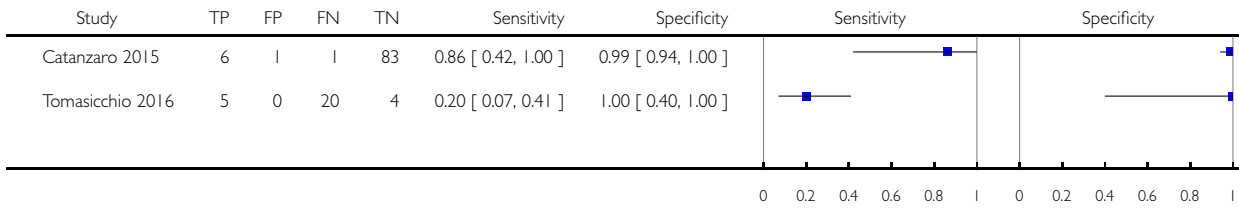
Test: 27 Ofloxacin, smear positive



### Test 28. Ofloxacin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

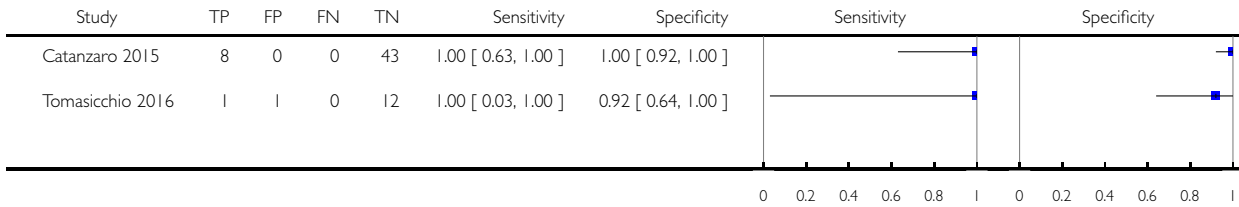
Test: 28 Ofloxacin, smear negative



### Test 29. Ofloxacin, smear grade = scanty.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

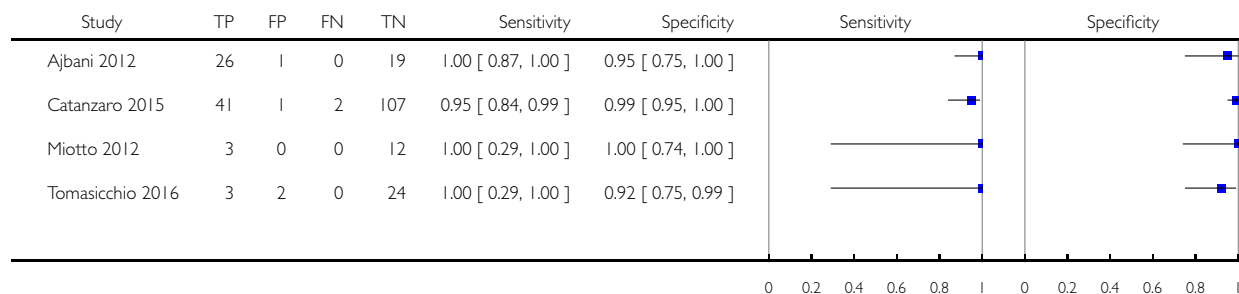
Test: 29 Ofloxacin, smear grade = scanty



### Test 30. Ofloxacin, smear grade = 1+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

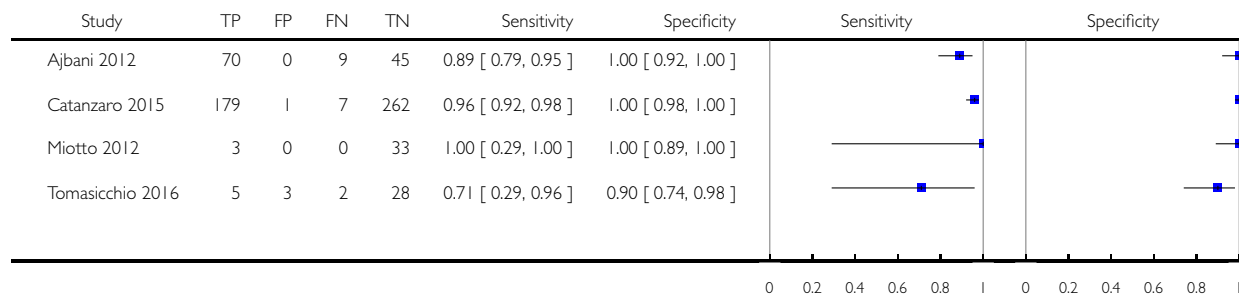
Test: 30 Ofloxacin, smear grade = 1+



### Test 31. Ofloxacin, smear grade ≥ 2+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

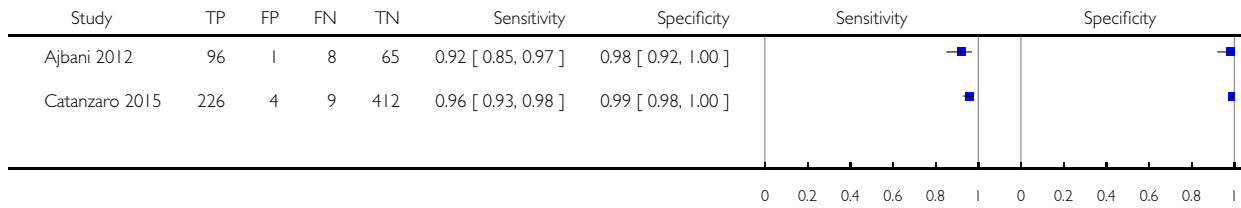
Test: 31 Ofloxacin, smear grade ≥ 2+



### Test 32. Moxifloxacin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

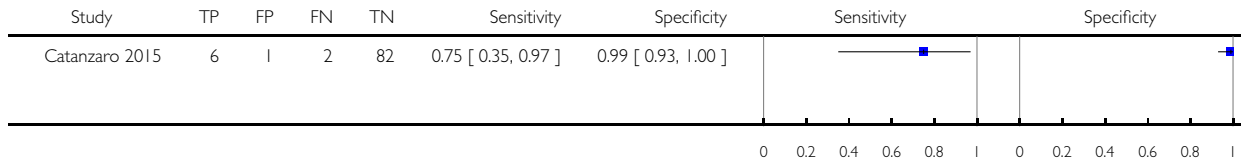
Test: 32 Moxifloxacin, smear positive



### Test 33. Moxifloxacin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

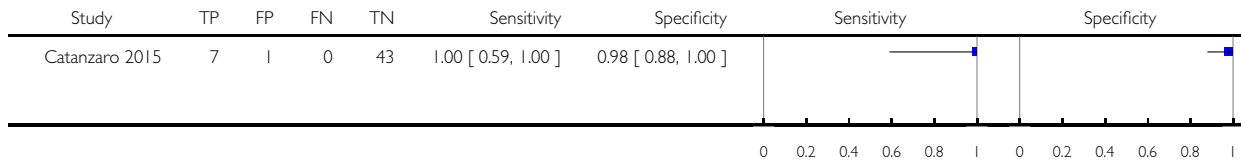
Test: 33 Moxifloxacin, smear negative



### Test 34. Moxifloxacin, smear grade = scanty.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

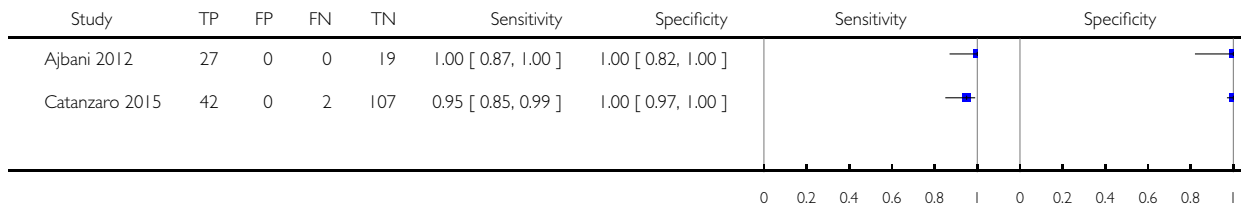
Test: 34 Moxifloxacin, smear grade = scanty



### Test 35. Moxifloxacin, smear grade = I+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

Test: 35 Moxifloxacin, smear grade = I+

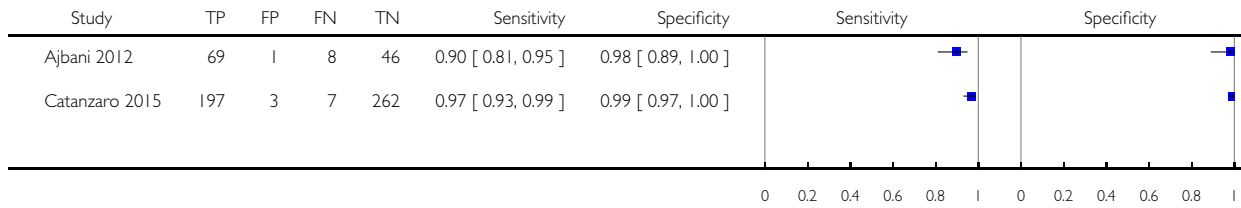




### Test 36. Moxifloxacin, smear grade $\geq 2+$ .

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

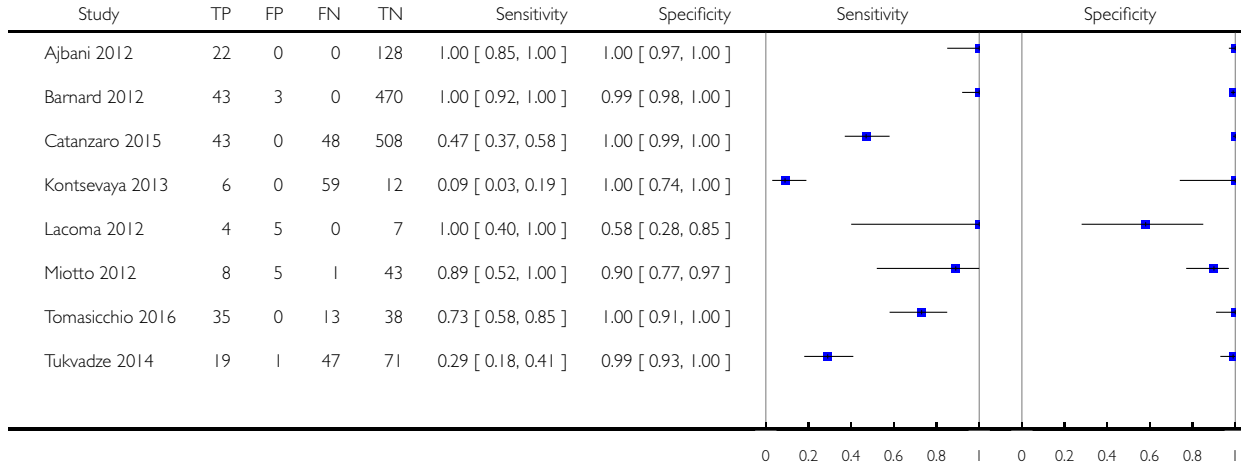
Test: 36 Moxifloxacin, smear grade  $\geq 2+$



### Test 37. Direct, SLID, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

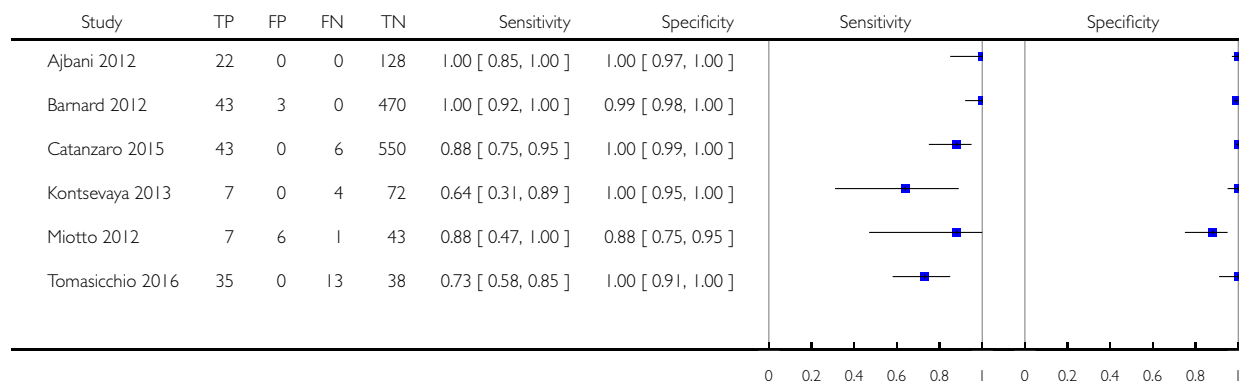
Test: 37 Direct, SLID, culture



### Test 38. Direct, amikacin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

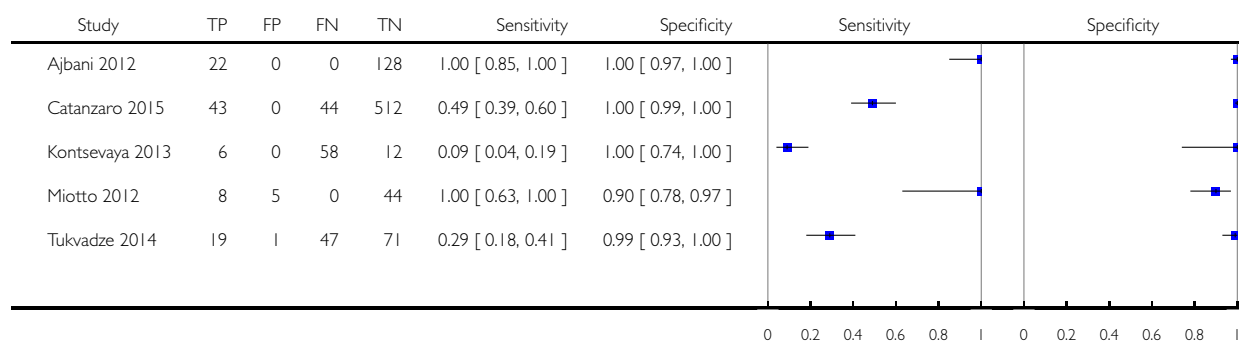
Test: 38 Direct, amikacin, culture



### Test 39. Direct, kanamycin, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

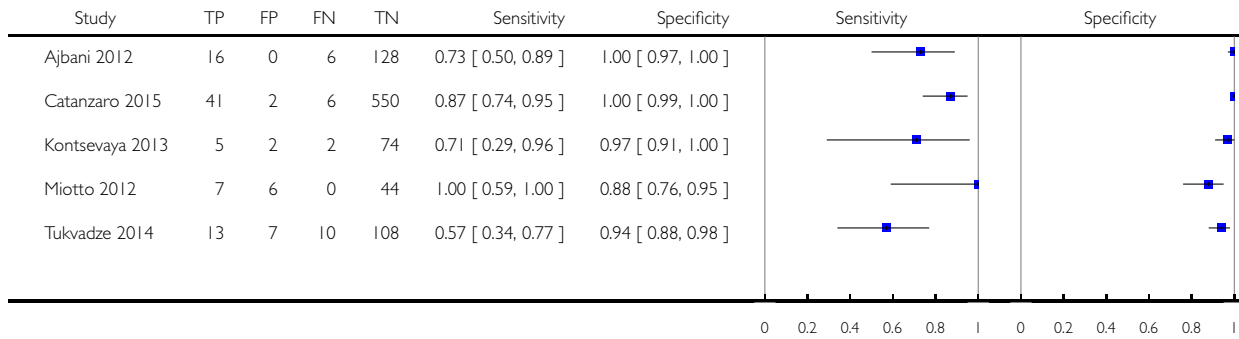
Test: 39 Direct, kanamycin, culture



### Test 40. Direct, capreomycin, culture.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

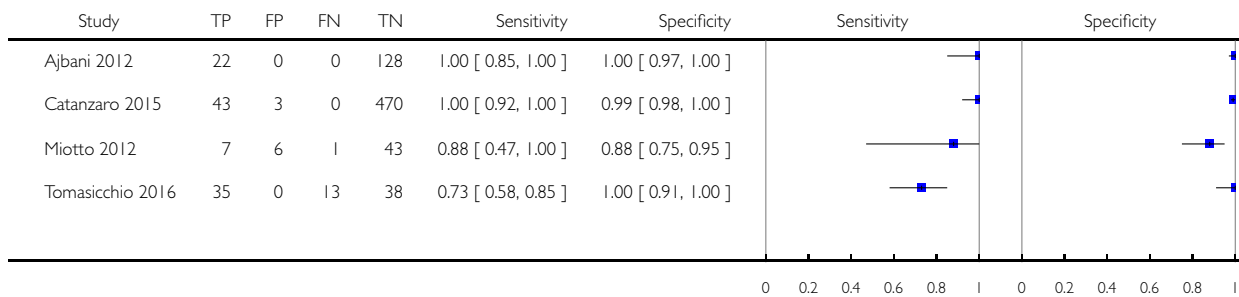
Test: 40 Direct, capreomycin, culture



### Test 41. Amikacin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

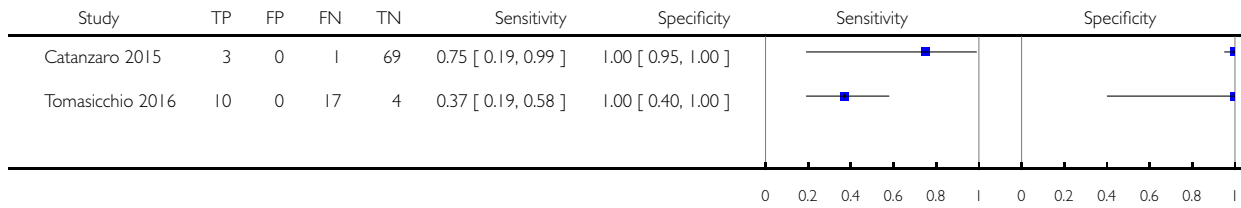
Test: 41 Amikacin, smear positive



### Test 42. Amikacin, smear negative.

Review: GenoType MTBDRs<sub>l</sub> assay for resistance to second-line anti-tuberculosis drugs

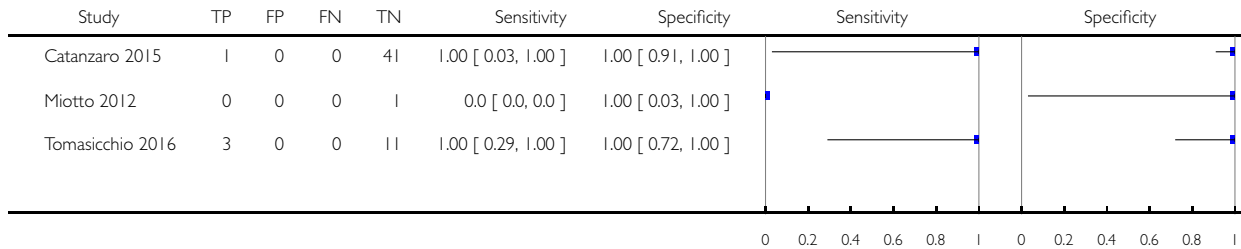
Test: 42 Amikacin, smear negative



### Test 43. Amikacin, smear grade = scanty.

Review: GenoType MTBDRs<sub>l</sub> assay for resistance to second-line anti-tuberculosis drugs

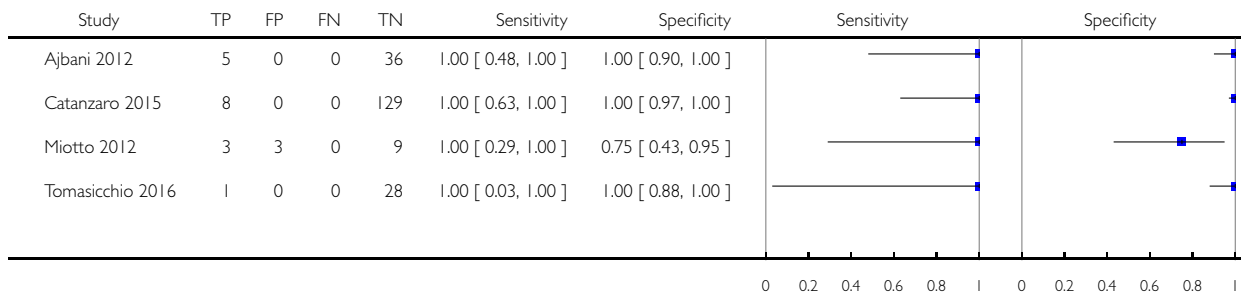
Test: 43 Amikacin, smear grade = scanty



### Test 44. Amikacin, smear grade= 1+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

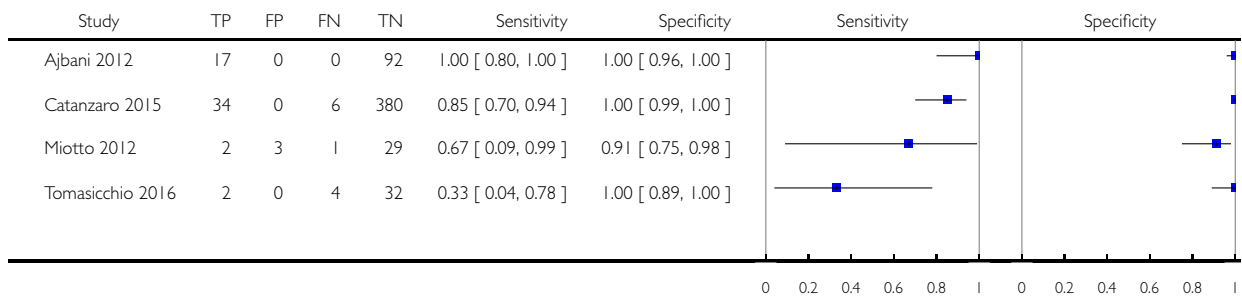
Test: 44 Amikacin, smear grade= 1+



### Test 45. Amikacin, smear grade $\geq$ 2+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

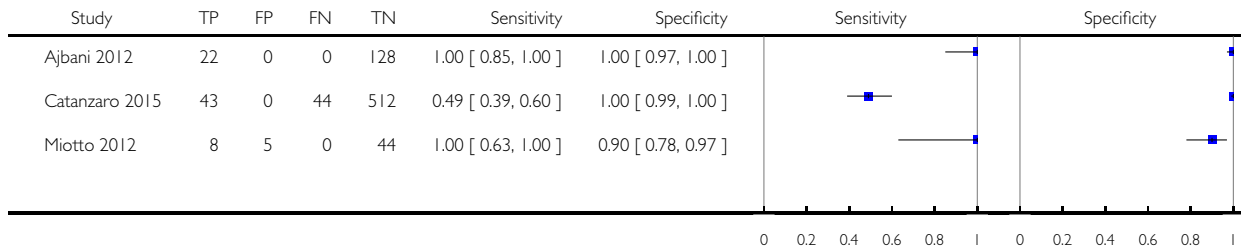
Test: 45 Amikacin, smear grade  $\geq$  2+



### Test 46. Kanamycin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

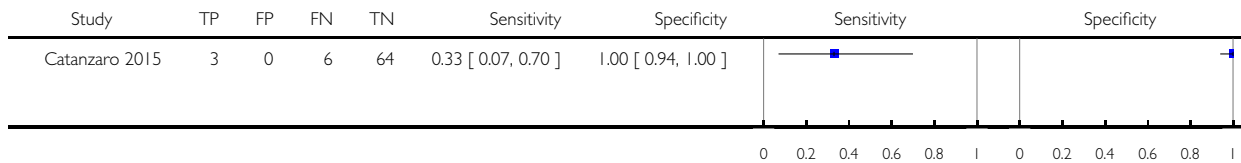
Test: 46 Kanamycin, smear positive



### Test 47. Kanamycin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

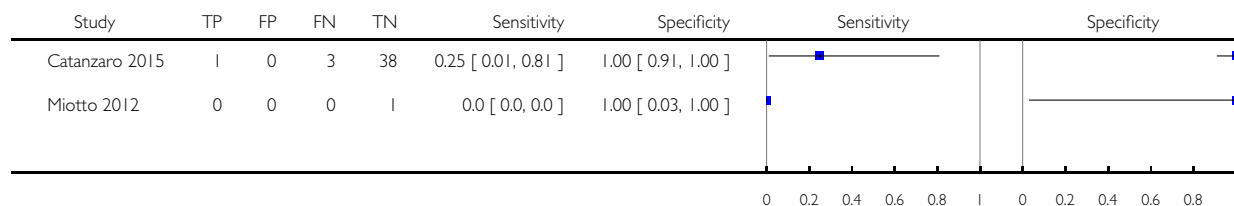
Test: 47 Kanamycin, smear negative



### Test 48. Kanamycin, smear grade = scanty.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

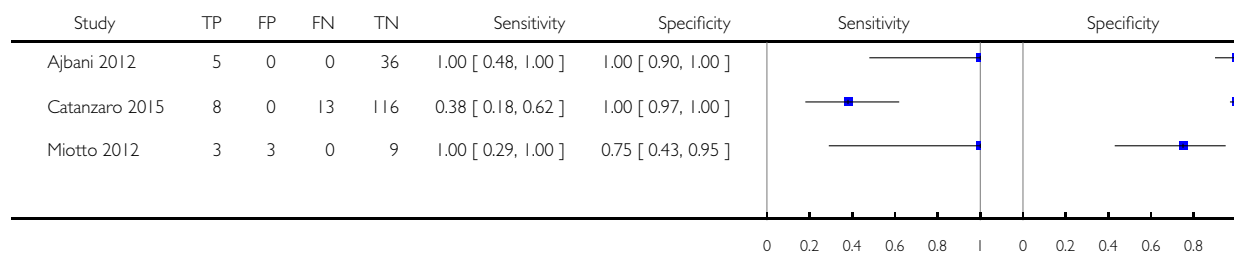
Test: 48 Kanamycin, smear grade = scanty



### Test 49. Kanamycin, smear grade = +1.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

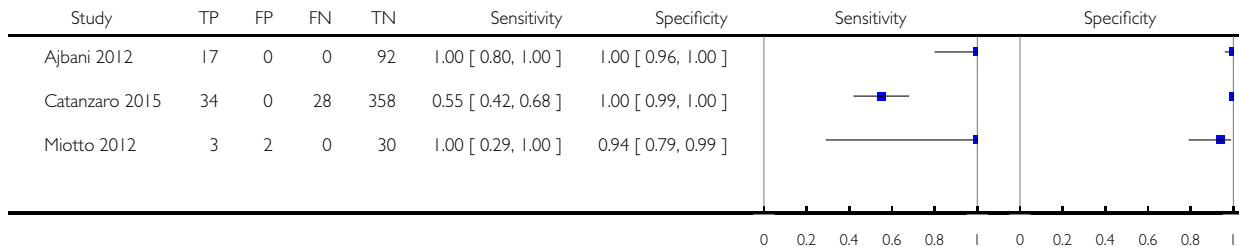
Test: 49 Kanamycin, smear grade = +1



### Test 50. Kanamycin, smear grade $\geq 2+$ .

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

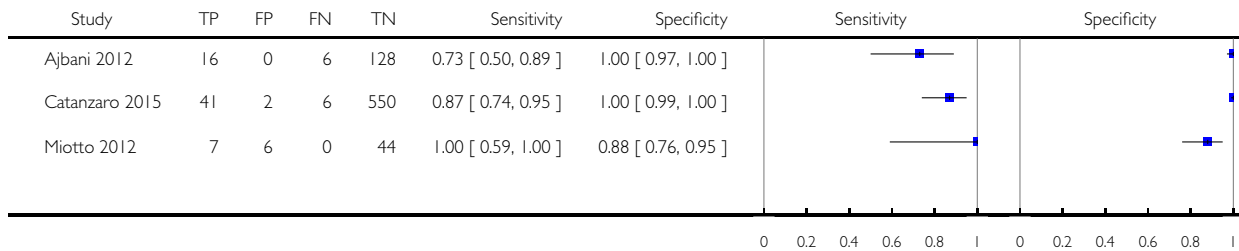
Test: 50 Kanamycin, smear grade  $\geq 2+$



### Test 51. Capreomycin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

Test: 51 Capreomycin, smear positive

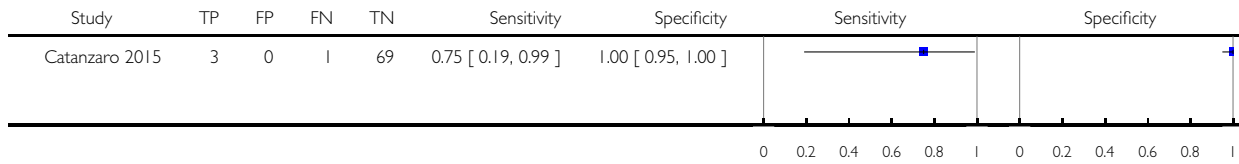




### Test 52. Capreomycin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

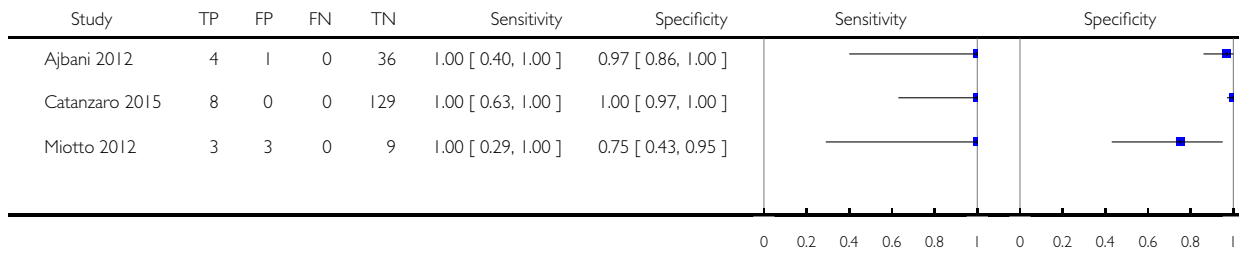
Test: 52 Capreomycin, smear negative



### Test 53. Capreomycin, smear grade = I+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

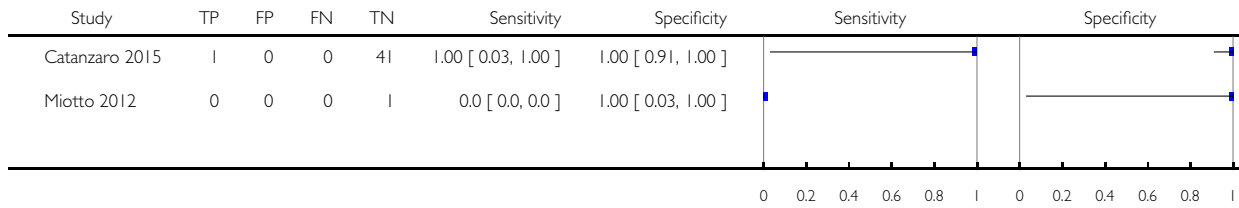
Test: 53 Capreomycin, smear grade = I+



### Test 54. Capreomycin, smear grade = scanty.

Review: GenoType MTBDRs<sub>l</sub> assay for resistance to second-line anti-tuberculosis drugs

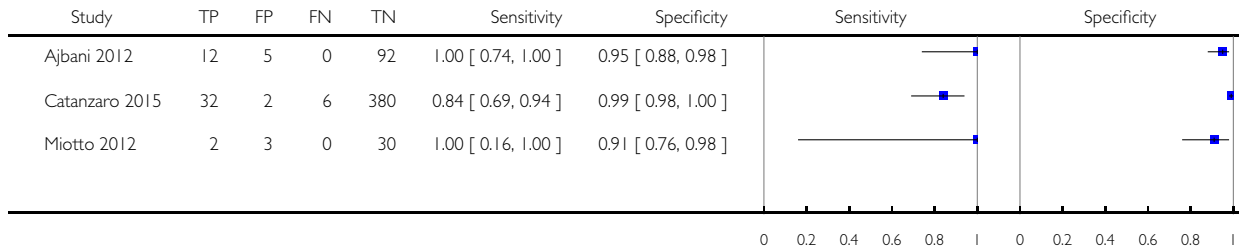
Test: 54 Capreomycin, smear grade = scanty



### Test 55. Capreomycin, smear grade $\geq 2+$ .

Review: GenoType MTBDRs<sub>l</sub> assay for resistance to second-line anti-tuberculosis drugs

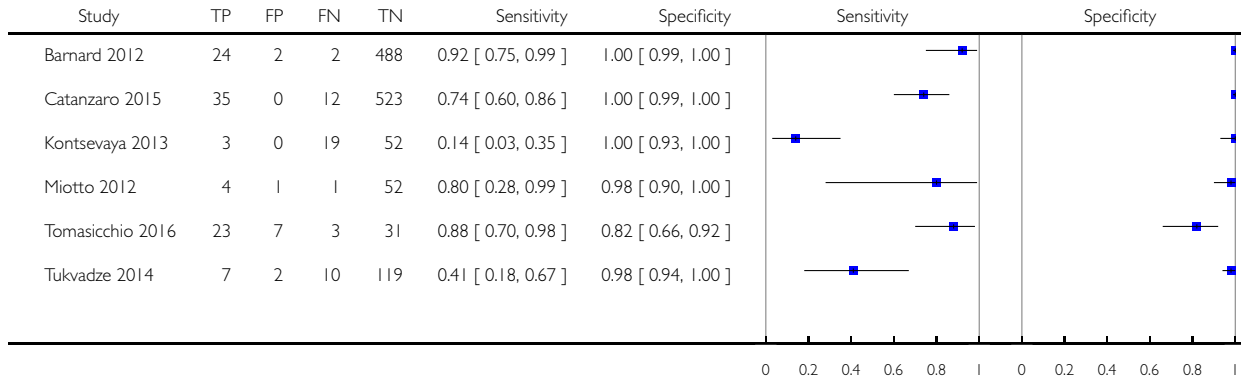
Test: 55 Capreomycin, smear grade  $\geq 2+$



### Test 56. Direct, XDR, culture.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

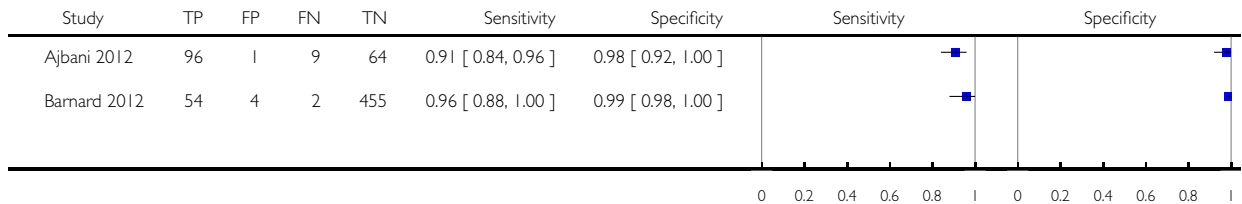
Test: 56 Direct, XDR, culture



### Test 57. Direct, FQ, culture followed by sequencing of discrepant.

Review: GenoType MTBDR<sub>s/l</sub> assay for resistance to second-line anti-tuberculosis drugs

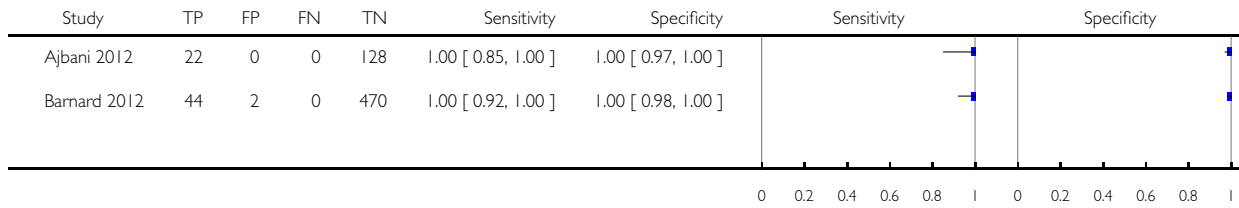
Test: 57 Direct, FQ, culture followed by sequencing of discrepant



### Test 58. Direct, SLID, culture followed by sequencing of discrepant.

Review: GenoType MTBDRs<sub>l</sub> assay for resistance to second-line anti-tuberculosis drugs

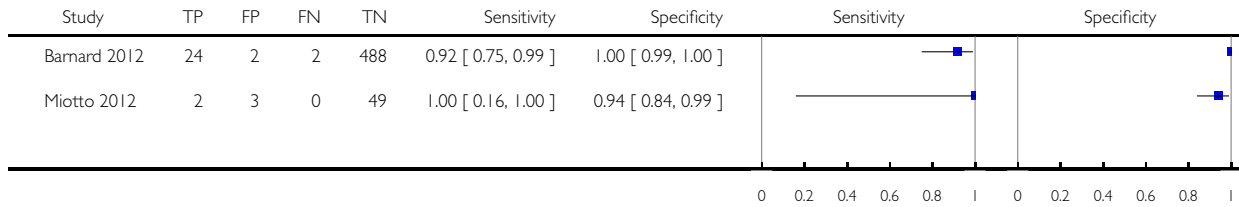
Test: 58 Direct, SLID, culture followed by sequencing of discrepant



### Test 59. Direct, XDR, culture followed by sequencing of discrepant.

Review: GenoType MTBDRs<sub>l</sub> assay for resistance to second-line anti-tuberculosis drugs

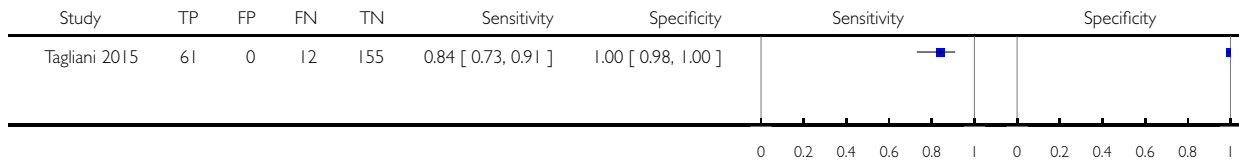
Test: 59 Direct, XDR, culture followed by sequencing of discrepant



### Test 60. V2, Indirect, FQ, culture.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

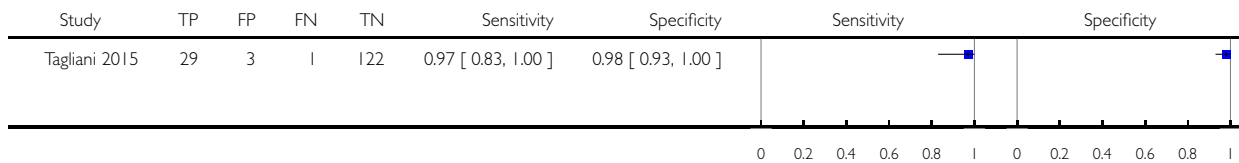
Test: 60 V2, Indirect, FQ, culture



### Test 61. V2, Direct, FQ, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

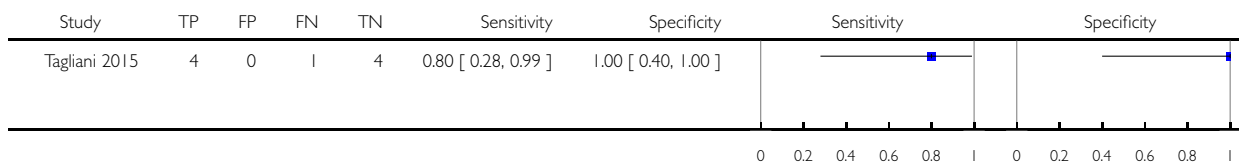
Test: 61 V2, Direct, FQ, smear positive



### Test 62. V2, Direct, FQ, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

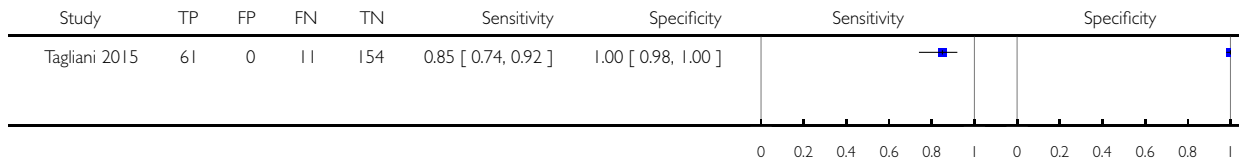
Test: 62 V2, Direct, FQ, smear negative



### Test 63. V2, Indirect, ofloxacin, culture.

Review: GenoType MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs

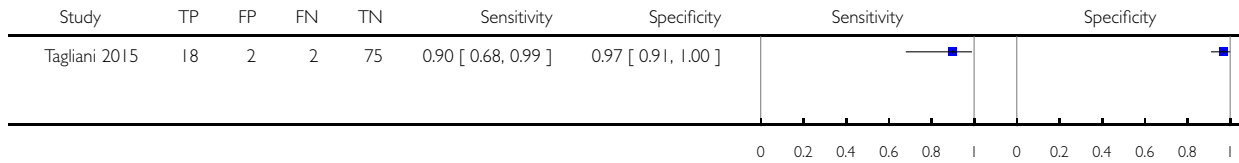
Test: 63 V2, Indirect, ofloxacin, culture



### Test 64. V2, Indirect, moxifloxacin, culture.

Review: GenoType MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs

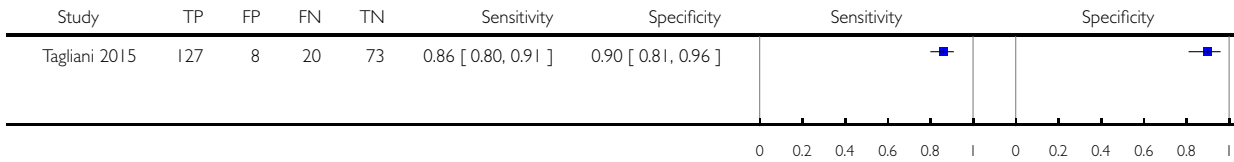
Test: 64 V2, Indirect, moxifloxacin, culture



### Test 65. V2, Indirect, SLID, culture.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

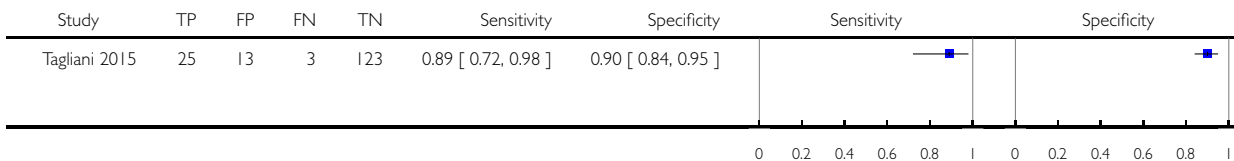
Test: 65 V2, Indirect, SLID, culture



### Test 66. V2, Direct, SLID, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

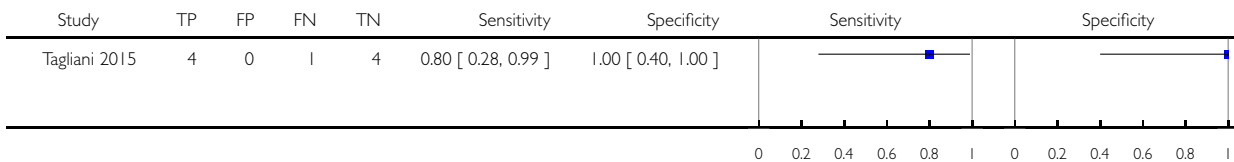
Test: 66 V2, Direct, SLID, smear positive



### Test 67. V2, Direct, SLID, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

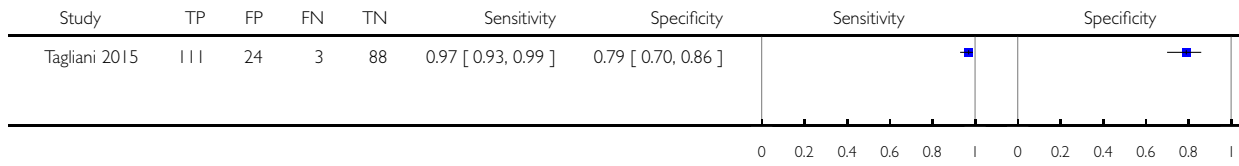
Test: 67 V2, Direct, SLID, smear negative



### Test 68. V2, Indirect, amikacin, culture.

Review: GenoType MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs

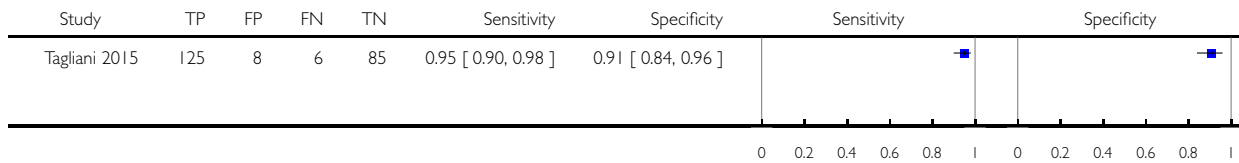
Test: 68 V2, Indirect, amikacin, culture



### Test 69. V2, Indirect, kanamycin, culture.

Review: GenoType MTBDR<sub>sl</sub> assay for resistance to second-line anti-tuberculosis drugs

Test: 69 V2, Indirect, kanamycin, culture

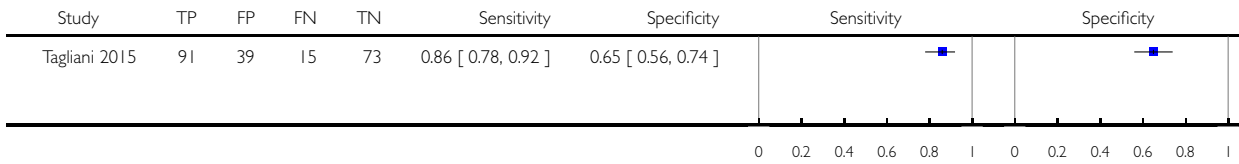




### Test 70. V2, Indirect, capreomycin, culture.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

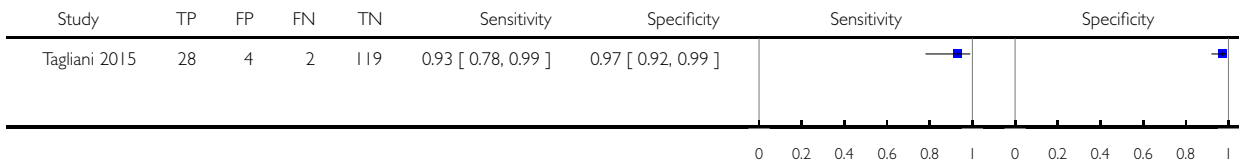
Test: 70 V2, Indirect, capreomycin, culture



### Test 71. V2, Ofloxacin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

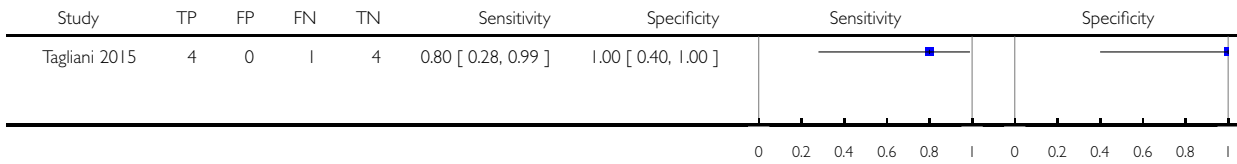
Test: 71 V2, Ofloxacin, smear positive



### Test 72. V2, Ofloxacin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

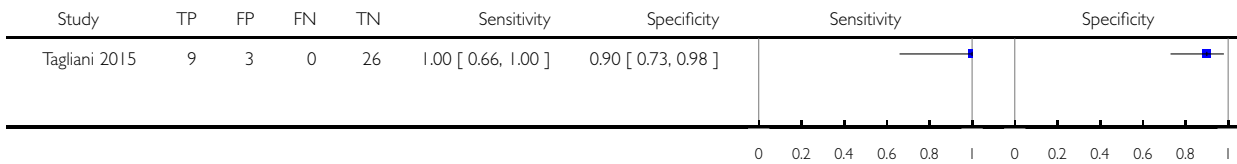
Test: 72 V2, Ofloxacin, smear negative



### Test 73. V2, Ofloxacin, smear grade = scanty.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

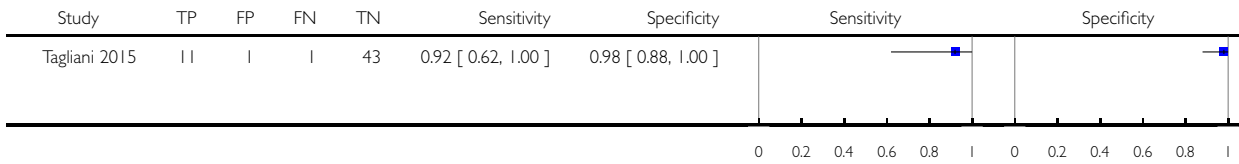
Test: 73 V2, Ofloxacin, smear grade = scanty



### Test 74. V2, Ofloxacin, smear grade = 1+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

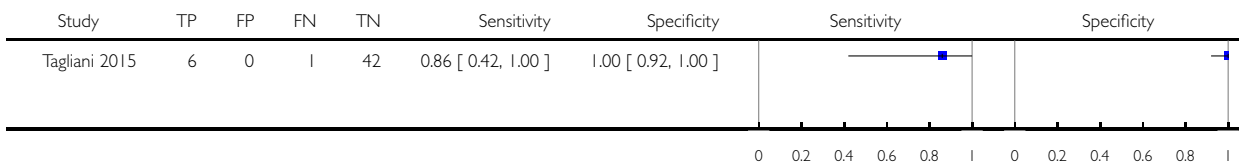
Test: 74 V2, Ofloxacin, smear grade = 1+



### Test 75. V2, Ofloxacin, smear grade ≥ 2+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

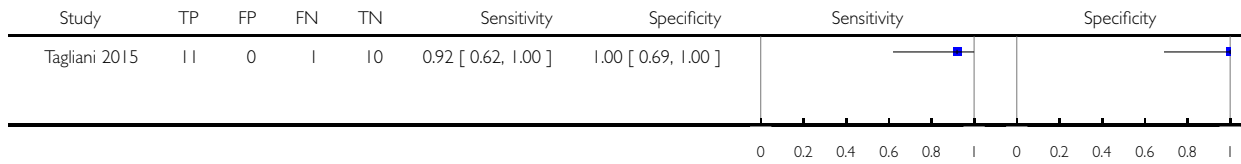
Test: 75 V2, Ofloxacin, smear grade ≥ 2+



### Test 76. V2, Moxifloxacin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

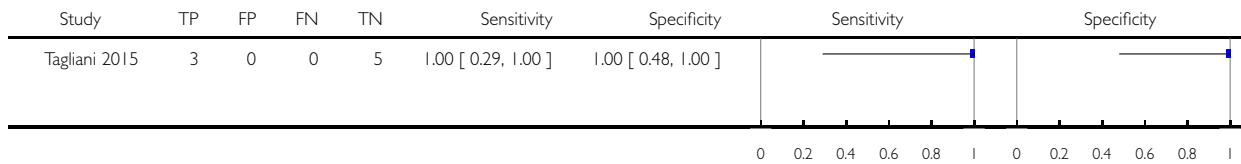
Test: 76 V2, Moxifloxacin, smear positive



### Test 77. V2, Moxifloxacin, smear grade $\geq 2+$ .

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

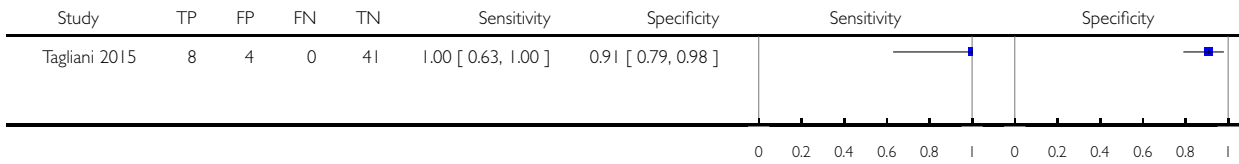
Test: 77 V2, Moxifloxacin, smear grade  $\geq 2+$



### Test 78. V2, Levofloxacin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

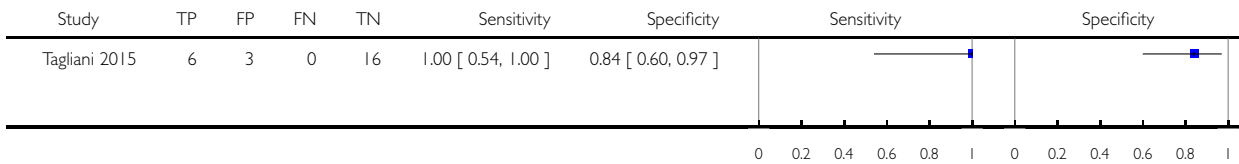
Test: 78 V2, Levofloxacin, smear positive



### Test 79. V2, Levofloxacin, smear grade = scanty.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

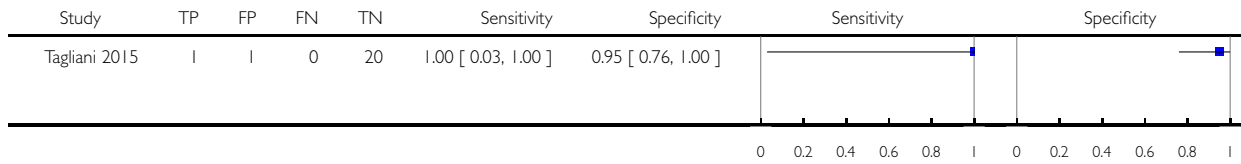
Test: 79 V2, Levofloxacin, smear grade = scanty



### Test 80. V2, Levofloxacin, smear grade = 1+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

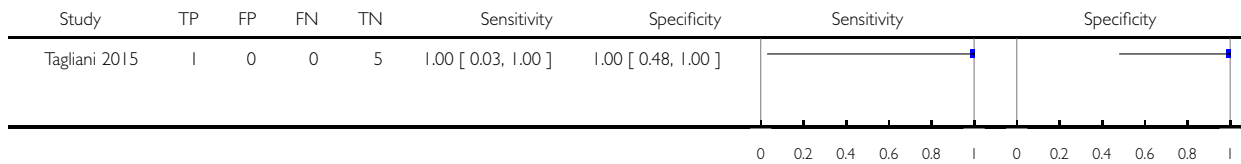
Test: 80 V2, Levofloxacin, smear grade = 1+



### Test 81. V2, Levofloxacin, smear grade ≥ 2+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

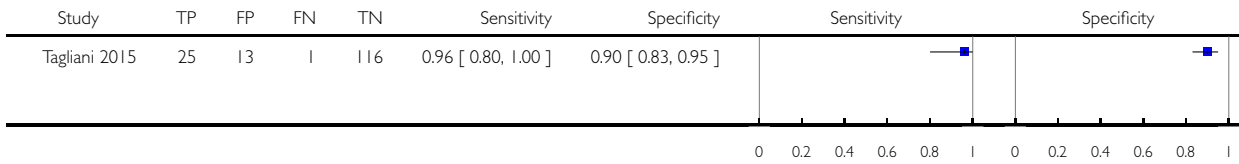
Test: 81 V2, Levofloxacin, smear grade ≥ 2+



### Test 82. V2, Amikacin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

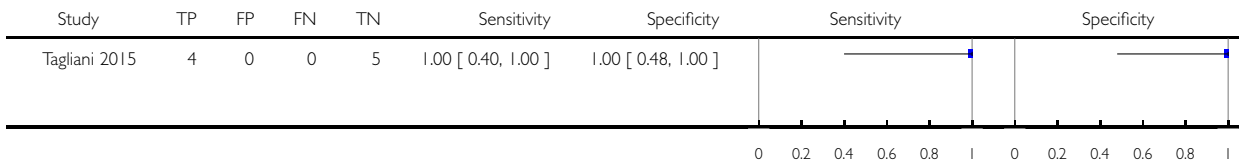
Test: 82 V2, Amikacin, smear positive



### Test 83. V2, Amikacin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

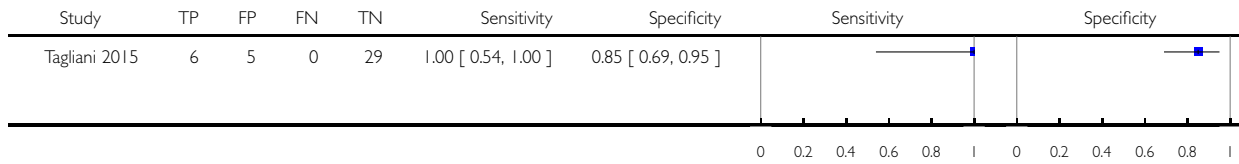
Test: 83 V2, Amikacin, smear negative



### Test 84. V2, Amikacin, smear grade = scanty.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

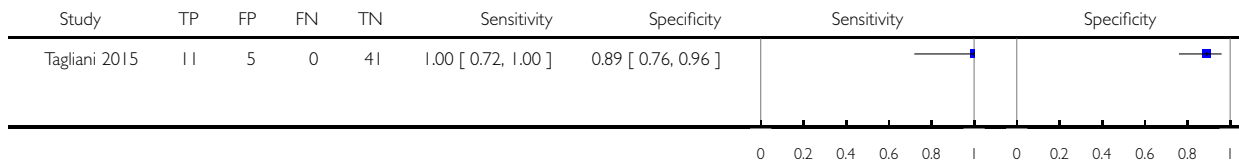
Test: 84 V2, Amikacin, smear grade = scanty



### Test 85. V2, Amikacin, smear grade = I+.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

Test: 85 V2, Amikacin, smear grade = I+

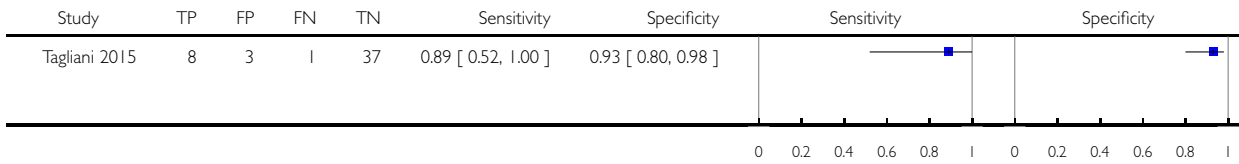




### Test 86. V2, Amikacin, smear grade $\geq 2+$ .

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

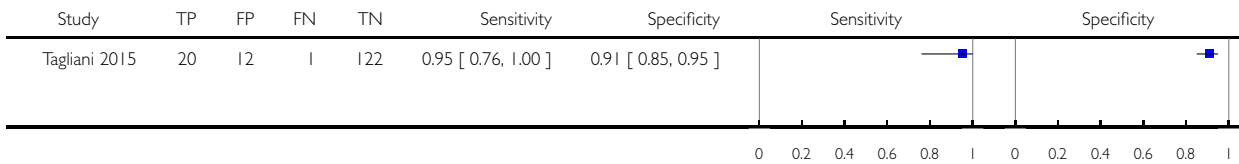
Test: 86 V2, Amikacin, smear grade  $\geq 2+$



### Test 87. V2, Kanamycin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

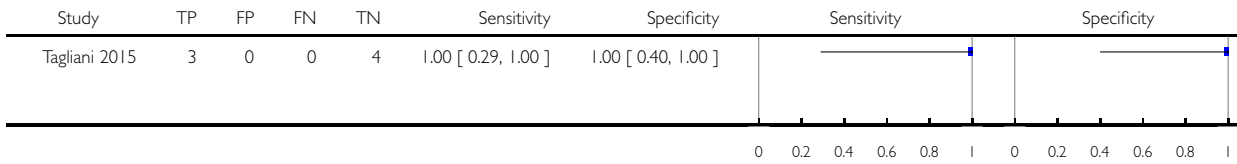
Test: 87 V2, Kanamycin, smear positive



### Test 88. V2, Kanamycin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

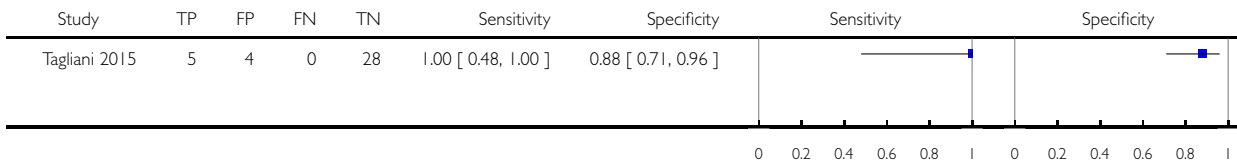
Test: 88 V2, Kanamycin, smear negative



### Test 89. V2, Kanamycin, smear grade = scanty.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

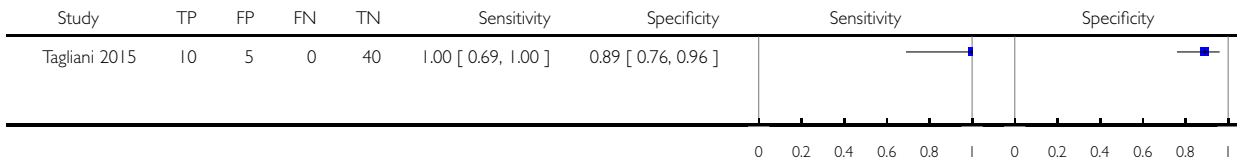
Test: 89 V2, Kanamycin, smear grade = scanty



### Test 90. V2, Kanamycin, smear grade = 1+

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

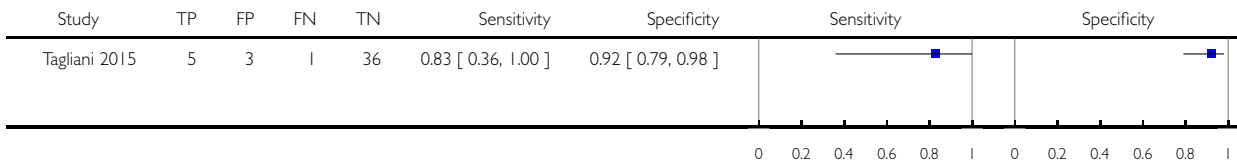
Test: 90 V2, Kanamycin, smear grade = 1+



### Test 91. V2, Kanamycin, smear grade ≥ 2+

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

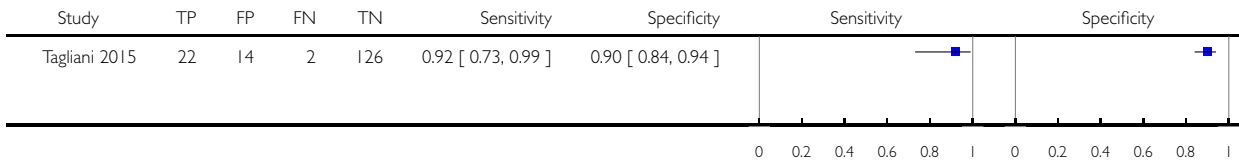
Test: 91 V2, Kanamycin, smear grade ≥ 2+



### Test 92. V2, Capreomycin, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

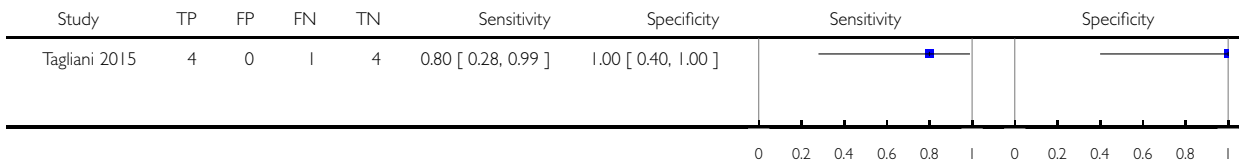
Test: 92 V2, Capreomycin, smear positive



### Test 93. V2, Capreomycin, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

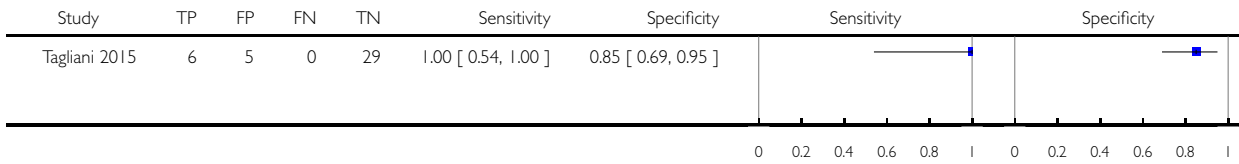
Test: 93 V2, Capreomycin, smear negative



**Test 94. V2, Capreomycin, smear grade = scanty.**

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

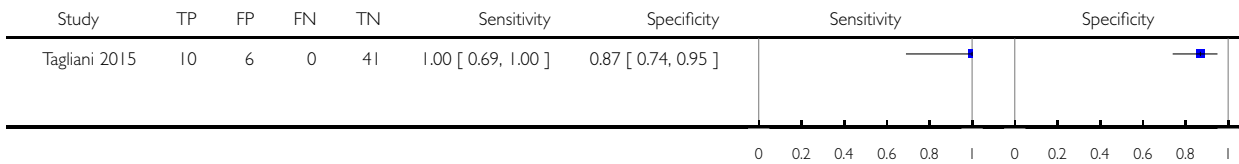
Test: 94 V2, Capreomycin, smear grade = scanty



**Test 95. V2, Capreomycin, smear grade = I+.**

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

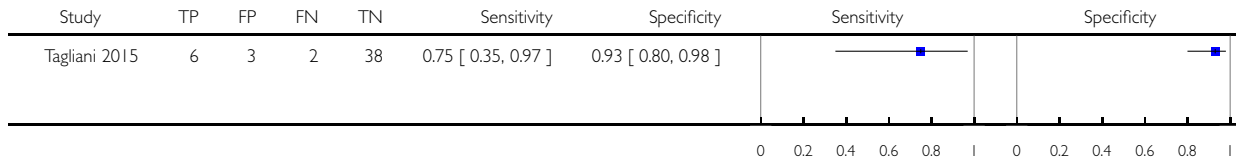
Test: 95 V2, Capreomycin, smear grade = I+



### Test 96. V2, Capreomycin, smear grade $\geq 2+$ .

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

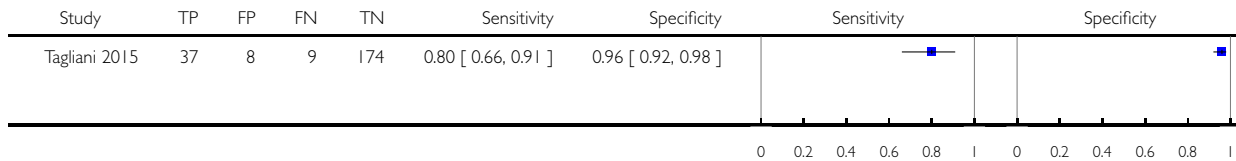
Test: 96 V2, Capreomycin, smear grade  $\geq 2+$



### Test 97. V2, Indirect, XDR, culture.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

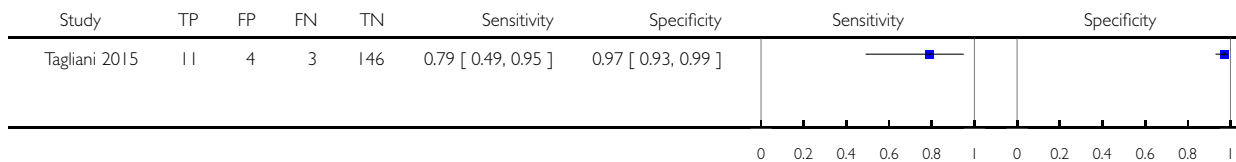
Test: 97 V2, Indirect, XDR, culture



### Test 98. V2, Direct, XDR, smear positive.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

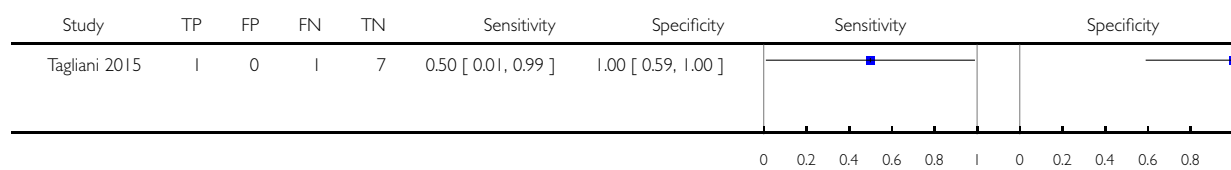
Test: 98 V2, Direct, XDR, smear positive



### Test 99. V2, Direct, XDR, smear negative.

Review: GenoType MTBDR<sub>s</sub>/ assay for resistance to second-line anti-tuberculosis drugs

Test: 99 V2, Direct, XDR, smear negative



## ADDITIONAL TABLES

Table 1. Characteristics of MTBDR<sub>s</sub>/ versions 1.0 and 2.0

| Detection             | Version 1.0: <i>M. tuberculosis</i> complex and resistances to FQs, SLIDs, and ethambutol | Version 2.0: <i>M. tuberculosis</i> complex and resistances to FQs and SLIDs                  |
|-----------------------|---|---|
| Samples               | Smear-positive specimens and culture isolates   | Smear-positive and smear-negative specimens and culture isolates                              |
| FQ resistance         | Mutations in resistance determining region of the <i>gyrA</i> gene                        | Mutations in resistance determining regions of the <i>gyrA</i> and <i>gyrB</i> genes          |
| SLID resistance       | Mutations in resistance determining region of the <i>rrs</i> gene                         | Mutations in resistance determining region <i>rrs</i> gene and the <i>eis</i> promoter region |
| Ethambutol resistance | Mutations in the <i>embB</i> gene   | Not included  |

Abbreviations: FQ: fluoroquinolone; SLID: second-line injectable drug.

MTBDR<sub>s</sub>/ reports on the presence of mutations within genes (*gyrA* and *rrs* for version 1.0 and, in addition, *gyrB* and the *eis* promoter for version 2.0), which are associated with resistance to a class of drugs. The presence of mutation(s) in these regions does not necessarily imply resistance to all the drugs within that class. Although specific mutations within these regions may be associated with different levels of resistance to each drug within these classes, the extent of this poorly understood

**Table 2. Map of review showing the number of studies evaluating MTBDR<sub>s/l</sub> version 1.0 by indirect testing, according to the reference standard and target condition**

| Target condition, drug resistance to... | Reference standard       |                     |                                 |   |
|---|--------------------------|---------------------|---------------------------------|---|
|   | Culture, n/N (%)         | Sequencing, n/N (%) | Sequencing and culture, n/N (%) | Culture followed by sequencing of discrepant results, n/N (%) |
| FQs                                     | 19/19 (100) <sup>1</sup> | 7/19 (37)           | 7/19 (37)                       | 3/19 (16)   |
| Ofloxacin                               | 13/19 (68)               | 0                   | 0                               | 0   |
| Moxifloxacin                            | 6/19 (32)                | 0                   | 0                               | 0   |
| Levofloxacin                            | 2/19 (11)                | 0                   | 0                               | 0   |
| SLIDs                                   | 16/16 (100) <sup>1</sup> | 7/16 (44)           | 7/16 (44)                       | 3/16 (19)   |
| Amikacin                                | 11/16 (69)               | 0                   | 0                               | 0   |
| Kanamycin                               | 9/16 (56)                | 0                   | 0                               | 0   |
| Capreomycin                             | 10/16 (63)               | 0                   | 0                               | 0   |
| XDR-TB                                  | 8/8 (100)                | 3/8 (38)            | 2/8 (25)                        | 0   |

Abbreviations: FQ: fluoroquinolone; SLID: second-line injectable drug; TB: tuberculosis; XDR-TB: extensively drug-resistant TB.

<sup>1</sup>A total of 19, 16, and 8 studies were included that evaluated MTBDR<sub>s/l</sub> for FQ resistance, SLID resistance, and XDR-TB, respectively, against culture-based DST. These form the denominators to generate percentages of studies that included a particular additional reference standard.

**Table 3. Map of review showing the number of studies evaluating MTBDR<sub>s/l</sub> version 1.0 by direct testing, according to the reference standard and target condition**

| Target condition, drug resistance to... | Reference standard     |                     |                                 |   |
|---|------------------------|---------------------|---------------------------------|---|
|   | Culture, n/N (%)       | Sequencing, n/N (%) | Sequencing and culture, n/N (%) | Culture followed by sequencing of discrepant results, n/N (%) |
| FQs                                     | 9/9 (100) <sup>1</sup> | 0                   | 0                               | 2/9 (22)  |
| Ofloxacin                               | 7/9 (78)               | 0                   | 0                               | 2/9 (11)  |
| Moxifloxacin                            | 2/9 (22)               | 0                   | 0                               | 0   |
| Levofloxacin                            | 0                      | 0                   | 0                               | 0   |



**Table 3. Map of review showing the number of studies evaluating MTBDRs/ version 1.0 by direct testing, according to the reference standard and target condition** (Continued)

|              |                        |   |   |          |
|--------------|------------------------|---|---|----------|
| Gatifloxacin | 0                      | 0 | 0 | 0        |
| SLIDs        | 8/8 (100) <sup>1</sup> | 0 | 0 | 2/8 (25) |
| Amikacin     | 6/8 (75)               | 0 | 0 | 1/8 (13) |
| Kanamycin    | 5/8 (63)               | 0 | 0 | 0        |
| Capreomycin  | 5/8 (63)               | 0 | 0 | 0        |
| XDR-TB       | 6/6 (100)              | 0 | 0 | 2/6 (33) |

Abbreviations: FQ: fluoroquinolone; SLID: second-line injectable drug; TB: tuberculosis; XDR-TB: extensively drug-resistant TB.

<sup>1</sup>We included a total of 9, 8, and 6 studies that evaluated MTBDRs/ for detection of FQ resistance, SLID resistance, and XDR-TB, respectively, against culture-based DST. These form the denominators to generate percentages of studies that included a particular additional reference standard.

**Table 4. Accuracy of MTBDRs/ version 1.0 for resistance to FQs and SLIDs and XDR-TB, by type of testing, culture-based DST reference standard**

| Pooled sensitivity (95% CI)                                       | Pooled specificity (95% CI) | Pooled sensitivity (95% CI)                                     | Pooled specificity (95% CI) | Pooled sensitivity P value <sup>1</sup> | Pooled specificity P value <sup>1</sup> |
|---|-----------------------------|---|-----------------------------|---|---|
| <b>FQs, indirect testing</b><br>(19 studies, 2223 participants)   |                             | <b>FQs, direct testing</b><br>(9 studies, 1771 participants)    |                             | 0.932                                   | 0.333                                   |
| 85.6% (79.2 to 90.4)  | 98.5% (95.7 to 99.5)        | 86.2% (74.6 to 93.0)  | 98.6% (96.9 to 99.4)        |   |   |
| <b>SLIDs, indirect testing</b><br>(16 studies, 1921 participants) |                             | <b>SLIDs, direct testing</b><br>(8 studies, 1639 participants)  |                             | 0.547                                   | 0.664                                   |
| 76.5% (63.3 to 86.0)  | 99.1% (97.3 to 99.7)        | 87.0% (38.1 to 98.6)  | 99.5% (93.6 to 100.0)       |   |   |
| <b>XDR-TB, indirect testing</b><br>(8 studies, 880 participants)  |                             | <b>XDR-TB, direct testing</b><br>(6 studies, 1420 participants) |                             | 0.888                                   | 0.855                                   |
| 70.9% (42.9 to 88.7)  | 98.8% (96.1 to 99.6)        | 69.4% (38.8 to 89.0)  | 99.4% (95.0 to 99.3)        |   |   |

Abbreviations: CI: confidence interval; DST: drug susceptibility testing; FQ: fluoroquinolone; SLID: second-line injectable drug; TB: tuberculosis; XDR-TB: extensively drug-resistant TB.

The accuracy estimates were derived from non-comparative studies of test accuracy in which different sets of studies were used. For example, for FQ resistance, the set of studies used for indirect testing differed from that used for direct testing.

<sup>1</sup>Likelihood ratio test for evidence of a significant difference between accuracy estimates.

**Table 5. Accuracy of MTBDRs<sub>1</sub> version 1.0 for resistance to select FQ and SLID drugs, culture-based DST reference standard**

| Pooled sensitivity (95% CI)  | Pooled specificity (95% CI) | Pooled sensitivity (95% CI)  | Pooled specificity (95% CI) | Pooled sensitivity P value <sup>1</sup> | Pooled specificity P value <sup>1</sup> |
|--|-----------------------------|--|-----------------------------|---|---|
| <b>Ofloxacin, indirect testing</b><br>(13 studies, 1927 participants)              |                             | <b>Ofloxacin, direct testing</b><br>(7 studies, 1667 participants)   |                             | 0.180                                   | 0.161                                   |
| 85.2% (78.5 to 90.1)   | 98.5% (95.6 to 99.5)        | 90.9% (84.7 to 94.7)   | 98.9% (97.8 to 99.4)        |   |   |
| <b>Moxifloxacin, indirect testing</b><br>(6 studies, 419 participants)             |                             | <b>Moxifloxacin, direct testing</b><br>(2 studies, 821 participants) |                             | 0.820                                   | 0.365                                   |
| 94.0% (82.2 to 98.1)   | 96.6% (85.2 to 99.3)        | 95.0% (92.1 to 96.9)   | 99.0% (97.5 to 99.6)        |   |   |
| <b>Levofloxacin, indirect testing<sup>2</sup></b><br>(2 studies, 169 participants) |                             | <b>Levofloxacin, direct testing</b><br>(0 studies, 0 participants)   |                             | Not applicable                          | Not applicable                          |
| Not applicable   | Not applicable              | Not applicable   | Not applicable              |   |   |
| <b>Amikacin, indirect testing</b><br>(11 studies, 1301 participants)               |                             | <b>Amikacin, direct testing</b><br>(6 studies, 1491 participants)    |                             | 0.338                                   | 0.213                                   |
| 84.9% (79.2 to 89.1)   | 99.1% (97.6 to 99.6)        | 91.9% (71.5 to 98.1)   | 99.9% (95.2 to 100.0)       |   |   |
| <b>Kanamycin, indirect testing</b><br>(9 studies, 1342 participants)               |                             | <b>Kanamycin, direct testing</b><br>(5 studies, 1020 participants)   |                             | 0.836                                   | 0.445                                   |
| 66.9% (44.1 to 83.8)   | 98.6% (96.1 to 99.5)        | 78.7% (11.9 to 99.0)   | 99.7% (93.8 to 100.0)       |   |   |
| <b>Capreomycin, indirect testing</b><br>(10 studies, 1406 participants)            |                             | <b>Capreomycin, direct testing</b><br>(5 studies, 1027 participants) |                             | 0.841                                   | 0.353                                   |
| 79.5% (58.3 to 91.4)   | 95.8% (93.4 to 97.3)        | 76.6% (61.1 to 87.3)   | 98.2% (92.5 to 99.6)        |   |   |

Abbreviations: CI: confidence interval; DST: drug susceptibility testing; FQ: fluoroquinolones; SLID: second-line injectable drug. We derived the accuracy estimates from non-comparative studies of test accuracy in which different sets of studies were used. For example, for ofloxacin resistance, the set of studies used for indirect testing differed from that used for direct testing.

<sup>1</sup>Likelihood ratio test for evidence of a significant difference between accuracy estimates.

<sup>2</sup>Sensitivity and specificity (95% confidence intervals (CIs)) were 80% (56 to 94) and 96% (80 to 100) for [Chikamatsu 2012](#) and 100% (96 to 100) and 100% (88 to 100) for [Kambli 2015b](#). We did not perform a meta-analysis.

**Table 6. Accuracy of MTBDRs<sub>1</sub> version 1.0 by indirect testing for FQ and SLID resistance and XDR-TB, by reference standard**

| Pooled sensitivity (95% CI)                             | Pooled specificity (95% CI) | Pooled sensitivity (95% CI)  | Pooled specificity (95% CI) | Pooled sensitivity P value <sup>1</sup> | Pooled specificity P value <sup>1</sup> |
|---|-----------------------------|--|-----------------------------|---|---|
| <b>FQ, culture</b><br>(6 studies, 873 participants)     |                             | <b>FQ, sequencing</b><br>(6 studies, 873 participants)                 |                             | < 0.001                                 | 0.735                                   |
| 82.4% (77.6 to 86.3)                                    | 98.8% (94.3 to 99.8)        | 99.3% (81.2 to 100.0)  | 99.3% (90.8 to 100.0)       |   |   |
| <b>FQ, culture</b><br>(7 studies, 1211 participants)    |                             | <b>FQ, sequencing and culture</b><br>(7 studies, 1211 participants)    |                             | 0.664                                   | 0.070                                   |
| 81.8% (77.2 to 85.7)                                    | 99.0% (95.0 to 99.8)        | 82.0% (77.7 to 85.6)   | 99.8% (98.5 to 100.0)       |   |   |
| <b>SLIDs, culture</b><br>(6 studies, 873 participants)  |                             | <b>SLIDs sequencing</b><br>(6 studies, 873 participants)               |                             | 0.034                                   | 0.456                                   |
| 74.6% (66.2 to 81.5)                                    | 99.9% (71.8 to 100.0)       | 97.0% (77.0 to 99.7)   | 99.5% (94.5 to 100.0)       |   |   |
| <b>SLIDs, culture</b><br>(6 studies, 1159 participants) |                             | <b>SLIDs, sequencing and culture</b><br>(6 studies, 1159 participants) |                             | 0.458                                   | 0.203                                   |
| 70.5% (52.0 to 84.1)                                    | 99.8% (93.8 to 100.0)       | 61.3% (45.8 to 74.8%)  | 99.9% (99.0 to 100.0)       |   |   |
| <b>XDR-TB, culture</b><br>(8 studies, 880 participants) |                             | <b>XDR-TB, sequencing<sup>2</sup></b><br>(4 studies, 630 participants) |                             | Could not determine                     | Could not determine                     |
| 70.9% (42.9 to 88.8)                                    | 98.8% (96.1 to 99.6)        | 100% (94.6 to 100.0)   | 97.9% (96.3 to 98.8)        |   |   |

Abbreviations: CI: confidence interval; FQ: fluoroquinolones; SLID: second-line injectable drug; TB: tuberculosis; XDR-TB: extensively drug-resistant TB.

For detection of FQ and SLID resistance, the accuracy estimates were derived from comparative studies of test accuracy in which the same set of studies was used. For example, for FQ resistance, the set of studies using culture-based drug susceptibility testing (DST) as a reference standard was the same as that using sequencing as a reference standard. For detection of XDR-TB, the accuracy estimates were derived from non-comparative studies of test accuracy in which different sets of studies were used.

<sup>1</sup>Likelihood ratio test for evidence of a significant difference between accuracy estimates.

<sup>2</sup>Accuracy estimates were obtained with fixed-effect model.

**Table 7. Sensitivity analyses MTBDRs/ version 1.0, fluoroquinolone resistance**

| Culture, indirect testing  |                             |                             | Culture, direct testing          |                             |                             | Pooled sensitivity<br>P value <sup>1</sup> | Pooled specificity<br>P value <sup>1</sup> |
|--|-----------------------------|-----------------------------|----------------------------------|-----------------------------|-----------------------------|--|--|
| Number of studies (participants)   | Pooled sensitivity (95% CI) | Pooled specificity (95% CI) | Number of studies (participants) | Pooled sensitivity (95% CI) | Pooled specificity (95% CI) |  |  |
| <b>All studies of fluoroquinolones</b>   |                             |                             |                                  |                             |                             |  |  |
| 19 studies (2223)  | 85.6% (79.2 to 90.4)        | 98.5% (95.7 to 99.5)        | 9 studies (1771)                 | 86.2% (74.6 to 93.0)        | 98.6% (96.9 to 99.4)        | 0.932                                      | 0.333                                      |
| <b>Was a consecutive or random sample of patients/specimens enrolled? Yes</b>                                  |                             |                             |                                  |                             |                             |  |  |
| 14 studies (1979)  | 84.1% (75.7 to 90.0)        | 99.0% (94.8 to 99.8)        | 9 studies (1771)                 | 86.2% (75.2 to 92.8)        | 98.9% (97.7 to 99.5)        | 0.725                                      | 0.506                                      |
| <b>Was a case-control design avoided? Yes</b>  |                             |                             |                                  |                             |                             |  |  |
| 13 studies (1389)  | 88.9 (79.4 to 94.3)         | 98.5% (93.4 to 99.7)        | 8 studies (1721)                 | 85.6% (72.4 to 93.1)        | 98.5% (96.9 to 99.3)        | 0.613                                      | 0.417                                      |
| <b>Were the index test results interpreted without knowledge of the results of the reference standard? Yes</b> |                             |                             |                                  |                             |                             |  |  |
| 12 studies (1796)  | 86.1% (75.9 to 92.5)        | 99.3% (94.4 to 99.9)        | 7 studies (982)                  | 83.8% (68.5 to 92.5)        | 98.1% (96.4 to 99.0)        | 0.768                                      | 0.946                                      |
| <b>Was the test applied in the manner recommended by the manufacturer? Yes</b>                                 |                             |                             |                                  |                             |                             |  |  |
| 18 studies (2171)  | 85.6% (78.6 to 90.6)        | 98.6% (95.6 to 99.6)        | 7 studies (982)                  | 83.8% (68.5 to 92.5)        | 98.1% (96.4 to 99.0)        | 0.736                                      | 0.652                                      |

Abbreviations: CI: confidence interval.

We derived the accuracy estimates from non-comparative studies of test accuracy in which different sets of studies were used.

<sup>1</sup>Likelihood ratio test for evidence of a significant difference between accuracy estimates.

**Table 8. Sensitivity analyses MTBDRs/ version 1.0, second-line injectable drug resistance**

| Culture, indirect testing                          |                             |                             | Culture, direct testing          |                             |                             | Pooled sensitivity<br>P value <sup>1</sup> | Pooled specificity<br>P value <sup>1</sup> |
|--|-----------------------------|-----------------------------|----------------------------------|-----------------------------|-----------------------------|--|--|
| Number of studies (participants)                   | Pooled sensitivity (95% CI) | Pooled specificity (95% CI) | Number of studies (participants) | Pooled sensitivity (95% CI) | Pooled specificity (95% CI) |  |  |
| <b>All studies of second-line injectable drugs</b> |                             |                             |                                  |                             |                             |  |  |

**Table 8. Sensitivity analyses MTBDRs/ version 1.0, second-line injectable drug resistance** (Continued)

|  |                         |                         |                     |                         |                          |       |       |
|--|-------------------------|-------------------------|---------------------|-------------------------|--------------------------|-------|-------|
| 16 studies<br>(1921)   | 76.5% (63.3<br>to 86.0) | 99.1% (97.3<br>to 99.7) | 8 studies<br>(1639) | 87.0% (38.1<br>to 98.6) | 99.5% (93.6<br>to 100.0) | 0.547 | 0.664 |
| <b>Was a consecutive or random sample of patients/specimens enrolled? Yes</b>                                  |                         |                         |                     |                         |                          |       |       |
| 11 studies<br>(1869)   | 77.3% (58.9<br>to 89.0) | 99.2% (96.4<br>to 99.8) | 8 studies<br>(1639) | 87.0% (38.1<br>to 98.6) | 99.5% (93.6<br>to 100.0) | 0.896 | 0.873 |
| <b>Was a case-control design avoided? Yes</b>  |                         |                         |                     |                         |                          |       |       |
| 10 studies<br>(1088)   | 80.2% (57.1<br>to 92.4) | 98.6% (95.3<br>to 99.6) | 8 studies<br>(1639) | 87.0% (38.1<br>to 98.6) | 99.5% (93.6<br>to 100.0) | 0.822 | 0.889 |
| <b>Were the index test results interpreted without knowledge of the results of the reference standard? Yes</b> |                         |                         |                     |                         |                          |       |       |
| 10 studies<br>(1513)   | 75.4% (57.0<br>to 87.7) | 99.0% (96.0<br>to 99.7) | 6 studies (902)     | 96.1% (40.0<br>to 99.9) | 99.2% (82.3<br>to 100.0) | 0.471 | 0.573 |
| <b>Was the test applied in the manner recommended by the manufacturer? Yes</b>                                 |                         |                         |                     |                         |                          |       |       |
| 15 studies<br>(1869)   | 77.2% (62.6<br>to 87.2) | 99.0% (97.1<br>to 99.7) | 6 studies (902)     | 96.1% (40.0<br>to 99.9) | 99.2% (82.3<br>to 100.0) | 0.228 | 0.926 |

Abbreviations: CI: confidence interval.

We derived the accuracy estimates from non-comparative studies of test accuracy in which different sets of studies were used.

<sup>1</sup>Likelihood ratio test for evidence of a significant difference between accuracy estimates.

## APPENDICES

### Appendix I. Glossary of terms

#### *Amplification*

Amplification is replication of a DNA fragment to generate copies. Both the original and the newly synthesized copies can be described as the amplicons.

#### *Codon*

A codon is a sequence of three DNA or RNA bases that corresponds to a specific amino acid or a signal to start or stop transcription or translation. The DNA in coding regions of the genome is read in groups of three bases (A, G, C, T).

#### *Conjugate and amplification bands (controls)*

The conjugate band is a control to make sure that the DNA probes immobilized on the MTBDRs/ strip test can bind *M. tuberculosis* DNA and that this is detectable, to ensure that the test is working properly. If this is not present, we will be unsure if the results of the test are due to something going wrong with the test that prevents binding of the bands, or a real phenomenon. Hence, an MTBDRs/ test is indeterminate if the conjugate band is missing. Similarly, the amplification band is a control to make sure that the amplification of *M. tuberculosis* DNA (which is done in order to bring target DNA sequences to a detectable level) was successful. If this band is

absent, and we do not see any probes corresponding to mutations appearing, we cannot discount a failure to amplify DNA rather than the presence of a susceptible strain.

#### *Culture isolates*

Culture isolates refers to *M. tuberculosis* cells from a clinical specimen that have been grown. For TB diagnosis, a volume of the clinical specimen is processed and incubated under conditions that promote *M. tuberculosis* growth. The cells that are grown are referred to a culture isolate.

#### *Drug susceptibility testing*

Drug susceptibility tests determine whether *M. tuberculosis* cells are sensitive or resistant to antibiotics. Testing may be undertaken using phenotypic or genotypic analyses. Phenotypic testing requires growth of TB in the presence of drugs that will inhibit the growth of a sensitive organism or have no impact on growth of a resistant organism. Genotypic testing involves detecting predetermined mutations in DNA that are known to make the organism resistant to a drug. When mutations causing drug resistance are not known, genotypic DST is not useful. Genotype<sup>®</sup> MTBDR<sub>sl</sub> is a genotypic test.

#### *DNA sequencing*

DNA sequencing is a process to determine the nucleotide (A, G, C, T) sequence of fragments of DNA. By comparison of DNA sequences from distinct TB isolates, variations known as mutations can be identified. Some mutations in *M. tuberculosis* are known to cause drug resistance.

#### *Hybridization*

Hybridization is the process of allowing two nucleotide fragments to bind together if they share DNA sequences that are complementary. Hybridisation reactions can be designed such that when one strand binds to another, the reaction may produce a detectable colour change, which is what occurs in MTBDR<sub>sl</sub>.

#### *Locus*

A locus is the position of a genetic feature in the DNA sequence, like a genetic street address. Loci are standardized between genomes by reference to a common reference genome, such as H37Rv for *M. tuberculosis*.

#### *Promoter region*

A promoter region is a sequence of DNA where the transcriptional machinery binds before transcribing the DNA into RNA that may then be translated into an amino acid sequence.

#### *Resistance-determining region*

A region of the *M. tuberculosis* genome where mutations commonly cause resistance to a specific drug. We have adapted these definitions from [NIH 2015](#).

## **Appendix 2. Detailed search strategy**

### **MEDLINE (PubMed)**

1. MTBDR\*.ti/ab.
2. Genotype MTBDR\*.ti/ab
3. or/1-2
4. exp Tuberculosis, Pulmonary/
5. exp Tuberculosis, Multidrug-Resistant/
6. MDR-TB.ti/ab
7. XDR-TB.ti/ab
8. Mycobacterium tuberculosis/
9. TB.ti/ab
10. tuberculosis.ti/ab
11. or/4-10
12. 3 and 11

### **Embase (OVID)**

1. tuberculosis.mp. or lung tuberculosis/ or Mycobacterium tuberculosis/ or multidrug resistant tuberculosis/
2. (MDR-TB or XDR-TB).mp.

3. exp Mycobacterium tuberculosis/
4. 1 or 2 or 3
5. (MTBDR\* or "Genotype MTBDR\*").mp
6. 4 and 5

### **Web of Knowledge (SCI-expanded, Conference Proceedings science) and BIOSIS previews**

Topic=(MTBDR\*) AND Topic=(tuberculosis OR TB OR MDR-TB OR XDR-TB)

### **LILACS**

(tuberculosis OR TB OR mycobacterium OR MDR-TB OR XDR-TB) (Words) AND (MTBDR\$) (Wor

### **SCOPUS**

(tuberculosis OR TB OR mycobacterium OR MDR-TB OR XDR-TB ) (title, abstract, keywords) AND (MTBDR\*) (title, abstract, keywords)

### **CIDG Specialized Register**

(tuberculosis OR TB OR mycobacterium OR MDR-TB OR XDR-TB) AND (MTBDR\*)

### **ProQuest Dissertations & Theses A&I search strategy**

ab(tuberculosis) AND ab((diagnostic test\* OR RDT\* OR MTBDR\*))

### **metaRegister of Controlled Trials (mRCT)**

(tuberculosis OR TB OR mycobacterium OR MDR-TB OR XDR-TB) AND (MTBDR\*)

### **World Health Organization International Clinical Trials Registry Platform**

(tuberculosis OR TB OR mycobacterium OR MDR-TB OR XDR-TB) AND (MTBDR\*)

## **Appendix 3. QUADAS-2 rules and interpretation**

### **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 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 was not clearly reported.

#### **Signaling question 2: was a case-control design avoided?**

We scored 'yes' if the study enrolled only TB patients with suspected resistance to second-line drugs, including patients with confirmed multidrug-resistant tuberculosis (MDR-TB); 'no' if the study enrolled TB patients with confirmed resistance to second-line drugs; and 'unclear' for all other scenarios or if it was not clearly reported.

**Signaling question 3: did the study avoid inappropriate exclusions?**

An inappropriate exclusion might occur if, after the laboratory technician runs the index and reference tests, he or she does not record the test results in the study. This might occur if there were resource constraints as one might find in practice, but we did not expect this to occur in the research studies included in this review. We scored 'yes' for all studies.

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

We judged 'low' concern if the selected specimens matched the review question, which reflects the way the test will be used in practice. We judged 'high' concern if the selected specimens or isolates did not represent those for whom the test will be used in practice, such as in individuals who are not suspected of having drug-resistant TB. We judged 'unclear' concern if 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 this question 'yes' if the reader of the assay was blinded to results of reference tests. We scored 'no' if the reader of the assay was not blinded to the results of reference tests. If the specimens were from a biobank (repository that stores biological specimens) comprised of specimens with known second-line drug resistance and the identity of these specimens was known to the assay reader, we also answered 'no'. We scored 'unclear' if it was not stated in the paper or if the study authors failed to answer this question.

**Signaling question 2: if a threshold was used, was it prespecified?**

A threshold is prespecified in all versions of MTBDR<sub>sl</sub>. 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 the study to be of 'low concern' for applicability if the test was performed as recommended by the manufacturer. We judged the study to be of 'low concern' for applicability if the test was performed as recommended by the manufacturer. We judged the study to be of 'high concern' when the test was applied differently than recommended by the manufacturer, for example when the test was applied to pooled sputa, and we judged the study to be of 'unclear concern' when we could not tell. When available, we selected the high level WHO-recommended concentration for moxifloxacin because we felt this concentration was clinically meaningful.

**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?**

Culture-based DST is not 100% accurate for detection of drug resistance, especially resistance to second-line drugs. However, it is the test currently endorsed by WHO when performed using WHO-recommended critical drug concentrations (defined below [WHO 2012](#)). Therefore, for culture-based DST, we answered 'yes' if study authors used WHO critical concentrations for every test evaluation, 'no' if study authors did not use critical concentrations (or if for the culture method used, WHO-recommended critical concentrations did not exist) in any of their test evaluations, and 'unclear' if study authors used critical concentrations in only some evaluations and not others or the authors did not specify the critical concentrations used. We selected the WHO-recommended high level concentration for moxifloxacin because we felt this concentration was clinically meaningful.



We used the currently-recommended WHO critical concentrations as a benchmark for judging risk of bias (WHO 2012). For *M. tuberculosis*, the antimicrobial susceptibility testing 'critical concentration' is defined as "for each drug, the lowest concentration that inhibits 95% (90% for pyrazinamide) of wild-type strains of *M. tuberculosis* that have not been exposed to the drug, but that simultaneously does not inhibit strains of *M. tuberculosis* that are considered resistant that are isolated from patients who are not responding to therapy" (CSLI 2011). However, there is a lack of consensus about this definition (Ångeby 2012).

Genetic sequencing (gene sequencing of loci known to be associated with drug resistance) is considered by researchers in this field to be the best reference standard for testing for the presence of drug resistance. Although sequencing may not be performed for all regions of the TB genome associated with resistance, we consider this to be a concern about the setting in which the test is applied, rather than a concern about risk of bias.

**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 an automated result (for example, MGIT 960 DST), blinding was explicitly stated, or it was clear that the reference test 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 MTBDR<sub>sl</sub> assay result. We scored 'unclear' if it was not stated in the paper or if the authors failed to answer this question.

**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 because the method used (phenotypic testing with and without drugs) was appropriate.

**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 expected the reference standard test to be undertaken at the same time as the index test (i.e. each performed on a paired sample for most studies). However, we expected some studies to include specimens from patients who had received a reference test on an earlier sample. The sample applies to some culture isolates, whose drug susceptibility profile might have been confirmed prior to the index test being available. We answered 'yes' if the tests were paired or were separated by a few days. We answered 'no' if reference and index tests were not done on paired samples and were separated by several months. As patients suspected of second-line drug resistance are often on some form of anti-TB therapy, it is possible that variation in the microbial population of specimens collected at different time points may occur. We scored 'unclear' if it was not stated in the paper or if the authors failed to answer this question.

**Signaling question 2: did all patients receive the same reference standard?**

Our reference standard for the primary objectives was culture-based DST and we anticipated this reference standard to be used in all studies. We answered 'yes' if culture-based DST was applied to all patients or a random sample of patients, 'no' if the reference standard was only applied to a selective group of patients, and 'unclear' if it was not stated in the paper or if the authors failed to answer this question.

**Signaling question 3: were all patients included in the analysis?**

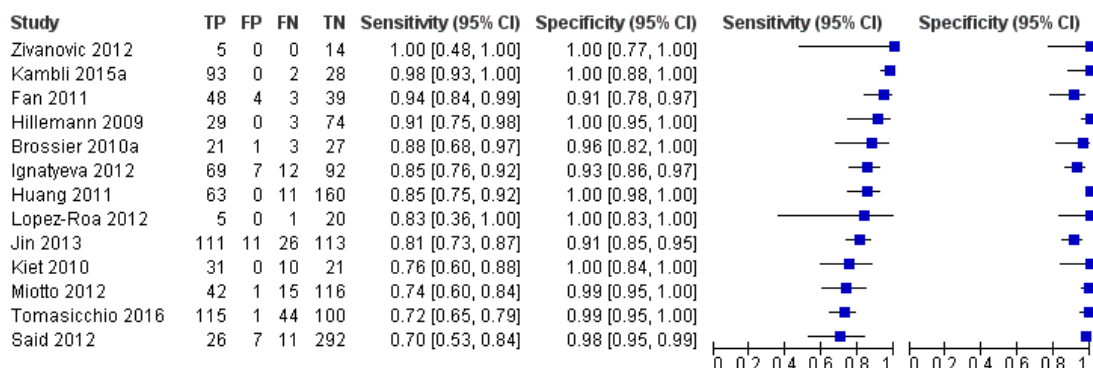
We determined the answer to this question by comparing the number of participants enrolled with the number of patients included in the 2 x 2 tables. We noted if the study authors reported the number of indeterminate assay results. We scored 'yes' if the number of participants enrolled was clearly stated 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. We scored 'unclear' if not enough information was given to assess whether participants were excluded from the analysis.

## Appendix 4. Fluoroquinolone resistance, individual drugs, indirect testing

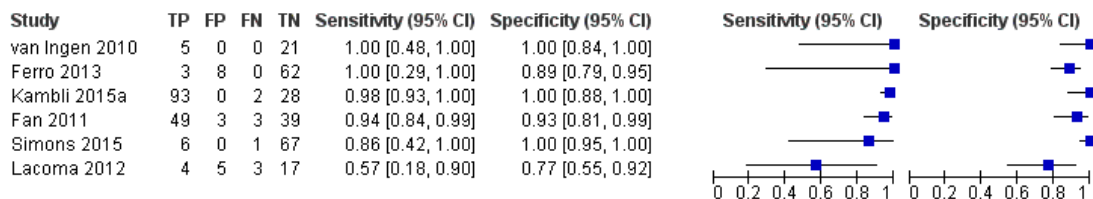
Figure 16 shows forest plots of MTBDR<sub>s</sub>/ sensitivity and specificity for ofloxacin, moxifloxacin, and levofloxacin resistance detection when performed indirectly and using culture-based DST as a reference standard.

**Figure 16. Forest plots of MTBDR<sub>s</sub> sensitivity and specificity for ofloxacin, moxifloxacin, and levofloxacin resistance, the test performed indirectly against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

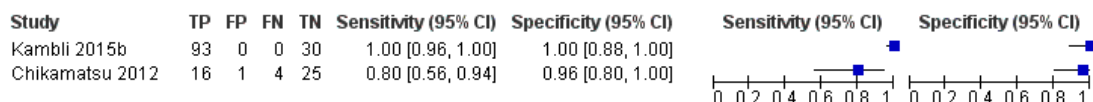
### Indirect, ofloxacin, culture



### Indirect, moxifloxacin, culture



### Indirect, levofloxacin, culture



## Appendix 5. Drug concentrations used in culture-based drug susceptibility testing

**Table 1. Ofloxacin, levofloxacin, and moxifloxacin drug concentrations used in culture-based drug susceptibility testing in relation to the WHO-recommended critical concentrations**

| Study           | Reference standard                 | Concentration used ( $\mu\text{g/mL}$ ) <sup>1</sup> | Used WHO-recommended critical concentration                               |
|-----------------|------------------------------------|--|---|
| Ajvani 2012     | MGIT 960                           | Ofloxacin: 2.0                                       | Yes   |
|                 |                                    | Moxifloxacin: 0.25                                   | No  |
| Barnard 2012    | Middlebrook 7H11 (agar proportion) | Ofloxacin: 2.0                                       | Yes   |
| Brossier 2010a  | LJ (agar proportion)               | Ofloxacin: 2.0                                       | Yes   |
| Catanzaro 2015  | MGIT 960                           | Ofloxacin: 2.0                                       | Yes   |
|                 |                                    | Moxifloxacin: 0.25                                   | No  |
| Chikamatsu 2012 | Ogawa                              | Levofloxacin: 1.0                                    | For Ogawa, there are no WHO-recommended concentrations                    |
| Fan 2011        | MGIT 960                           | Ofloxacin: 2.0                                       | Yes   |
|                 |                                    | Moxifloxacin: 0.25                                   | No  |
| Ferro 2013      | Middlebrook 7H10 (agar proportion) | Moxifloxacin: 2.0                                    | Yes   |
| Hillemann 2009  | MGIT 960 and LJ                    | Ofloxacin: 2.0 for both media                        | Yes for MGIT 960; no for LJ   |
| Huang 2011      | MGIT 960 and Middlebrook 7H11      | Ofloxacin: 2.0                                       | Yes for both media  |
| Ignatyeva 2012  | MGIT 960                           | Ofloxacin: 2.0                                       | Yes   |
| Jin 2013        | LJ and BacT/ALERT 3D               | Ofloxacin: 5.0 (LJ); 50 (BacT/ALERT 3D)              | No for LJ; for BacT/ALERT 3D, there are no WHO-recommended concentrations |
| Kambli 2015b    | MGIT 960                           | Levofloxacin: 1.5                                    | Yes   |
| Kiet 2010       | LJ                                 | Ofloxacin: 2.0                                       | No  |
| Kontsevaya 2013 | MGIT 960                           | Ofloxacin: 2.0                                       | Yes   |
|                 |                                    | Moxifloxacin: 0.25                                   | No  |

(Continued)

|                  |   |  |   |
|------------------|---|--|---|
| Lacoma 2012      | BACTEC 460TB                                  | Moxifloxacin: 0.5                                | For BACTEC460, there are no WHO-recommended concentrations                                  |
| Lopez-Roa 2012   | MGIT 960                                      | Ofloxacin: 2.0                                   | Yes   |
| Miotto 2012      | MGIT 960                                      | Ofloxacin: 2.0                                   | Yes   |
| Said 2012        | Middlebrook 7H11                              | Ofloxacin: 2.0                                   | Yes   |
| Simons 2015      | MGIT 960 and Middlebrook 7H10 (agar dilution) | Moxifloxacin: 0.5 (MGIT); 1.0 (Middlebrook 7H10) | For moxifloxacin using MGIT 960, the study used the WHO-recommended low-level concentration |
| Tagliani 2015    | MGIT 960 and LJ (agar proportion)             | Ofloxacin: 2.0 (LJ: 4.0)                         | Yes   |
|                  |   | Moxifloxacin: 0.5                                | No  |
|                  |   | Levofloxacin: 1.5                                | Yes   |
| Tomasicchio 2016 | MGIT 960                                      | Ofloxacin: 2.0                                   | Yes   |
| Tukvadze 2014    | LJ (proportion method)                        | Ofloxacin: 2.0                                   | No  |
| van Ingen 2010   | Middlebrook 7H10 (agar proportion)            | Moxifloxacin: 1.0                                | No  |
| Zivanovic 2012   | MGIT 960 and LJ agar proportion               | Ofloxacin: 2.0 for both media                    | Yes for MGIT 960; no for LJ   |

Abbreviations: LJ: Löwenstein-Jensen; MGIT: Mycobacteria Growth Indicator Tube; WHO: World Health Organization.

<sup>1</sup>We used the high level concentration when available for all drugs. For a discussion of critical concentrations in LJ media, see [Rigouts 2016](#).

**Table 2. Amikacin, kanamycin, and capreomycin drug concentrations used in culture-based drug susceptibility testing in relation to the WHO-recommended critical concentrations**

| Study       | Reference standard | Concentration used (µg/mL) | Used WHO-recommended critical concentration |
|-------------|--------------------|----------------------------|---|
| Ajvani 2012 | MGIT 960           | Amikacin: 1.0              | Yes   |
|             |                    | Kanamycin: 2.5             | Yes   |
|             |                    | Capreomycin: 2.5           | Yes   |

(Continued)

|                 |                                    |   |   |
|-----------------|------------------------------------|---|---|
| Barnard 2012    | Middlebrook 7H11 (agar proportion) | Amikacin: 4.0                                   | There are no WHO-recommended concentrations for amikacin using Middlebrook 7H11       |
| Brossier 2010a  | LJ (agar proportion)               | Amikacin: 20.0                                  | No  |
|                 |                                    | Kanamycin: 20.0                                 | No  |
|                 |                                    | Capreomycin: 20.0                               | No  |
| Catanzaro 2015  | MGIT 960                           | Amikacin: 1.0                                   | Yes   |
|                 |                                    | Kanamycin: 2.5                                  | Yes   |
|                 |                                    | Capreomycin: 2.5                                | Yes   |
| Chikamatsu 2012 | Ogawa                              | Amikacin: unknown                               | For Ogawa, there are no WHO-recommended concentrations                                |
|                 |                                    | Kanamycin: unknown                              |   |
|                 |                                    | Capreomycin: unknown                            |   |
| Fan 2011        | MGIT 960                           | Amikacin: 1.0                                   | Yes   |
| Ferro 2013      | Middlebrook 7H10 (agar proportion) | Amikacin: 5.0                                   | No  |
|                 |                                    | Kanamycin: 5.0                                  | Yes   |
| Hillemann 2009  | MGIT 960 and LJ (agar proportion)  | Amikacin: 1.0 for MGIT 960 and 40.0 for LJ      | Yes for MGIT 960; no for LJ   |
|                 |                                    | Capreomycin: 2.5 for MGIT 960 and 40.0 for LJ   | Yes for both types of media   |
| Huang 2011      | Middlebrook 7H11 and MGIT 960      | Amikacin: 1.0 for MGIT 960 and 6.0 for 7H11     | Yes for MGIT 960; for 7H11, there is no WHO-recommended concentration for amikacin    |
|                 |                                    | Kanamycin: 2.5 for MGIT 960 and 6.0 for 7H11    | Yes for MGIT 960 and 7H11   |
|                 |                                    | Capreomycin: 2.5 for MGIT 960 and 10.0 for 7H11 | Yes for MGIT 960; for 7H11, there is no WHO-recommended concentration for capreomycin |
| Ignatyeva 2012  | MGIT 960                           | Amikacin: 1.0                                   | Yes   |
|                 |                                    | Kanamycin: 5.0                                  | No  |

(Continued)

|                 |   |   |  |
|-----------------|---|---|--|
|                 |   | Capreomycin: 2.5  | Yes  |
| Jin 2013        | LJ and BacT/ALERT 3D                          | Kanamycin: 10.0   | No   |
|                 |   | Capreomycin: unknown                                    | For BacT/ALERT 3D, there are no WHO-recommended concentrations |
| Kiet 2010       | LJ  | Kanamycin: 20.0   | No   |
| Kontsevaya 2013 | MGIT 960                                      | Amikacin: 1.0   | Yes  |
|                 |   | Kanamycin: 5.0  | No   |
|                 |   | Capreomycin: 2.5  | Yes  |
| Lacoma 2012     | BACTEC 460TB                                  | Kanamycin: 5.0  | For BACTEC 460, there are no WHO-recommended concentrations    |
|                 |   | Capreomycin: 1.25                                       |  |
| Lopez-Roa 2012  | Middlebrook 7H10 and MGIT 960                 | Amikacin: 4.0 (7H10); 1.0 (MGIT 960)                    | Yes for both media   |
| Miotto 2012     | MGIT 960                                      | Amikacin: 1.0   | Yes  |
|                 |   | Kanamycin: 5.0  | No   |
|                 |   | Capreomycin: 2.5  | Yes  |
| Said 2012       | Middlebrook 7H11                              | Kanamycin: 5.0  | No   |
|                 |   | Capreomycin: 10.0                                       | No   |
| Simons 2015     | MGIT 960 and Middlebrook 7H10 (agar dilution) | Amikacin: 1.0 for MGIT and 5.0 for Middlebrook 7H10     | Yes for MGIT 960; no for 7H10                                  |
|                 |   | Capreomycin: 2.5 for MGIT and 10.0 for Middlebrook 7H10 | Yes for MGIT 960; no for 7H10                                  |
| Tagliani 2015   | MGIT 960 and LJ (agar proportion)             | Amikacin: 1.0 for MGIT 960 and 30.0 for LJ              | Yes for both types of media                                    |
|                 |   | Kanamycin: 2.5 for MGIT 960 and 30.0 for LJ             | Yes for both types of media                                    |
|                 |   | Capreomycin: 2.5 for MGIT 960 and 40.0 for LJ           | Yes for both types of media                                    |
| Tomasichio 2016 | MGIT 960                                      | Amikacin: 1.0   | Yes  |

(Continued)

|                |                                    |   |                             |
|----------------|------------------------------------|---|-----------------------------|
| Tukvadze 2014  | LJ                                 | Kanamycin: 30.0                               | Yes                         |
|                |                                    | Capreomycin: 40.0                             | Yes                         |
| van Ingen 2010 | Middlebrook 7H10 (agar proportion) | Amikacin: 5.0                                 | No                          |
|                |                                    | Capreomycin: 10.0                             | No                          |
| Zivanovic 2012 | MGIT 960 and LJ agar proportion    | Amikacin: 1.0 for MGIT 960 and 40.0 for LJ    | Yes for MGIT 960; no for LJ |
|                |                                    | Capreomycin: 2.5 for MGIT 960 and 40.0 for LJ | Yes for both types of media |

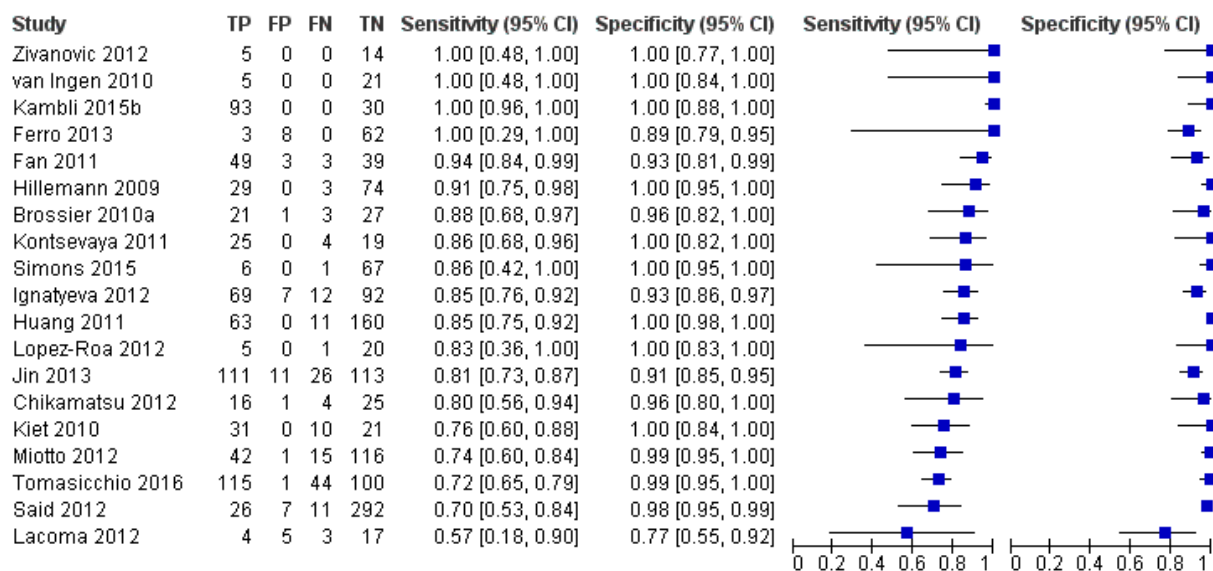
Abbreviations: LJ: Löwenstein-Jensen; MGIT: Mycobacteria Growth Indicator Tube; WHO: World Health Organization.

## Appendix 6. Fluoroquinolone resistance, different reference standards

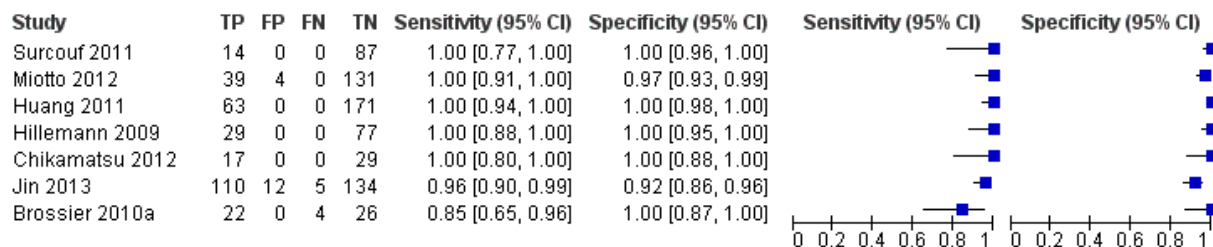
Figure 17 shows forest plots of MTBDRs/ sensitivity and specificity for fluoroquinolone (FQ) resistance detection when performed indirectly using different reference standards.

**Figure 17. Forest plots of MTBDRsl sensitivity and specificity for fluoroquinolone (FQ) resistance, the test performed indirectly against different reference standards. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

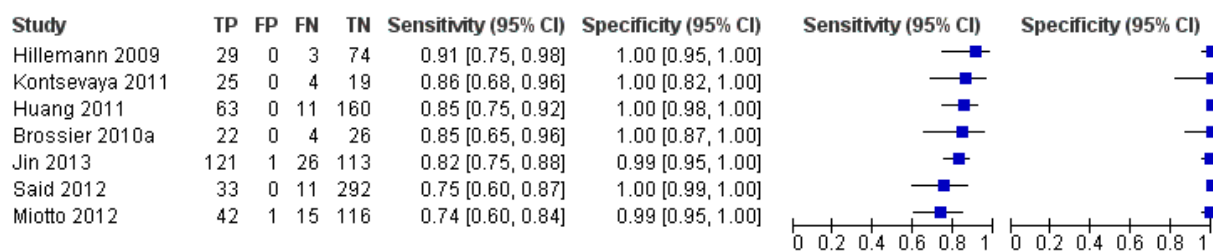
**Indirect, FQ, culture**



**Indirect, FQ, sequencing**



**Indirect, FQ, sequencing and culture**

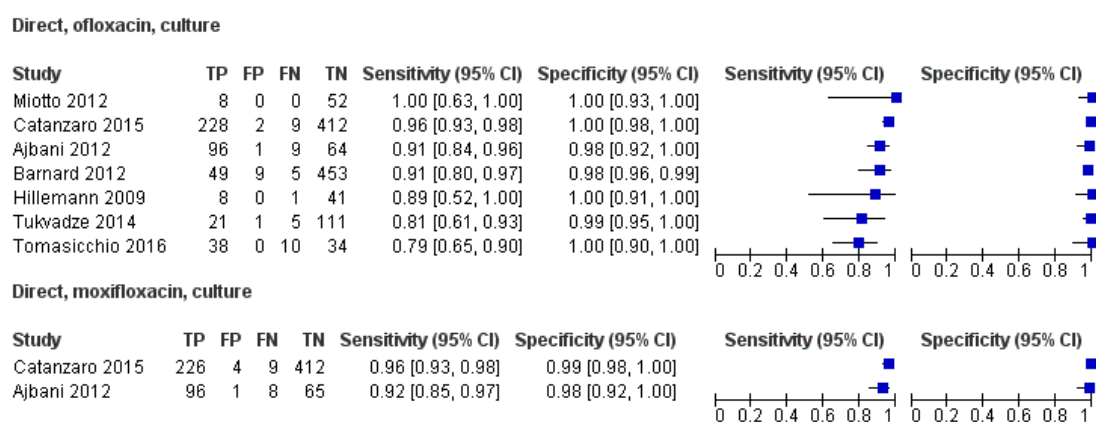




## Appendix 7. Fluoroquinolone resistance, individual drugs, direct testing

Figure 18 shows forest plots of MTBDRsI sensitivity and specificity for ofloxacin and moxifloxacin resistance detection when performed directly and using culture-based drug susceptibility testing (DST) as a reference standard.

**Figure 18. Forest plots of MTBDRsI sensitivity and specificity for ofloxacin, moxifloxacin, and levofloxacin resistance, the test performed directly against culture-based drug susceptibility testing (DST) as a reference standard. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

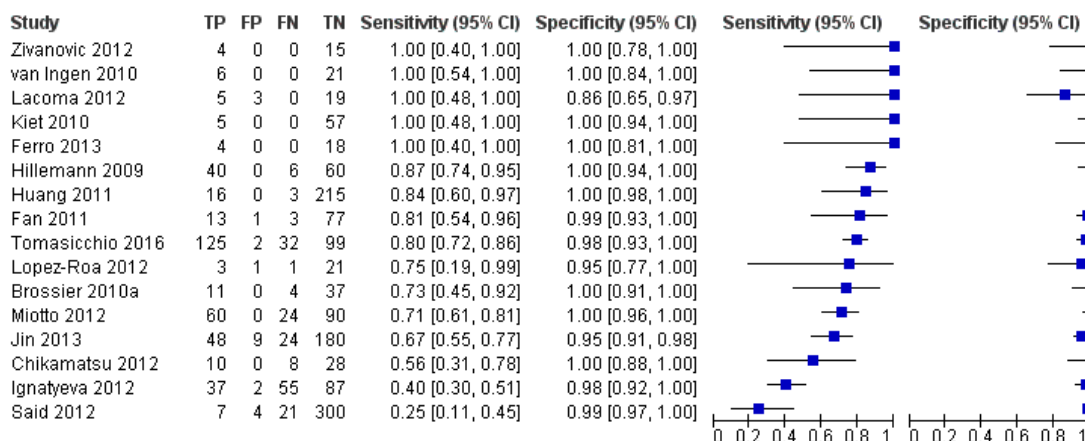


## Appendix 8. Second-line injectable drug resistance, different reference standards

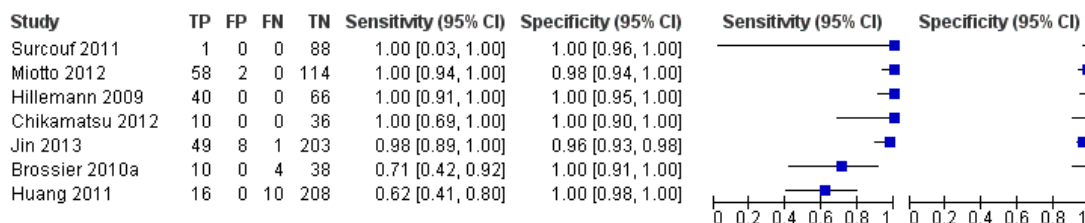
Figure 19 shows forest plots of MTBDRsI sensitivity and specificity when performed indirectly for second-line injectable drug (SLID) resistance detection and using different reference standards.

**Figure 19. Forest plots of MTBDRsl sensitivity and specificity for second-line injectable drug (SLID) resistance, the test performed indirectly against three different reference standards. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line). The individual studies are ordered by decreasing sensitivity.**

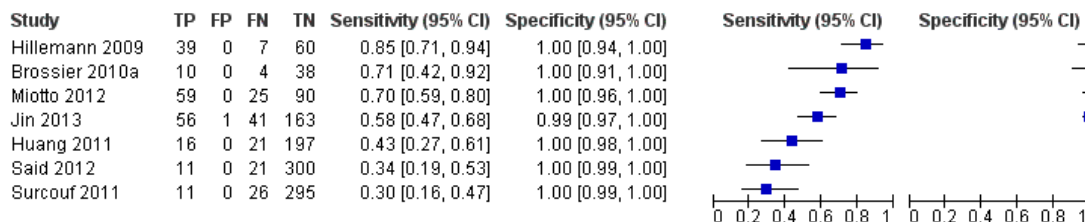
**Indirect, SLID, culture**



**Indirect, SLID, sequencing**



**Indirect, SLID, sequencing and culture**



## WHAT'S NEW

Last assessed as up-to-date: 21 September 2015.

| Date         | Event  | Description  |
|--------------|--|--|
| 25 July 2016 | New search has been performed                          | We updated the search and included six new studies. We updated tuberculosis (TB) surveillance data in the <a href="#">Background</a> , and revised the sections of <a href="#">Index tests</a> , <a href="#">Target conditions</a> , and <a href="#">Statistical analysis and data synthesis</a> . |
| 25 July 2016 | New citation required but conclusions have not changed | Review updated with six new included studies.  |

## CONTRIBUTIONS OF AUTHORS

GT and KRS wrote the first draft of the protocol. KD and MR contributed methodological advice. RW and JP gave advice on protocol content. GT and JP reviewed the studies and extracted the accuracy data. GT, JP, and KRS assessed the methodological quality of the included studies. MR performed the statistical analyses. GT, JP, MR, KD, and KRS interpreted the findings. GT and KRS wrote the first draft of the review and prepared the 'Summary of findings' tables. All review authors contributed to the final manuscript.

## DECLARATIONS OF INTEREST

The review authors have no financial involvement with any organization or entity with a financial interest in, or financial conflict with, the subject matter or materials discussed in the review apart from those disclosed.

## SOURCES OF SUPPORT

### Internal sources

- Liverpool School of Tropical Medicine, UK.

### External sources

- Medical Research Council, South Africa.  
Career Development Award held by GT.
- National Research Foundation, South Africa.  
Core support to KD.
- ACTG/IMPAACT/HVTN International Tuberculosis Speciality Laboratory, South Africa.  
Grant award for support to MB
- Department for International Development (DFID), UK.  
Grant: 5242
- United States Agency for International Development (USAID) administered by the World Health Organization (WHO) Global TB Programme, USA.
- The Bill and Melinda Gates Foundation, USA.

## DIFFERENCES BETWEEN PROTOCOL AND REVIEW

For both the original and updated reviews, we added an additional reference standard defined as two reference tests used together: phenotypic culture-based drug susceptibility testing (DST) and genetic sequencing of the same samples. We added the question, "Was a case-control design avoided?" to the sensitivity analyses. We stated in the protocol, [Theron 2013](#), that we would perform sensitivity analyses for each target condition, using the subset of studies that provided one result per patient. However, these studies did not provide sufficient data for such analyses. We made several revisions in the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. After further consultation with technical experts, we changed how we assessed risk of bias with regard to the reference standard domain of QUADAS-2. We decided to distinguish between studies that used culture-based DST with a WHO-recommended critical concentration in order to define resistance (answered as 'yes' if study consistently used recommended critical concentrations), those studies which did not (answered as 'no'), and those studies which used critical concentrations in some evaluations and not others or the authors did not specify the critical concentrations used (answered as 'unclear'). For concerns of applicability in the reference standard, we answered 'unclear' if sequencing was used as the reference standard and the type of sequencing did not examine the genes known to be associated with resistance (for example, *gyrB* for the FQs and *eis* for the SLIDs). For signalling question 2 in the flow and timing domain, "Did all patients receive the same reference standard?", we answered 'no' when culture followed by sequencing of discrepant results was performed. We also clarified concerns about applicability for the index test. We considered studies to be of 'high concern' when the test was applied differently than recommended by the manufacturer, for example when the test was applied to pooled sputa. We added an investigation of heterogeneity in relation to microscopy smear grade.

## INDEX TERMS

### Medical Subject Headings (MeSH)

Antitubercular Agents [\*therapeutic use]; Cross-Sectional Studies; Extensively Drug-Resistant Tuberculosis [\*diagnosis; drug therapy]; Fluoroquinolones [therapeutic use]; Genotype; Microbial Sensitivity Tests [\*methods; standards]; *Mycobacterium tuberculosis* [\*drug effects]; Sensitivity and Specificity

### MeSH check words

Adult; Humans