USE OF INTERFERON-γ RELEASE ASSAYS (IGRAs) IN TUBERCULOSIS CONTROL
IN LOW- AND MIDDLE-INCOME SETTINGS

EXPERT GROUP MEETING REPORT
20-21 JULY 2010

This report contains the collective views of an international group of experts, and does not necessarily represent the decisions or the stated policy of the World Health Organization. Mention of a technology does not imply endorsement of any specific commercial product.
Executive summary

Background

Research over the past decade has resulted in the development of two commercial interferon-gamma release assays (IGRAs). Both assays work on the principle that the T-cells of an individual who have acquired TB infection will respond to re-stimulation with M. tuberculosis-specific antigens by secreting interferon-gamma. The QuantiFERON-TB Gold (QFT-G, Cellestis, Australia) and the newer generation QuantiFERON-TB Gold In-Tube (QFT-GIT, Cellestis, Australia) are whole-blood based enzyme-linked immunosorbent assays (ELISA) measuring the amount of IFN-\(\gamma\) produced in response to three M. tuberculosis antigens (QFT-G: ESAT-6 and CFP-10; QFT-GIT: ESAT-6, CFP-10 and TB7.7). In contrast, the enzyme-linked immunospot (ELISPOT)-based T-SPOT.TB (Oxford Immunotec) measures the number of peripheral mononuclear cells that produce INF-\(\gamma\) after stimulation with ESAT-6 and CFP-10.

In recent years, IGRAs have become widely endorsed in high-income countries for diagnosis of latent TB infection (LTBI) and several guidelines (albeit equivocal) on their use have been issued. Currently, there are no guidelines for their use in high TB- and HIV-burden settings, typically found in low- and middle-income countries, where IGRA use are being marketed and promoted, especially in the private sector. Systematic reviews have suggested that IGRA performance differs in high- versus low TB and HIV incidence settings, with relatively lower sensitivity in high-burden settings. The majority of IGRA studies have been performed in high-income countries and mere extrapolation to low- and middle-income settings with high background TB infection rates is not appropriate. The WHO Stop TB Department has therefore commissioned systematic reviews on the use of IGRAs in low- and middle-income settings, in pre-defined target groups, with funding support from the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and TREAT-TB/The Union. The target groups and major findings are briefly summarised below.

Summary of results

Use of IGRAs in diagnosis of active TB: IGRAs were explicitly designed to replace the TST in diagnosis of LTBI, and were not intended for diagnosis of active TB. Because IGRAs (like the TST) cannot distinguish LTBI from active TB, these tests are expected to have poor specificity for active TB in high-burden settings due to a high background prevalence of LTBI. Nineteen studies simultaneously estimating sensitivity and specificity among 2,067 TB suspects demonstrated a pooled sensitivity of 83% (95% CI 70% - 91%) and pooled specificity of 58% (95% CI 42% - 73%) for T-SPOT (8 studies), and a pooled sensitivity of 73% (95% CI 61% - 82%) and pooled specificity of 49% (95% CI 40% - 58%) for QFT-GIT (11 studies). There was no consistent evidence that either IGRA was more sensitive than the TST for diagnosis of active TB diagnosis. Two studies evaluated the incremental value of IGRAs and found no meaningful contribution of IGRAs for diagnosis of active TB beyond readily available patient data and conventional microbiological tests.

Expert Group consensus: The quality of evidence for use of IGRAS in diagnosis of active TB was low and it is recommended that these tests should not be used as a replacement for conventional microbiological diagnosis of pulmonary and extra-pulmonary TB in low- and middle-income countries [strong recommendation]. The Expert Group also noted that current evidence did not support the use of IGRAs as part of the diagnostic workup of adults suspected of active TB in low-and middle-income countries, irrespective of HIV status. This recommendation places a high value on avoiding the consequences of unnecessary treatment (high false-positives) given the low specificity of IGRAs in these settings.

Use of IGRAs in children: Only two small studies were identified which prospectively estimated incidence of active TB in children who had been tested with QFT. Conflicting results were reported. When the reference standard for LTBI was exposure, all three tests (TST, QFT and T-SPOT) seemed to be associated with the
level of exposure (categorised either dichotomously or by an exposure gradient); however, methodological inconsistencies between the studies regarding the selection and definition of reference standards for active TB and exposure limited the comparability of studies and results. Estimates of association were very similar, suggesting no difference in performance between TST and IGRAs for diagnosis of LTBI and active TB in children.

**Expert Group consensus:** The quality of evidence for use of IGRAS in children was very low and it is recommended that these tests should not be used as an alternative to TST in paediatric TB in low and middle-income countries for the diagnosis of latent TB infection, or as an alternative to TST in the workup of a diagnosis of active TB disease in children, irrespective of HIV status (strong recommendation). The Expert Group also notes that there may be additional harms associated with blood collection in children and that issues such as acceptability and cost have not been adequately addressed in any studies.

**Use of IGRAs in HIV-infected individuals:** 36 studies were identified that included 5,400 HIV-infected individuals. In persons with active TB (used as a surrogate reference standard for LTBI), pooled sensitivity estimates were higher for TSPOT (72%, 95% CI 62% - 81%, 8 studies) than for QFT-GIT (61%, 95% CI 41% - 75%, 8 studies). Large prospective cohort studies have established that persons with a positive TST have a 1.4 to 1.7-fold higher rate of active TB within one year compared to persons with a negative TST result. Three studies evaluating the predictive value of IGRAs in HIV-infected individuals showed that IGRAs have poor positive predictive value but high negative predictive value for active TB. While these results suggest that a negative IGRA result is reassuring (no person with a negative IGRA result developed culture-positive TB), the studies had serious limitations, including small sample sizes with short-duration of follow-up and differential evaluation and/or follow-up of persons with positive and negative IGRA results. Neither IGRA was consistently more sensitive than TST in head-to-head comparisons, and the impact of advance immunosuppression on IGRA validity remains unclear: Two studies reported TST and IGRA data stratified by CD4 count. In one study, the proportion of positive results among those with CD4 cell count <200 decreased by 27% (95% CI -61, 8) with TSPOT and 35% (95% CI -59, -11) with TST. In the other study, the proportion of positive results among those with CD4 cell count <200 decreased by 31% (95% CI (-53, -9) with TSPOT and increased by 15% (95% CI (-11, 41) with TST. All tests therefore seem to be affected by CD4+ cell count, and additional studies from low/middle income countries are needed.

**Expert Group consensus:** The quality of evidence for use of IGRAS in individuals living with HIV infection was very low and recommended that these tests should not be used as a replacement for TST for the assessment of LTBI (strong recommendation). This recommendation also applies to HIV-positive children based on the generalisation of data from adults.

**Use of IGRAs in health care worker (HCW) screening:** Limited data was available on the utility of screening HCWs for LTBI in high incidence countries. Three cross-sectional studies were evaluated comparing IGRA and TST performance in HCWs in three countries, although TST was only performed in two of these. TST and IGRA positivity rates were high in HCWs, ranging from 40% to 66%. IGRA positivity was slightly lower than TST positivity in the two studies comparing TST and IGRAs; however, the difference in estimated prevalence was significant in one study only. Serial testing data, evidence on the predictive value of IGRAs in HCWs, as well as reproducibility data are still absent for high-incidence settings and limited even in low-incidence settings.

**Expert Group consensus:** The quality of evidence for use of IGRAS for screening of health care workers in low- and middle-income countries was very low and it is recommended that these tests should not be used in health care worker screening programmes (strong recommendation). The Expert Group also noted the lack of WHO policy on using the TST in health care worker screening programmes.

**Use of IGRAs in contact screening and outbreak investigations:** 16 studies (14 original manuscripts and 2 unpublished studies) were identified which evaluated IGRAs in contact screening and outbreak investigations in low- and middle income countries. Seventy-five percent (12/16) of contact studies included children in their study populations. The majority of studies were cross-sectional and looked at concordance between TST and IGRAs. Due to significant heterogeneity in study designs and outcomes
assessed in each study, it was not possible to pool the data. The majority of studies showed comparable LTBI prevalence by TST or IGRA in contacts and only 4 studies reported a statistically significant difference between positivity rates estimated by TST, SPOT.TB or QFT. The most commonly observed discordance was of the TST-positive/IGRA-negative type. Both IGRAs and the TST seemed to show positive associations with higher levels of exposure in cross-sectional studies, but the strength of the association (ie. adjusted odds ratio) varied across studies. Results indicated that concordance between TST and IGRAs ranged widely, with only moderate agreement. In high-income settings, IGRAs appear to be dynamic and are associated with conversions and reversions which has impact for serial testing of contacts; however no data exists for LMICs.

**Expert Group consensus:** The quality of evidence for use of IGRAS for LTBI screening in contact and outbreak investigations was very low and it is recommended that these tests should not be used as a replacement for TST, neither in adults nor children investigated as close contacts of patients with confirmed active TB (strong recommendation).

**Predictive value of IGRAs:** Three studies provided incidence rate ratios (IRR) of TB stratified by IGRA as well as TST status at baseline. The association with subsequent incident TB in test-positive individuals compared to test-negatives appeared higher for IGRA than for TST; however, this was not statistically significant (IGRA: IRR=3.24; 95CI 0.62-5.85; I^2=0%; p=0.90; TST: IRR=2.28; 95CI 0.83-3.73); The Expert Group also noted that both IGRAs and TST seemed to show positive associations between exposure gradient and test results but with variability in the strength of the association across populations irrespective of BCG vaccination. No statistically significant increase in incidence rates of TB in IGRA-positives compared to IGRA-negatives was observed and the vast majority of individuals (>95%) with a positive IGRA result did not progress to active TB disease during follow-up. Both IGRAs and the TST appeared to have only modest predictive value and did not help identify those who are at highest risk of progression to disease. The predictive value for serial testing could not be assessed as all three studies performed single time-point IGRA testing. Patient relevant outcomes based on sensitivity and specificity appeared comparable between IGRAs and the TST.

**Expert Group consensus:** The quality of evidence for the predictive value of IGRAS was very low and it is recommended that these assays should not be used to identify individuals at risk of active TB disease in low- and middle-income countries (strong recommendation).
Contents

1. BACKGROUND .................................................................................................................... 1

2. EVIDENCE BASE ................................................................................................................ 3
  2.1 EVIDENCE SYNTHESIS .................................................................................................. 3
  2.2 SYSTEMATIC REVIEWS AND META-ANALYSES .............................................................. 4
  2.3 GRADE EVALUATION ..................................................................................................... 6
  2.4 MEETING PROCEDURAL ISSUES ................................................................................... 7

3. RESULTS .............................................................................................................................. 8
  3.1 USE OF IGRAS IN DIAGNOSIS OF ACTIVE TB ............................................................... 8
    3.1.1 Objectives, reference standards and outcomes ............................................................ 8
    3.1.2 Search results ........................................................................................................... 9
    3.1.3 Data analysis ............................................................................................................ 9
    3.1.4 Study characteristics ............................................................................................... 10
    3.1.5 Study quality .......................................................................................................... 10
    3.1.6 Sensitivity and specificity estimation among TB suspects ......................................... 10
    3.1.7 Proportion of indeterminate IGRA results ............................................................... 17
    3.1.8 Incremental value of IGRAs for active TB ................................................................. 18
    3.1.9 Summary of findings and GRADE evidence profiles .................................................. 20
    3.1.10 Strengths and limitations of the evidence base ....................................................... 21
    3.1.11 Final recommendations .......................................................................................... 21
  3.2 USE OF IGRAS IN CHILDREN ...................................................................................... 29
    3.2.1 Objectives, reference standards and outcomes ............................................................ 29
    3.2.2 Data analysis ............................................................................................................ 31
    3.2.3 Search results ....................................................................................................... 32
    3.2.4 Study characteristics .............................................................................................. 32
    3.2.5 Study quality .......................................................................................................... 33
    3.2.6 Test failure and indeterminate results across studies and populations .................... 34
    3.2.7 Studies assessing incident TB .................................................................................. 35
    3.2.8 Studies assessing exposure ..................................................................................... 35
    3.2.9 Studies assessing active TB ..................................................................................... 39
    3.2.10 Summary of findings and GRADE evidence profiles .............................................. 47
    3.2.11 Strengths and limitations of the evidence base ....................................................... 47
    3.2.12 Final recommendations .......................................................................................... 48
  3.3 USE OF IGRAS FOR THE DIAGNOSIS OF LTBI IN HIV-INFECTED INDIVIDUALS ....... 60
    3.3.1 Objectives, reference standards and outcomes ............................................................ 60
    3.3.2 Search results ....................................................................................................... 60
    3.3.3 Study characteristics .............................................................................................. 61
    3.3.4 Study quality .......................................................................................................... 61
    3.3.5 Risk of progression to active TB .............................................................................. 61
    3.3.6 Sensitivity in culture-confirmed active TB patients ................................................... 62
    3.3.7 Agreement between IGRA and TST results ............................................................... 63
    3.3.8 Indeterminate IGRA results ..................................................................................... 64
    3.3.9 Impact of immunosuppression ............................................................................... 65
    3.3.10 Strengths and limitations of the evidence base ....................................................... 67
3.3.11 Research gaps .................................................................................................................. 68
3.3.12 Summary of findings and GRADE evidence profiles .......................................................... 68
3.3.13 Final Recommendations ....................................................................................................... 68
3.4 USE OF IGRAS FOR SCREENING OF HEALTH CARE WORKERS ........................................ 73
3.4.1 Objectives, reference standards and outcomes ......................................................................... 73
3.4.2 Search results .......................................................................................................................... 73
3.4.3 Study characteristics ................................................................................................................ 74
3.4.4 Study quality ............................................................................................................................ 74
3.4.5 IGRA vs. TST positivity rates in high-incidence countries ....................................................... 74
3.4.6 IGRA conversion and reversion rates in longitudinal, serial testing ...................................... 75
3.4.7 Association between occupational risk factors and test results in HCWs ............................... 76
3.4.8 Cost-effectiveness of IGRAs in HCW screening ..................................................................... 76
3.4.9 Strengths and limitations of the evidence base ....................................................................... 76
3.4.10 Research gaps ........................................................................................................................ 76
3.4.11 Summary of findings and GRADE evidence profiles ............................................................. 76
3.4.12 Final Recommendations ....................................................................................................... 77
3.5 USE OF IGRAS IN CONTACT SCREENING AND OUTBREAK INVESTIGATIONS .................... 83
3.5.1 Objectives, reference standards and outcomes ......................................................................... 83
3.5.2 Search results .......................................................................................................................... 83
3.5.3 Data analysis ............................................................................................................................ 84
3.5.4 Study characteristics ................................................................................................................ 84
3.5.5 Study quality ............................................................................................................................ 84
3.5.6 Agreement between IGRA and TST Results ........................................................................... 84
3.5.7 Correlation between test positivity and exposure ..................................................................... 85
3.5.8 Concordance between test results in longitudinal contact studies .......................................... 87
3.5.9 Indeterminate IGRA results .................................................................................................... 95
3.5.10 Strengths & limitations of the evidence base ....................................................................... 98
3.5.11 Summary of findings and GRADE evidence profiles ............................................................. 98
3.5.12 Final Recommendations ....................................................................................................... 98
3.6 THE PREDICTIVE VALUE OF IGRAS FOR INCIDENT ACTIVE TB ............................................ 102
3.6.1 Objectives, reference standards and outcomes ......................................................................... 102
3.6.2 Search results .......................................................................................................................... 103
3.6.3 Data analysis ............................................................................................................................ 103
3.6.4 Study characteristics ................................................................................................................ 103
3.6.5 Study quality ............................................................................................................................ 103
3.6.6 Incidence rates of TB during follow-up .................................................................................... 103
3.6.7 The association between IGRA and incident active TB ......................................................... 104
3.6.8 The predictive value of IGRA compared with TST ................................................................. 104
3.6.9 The influence of discordant-/ concordant TST/IGRA pairs at baseline on subsequent TB rates 106
3.6.10 Gradient between rates of TB and quantitative IGRA levels .................................................. 106
3.6.11 Patient relevant outcomes ...................................................................................................... 108
3.6.12 The predictive value of serial testing ..................................................................................... 109
3.6.13 Summary of findings and GRADE evidence profiles ............................................................. 109
3.6.14 Final Recommendations ....................................................................................................... 109
3.7 OPERATIONAL ASPECTS ON THE USE OF IGRAS IN HIGH TB BURDEN COUNTRIES .......... 117
ANNEX 1. MEETING PARTICIPANTS ............................................................................................... 119
ANNEX 2. DECLARATIONS OF INTEREST .................................................................................... 122
ANNEX 3. SELECTION OF STUDIES EVALUATING THE USE OF IGRA IN THE DIAGNOSIS OF ACTIVE TB ................................................................. 123
ANNEX 4. SELECTION OF STUDIES EVALUATING THE USE OF IGRAS IN CHILDREN .................. 135
ANNEX 5. SELECTION OF STUDIES EVALUATING THE USE OF IGRAS FOR THE DIAGNOSIS OF LTBI IN HIV-POSITIVE INDIVIDUALS .......... 141
ANNEX 6. SELECTION OF STUDIES EVALUATING THE USE OF IGRAS FOR TUBERCULOSIS SCREENING OF HEALTH CARE WORKERS .......... 152
ANNEX 7. SELECTION OF STUDIES EVALUATING THE USE OF IGRAS LTBI SCREENING IN CONTACT AND OUTBREAK INVESTIGATIONS .... 157
ANNEX 8. SELECTION OF STUDIES EVALUATING THE PREDICTIVE VALUE OF IGRAS FOR INCIDENT ACTIVE TB DISEASE IN LOW, MIDDLE AND HIGH INCOME COUNTRIES ................................................................. 165
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hierarchy of reference standards used to assess the evidence base.............</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Assessment of study quality using the QUADAS tool................................</td>
<td>10</td>
</tr>
<tr>
<td>3A-3C</td>
<td>Hierarchical summary receiver operating characteristics (HSROC) plot of studies that reported both sensitivity and specificity in active TB suspects</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB in HIV-positive persons with confirmed active tuberculosis in low- and middle-income countries</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB in HIV-negative persons with confirmed active tuberculosis in low- and middle-income countries</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>Percent sensitivity difference between IGRA and TST results........................</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>Proportion of indeterminate IGRA results among HIV-negative subjects in low- and middle-income countries</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>Proportion of indeterminate IGRA results among HIV-positive subjects in low- and middle-income countries</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>Receiver operating characteristic (ROC) curves for baseline clinical prediction model with and without addition of IGRA (unpublished)</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>Receiver operating characteristic (ROC) curves for baseline clinical prediction model with and without addition of QuantiFERON-Gold In-Tube and tuberculin skin test (unpublished)</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>Methodological quality: Exposure studies...............................................</td>
<td>34</td>
</tr>
<tr>
<td>12</td>
<td>Methodological quality: Studies assessing active TB................................</td>
<td>34</td>
</tr>
<tr>
<td>13</td>
<td>Pooled odds ratios (95% CIs): Concordance of test with dichotomous exposure</td>
<td>36</td>
</tr>
<tr>
<td>14</td>
<td>Pooled estimates: Correlation of tests with exposure gradients..................</td>
<td>38</td>
</tr>
<tr>
<td>15</td>
<td>Regression slopes by exposure gradient..................................................</td>
<td>38</td>
</tr>
<tr>
<td>16</td>
<td>Forest plots for TST sensitivity and specificity in active TB, separated by World Bank income index</td>
<td>44</td>
</tr>
<tr>
<td>17</td>
<td>Forest plots for QuantiFERON sensitivity and specificity in active TB, separated by World Bank income index</td>
<td>45</td>
</tr>
<tr>
<td>18</td>
<td>Forest plots for T-SPOT sensitivity and specificity in active TB, separated by World Bank income index</td>
<td>46</td>
</tr>
<tr>
<td>19</td>
<td>Sensitivity of IGRAs in HIV-positive individuals with confirmed active TB......</td>
<td>63</td>
</tr>
<tr>
<td>20</td>
<td>Percent concordance between IGRA and TST results...................................</td>
<td>64</td>
</tr>
<tr>
<td>21</td>
<td>Proportion of indeterminate IGRA results in HIV-infected persons screened for LTBI</td>
<td>65</td>
</tr>
<tr>
<td>22</td>
<td>Impact of immunosuppression on IGRA results..........................................</td>
<td>66</td>
</tr>
<tr>
<td>23</td>
<td>LTBI and TST prevalence in low- and middle-income settings.......................</td>
<td>75</td>
</tr>
<tr>
<td>24</td>
<td>Studies evaluating concordance between TST and IGRAs................................</td>
<td>86</td>
</tr>
<tr>
<td>25</td>
<td>Summary differences in prevalence of positive tests against proportion BCG vaccinated</td>
<td>87</td>
</tr>
<tr>
<td>26</td>
<td>Association between test positivity and TB exposure, as a dichotomous variable</td>
<td>88</td>
</tr>
<tr>
<td>27A</td>
<td>Association between TST positivity and exposure gradient..........................</td>
<td>93</td>
</tr>
<tr>
<td>27B</td>
<td>Association between IGRA positivity and exposure gradient........................</td>
<td>93</td>
</tr>
<tr>
<td>28</td>
<td>Crude TB incidence rates stratified by IGRA status..................................</td>
<td>105</td>
</tr>
<tr>
<td>29</td>
<td>Crude, unadjusted TB incidence rate ratios for IGRA-positives vs. IGRA-negatives</td>
<td>106</td>
</tr>
<tr>
<td>30</td>
<td>Crude, unadjusted incidence rate ratios for IGRAs compared to the TST............</td>
<td>107</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>QUADAS</td>
<td>5</td>
</tr>
<tr>
<td>Table 2</td>
<td>HEAD-TO-HEAD COMPARISON OF SENSITIVITY OF T-SPOT.TB VERSUS QUANTIFERON-TB GOLD IN-TUBE AMONG ACTIVE TB SUSPECTS</td>
<td>14</td>
</tr>
<tr>
<td>Table 3</td>
<td>GRADE EVIDENCE PROFILE: 1 DIAGNOSTIC ACCURACY OF CURRENTLY AVAILABLE COMMERCIAL INTERFERON-GAMMA RELEASE ASSAYS (QUANTIFERON-TB GOLD IN-TUBE [QFT-GIT], CELLESTIS, AUSTRALIA AND T-SPOT.TB [T-SPOT], OXFORD IMMUNOTEC, UNITED KINGDOM) FOR EVALUATION OF PATIENTS WITH PULMONARY TB IN LOW- AND MIDDLE-INCOME COUNTRIES</td>
<td>22</td>
</tr>
<tr>
<td>Table 4</td>
<td>GRADE SUMMARY OF FINDINGS — ROLE OF IGRA FOR EVALUATION OF PATIENTS WITH PULMONARY TB IN LOW- AND MIDDLE-INCOME COUNTRIES</td>
<td>25</td>
</tr>
<tr>
<td>Table 5</td>
<td>STUDIES WITH DICHOTOMOUS EXPRESSION OF EXPOSURE INCLUDED IN ANALYSIS OF TEST PERFORMANCE FOR LTBI, SORTED BY WORLD BANK INCOME INDEX</td>
<td>35</td>
</tr>
<tr>
<td>Table 6</td>
<td>STUDIES WITH EXPOSURE GRADIENTS INCLUDED IN ANALYSIS OF TEST PERFORMANCE FOR LTBI, SORTED BY WORLD BANK INCOME INDEX</td>
<td>37</td>
</tr>
<tr>
<td>Table 7</td>
<td>STUDIES INCLUDED IN ANALYSIS OF TEST PERFORMANCE IN TB DISEASE, SORTED BY WORLD BANK INCOME INDEX</td>
<td>40</td>
</tr>
<tr>
<td>Table 8</td>
<td>DIAGNOSTIC ACCURACY OF TST, QFT AND T-SPOT IN ACTIVE DISEASE IN CHILDREN, INCLUDING ALL ACCEPTABLE DEFINITIONS OF TB AND ALL COUNTRIES (HIC AND LMIC)</td>
<td>41</td>
</tr>
<tr>
<td>Table 9</td>
<td>GRADE EVIDENCE PROFILE: THE PERFORMANCE OF IGRA FOR THE DIAGNOSIS OF LATENT TUBERCULOSIS INFECTION IN CHILDREN IN LMIC</td>
<td>50</td>
</tr>
<tr>
<td>Table 10</td>
<td>GRADE SUMMARY OF FINDINGS — IGRA FOR LTBI IN CHILDREN</td>
<td>52</td>
</tr>
<tr>
<td>Table 11</td>
<td>GRADE EVIDENCE PROFILE: THE DIAGNOSTIC ACCURACY OF IGRA FOR THE DIAGNOSIS OF ACTIVE TUBERCULOSIS IN CHILDREN IN LMIC</td>
<td>53</td>
</tr>
<tr>
<td>Table 12</td>
<td>GRADE SUMMARY OF FINDINGS — IGRA FOR THE DIAGNOSIS OF ACTIVE TB IN CHILDREN</td>
<td>56</td>
</tr>
<tr>
<td>Table 13</td>
<td>RISK OF ACTIVE TUBERCULOSIS IN HIV-INFECTED INDIVIDUALS, STRATIFIED BY IGRA RESULT</td>
<td>62</td>
</tr>
<tr>
<td>Table 14</td>
<td>GRADE EVIDENCE PROFILE: THE ROLE OF IGRA IN THE DIAGNOSIS OF LATENT TUBERCULOSIS INFECTION IN HIV-INFECTED INDIVIDUALS</td>
<td>70</td>
</tr>
<tr>
<td>Table 15</td>
<td>GRADE SUMMARY OF FINDINGS — ROLE OF IGRA FOR DIAGNOSIS OF LTBI IN HIV INFECTED INDIVIDUALS</td>
<td>72</td>
</tr>
<tr>
<td>Table 16</td>
<td>GRADE EVIDENCE PROFILE: INTERFERON-γ RELEASE ASSAYS FOR TUBERCULOSIS SCREENING OF HEALTHCARE WORKERS IN LOW AND MIDDLE INCOME COUNTRIES</td>
<td>78</td>
</tr>
<tr>
<td>Table 17</td>
<td>GRADE SUMMARY OF FINDINGS — IGRA FOR TUBERCULOSIS SCREENING OF HEALTHCARE WORKERS IN LOW AND MIDDLE INCOME COUNTRIES</td>
<td>81</td>
</tr>
<tr>
<td>Table 18</td>
<td>RESULTS FROM STUDIES USING EXPOSURE GRADIENTS</td>
<td>90</td>
</tr>
<tr>
<td>Table 19</td>
<td>TST AND QFT CONVERSIONS IN CONTACT STUDIES IN LOW- AND MIDDLE-INCOME COUNTRIES</td>
<td>97</td>
</tr>
<tr>
<td>Table 20</td>
<td>GRADE EVIDENCE PROFILE: PERFORMANCE OF IGRA FOR THE DIAGNOSIS OF LTBI IN CONTACTS OF ACTIVE TB IN LOW- AND MIDDLE-INCOME COUNTRIES</td>
<td>99</td>
</tr>
<tr>
<td>Table 21</td>
<td>DISCORDANT-CONCORDANT TST/IGRA PAIRS AND INCIDENCE RATES OF TB</td>
<td>109</td>
</tr>
<tr>
<td>Table 22</td>
<td>GRADE EVIDENCE PROFILE: PREDICTIVE VALUE OF COMMERCIAL IGRA FOR INCIDENT ACTIVE TB IN LOW AND MIDDLE-INCOME COUNTRIES</td>
<td>111</td>
</tr>
<tr>
<td>Table 23</td>
<td>GRADE SUMMARY OF FINDINGS: PREDICTIVE VALUE OF COMMERCIAL IGRA FOR INCIDENT ACTIVE TB IN LOW AND MIDDLE-INCOME COUNTRIES</td>
<td>113</td>
</tr>
</tbody>
</table>
USE OF INTERFERON-\(\gamma\) RELEASE ASSAYS (IGRAs) IN TUBERCULOSIS CONTROL IN LOW- AND MIDDLE-INCOME SETTINGS

1. BACKGROUND

Tuberculosis (TB) continues to have a significant health impact worldwide, with one third of the world’s population estimated to be infected with *Mycobacterium tuberculosis*, resulting in so-called latent TB infection (LTBI). Until recently, the tuberculin skin test (TST) was the only tool available for LTBI diagnosis. The TST involves intradermal injection of purified protein derivative (PPD), a crude mixture of mycobacterial antigens, which stimulates a delayed type hypersensitivity response and causes induration at the injection site within 48 to 72 hours.

The identification of genes in the *Mycobacterium tuberculosis* genome that are absent from *M. bovis* BCG and most nontuberculous mycobacteria has supported the development of more specific and sensitive tests for detection of *M. tuberculosis*. *M. bovis* BCG has 16-gene deletions including the region of difference 1 (RD-1) that encodes for early secretory antigen target-6 (ESAT-6) and culture filtrate protein 10 (CFP-10).1, 2 24, 26 ESAT-6 and CFP-10 are strong targets of the cellular immune response in patients with *M. tuberculosis* infection.3, 4 27, 28 In such persons, sensitized memory/effector T cells produce interferon-gamma (IFN-\(\gamma\)) in response to these *M. tuberculosis* antigens, allowing a biologic basis for T-cell-based tests such interferon-gamma release assays (IGRAs).

Research over the past decade has resulted in the development of two commercial IGRAs. Both assays work on the principle that the T-cells of an individual who have acquired TB infection will respond to re-stimulation with *M. tuberculosis*-specific antigens by secreting interferon-gamma. The QuantiFERON-TB Gold (QFT-G, Cellestis, Australia) and the newer version QuantiFERON-TB Gold In-Tube (QFT-GIT, Cellestis, Australia) are whole-blood based enzyme-linked immunosorbent assays (ELISA) measuring the amount of IFN-\(\gamma\) produced in response to specific *M. tuberculosis* antigens (QFT-G: ESAT-6 and CFP-10, QFT-GIT: ESAT-&, CFP-10, TB7.7). In contrast, the enzyme-linked immunospot (ELISPOT)-based T-SPOT.TB (Oxford Immunotec, UK) measures the number of peripheral mononuclear cells that produce INF-\(\gamma\) after stimulation with ESAT-6 and CFP-10.

Both IGRAs and the TST are surrogate markers of *M. tuberculosis* infection, indicating a cellular immune response to recent or remote sensitization with *M. tuberculosis*. Currently, there is no gold standard for the detection of *M. tuberculosis* infection, and neither the TST nor IGRAs can distinguish TB infection from active TB disease.

Although routinely used, the TST has limited sensitivity and specificity. Factors related to the host, test administration and/or reading may diminish TST reactivity resulting in false-negative reactions and decreased TST sensitivity. Important factors associated with reduced TST sensitivity include malnutrition, young age, severe TB disease, HIV-related impaired cellular immunity, and other forms of immune suppression. Several factors are associated with decreased TST specificity and false-positive reactions including antigens shared between *M. tuberculosis* purified protein derivative (PPD), non-tuberculous mycobacteria (NTM) and BCG vaccine. Additionally, completing the TST requires two health care visits and measurement of reaction size is subjective, with documented poor inter-reader reliability. Nevertheless, the TST is the only test for which the risk of developing active TB in persons with a positive result has been well-defined.

IGRAs are the first new diagnostic test for latent tuberculosis infection (LTBI) in over 100 years. In previous systematic reviews it has been shown that, in low TB incidence settings, IGRAs have higher specificity than the TST, better correlation with surrogate measures of *M. tuberculosis* exposure, and less cross reactivity.
Use of IGRAs in diagnosis of active TB

with the BCG vaccine. IGRAs do, however, require fairly sophisticated laboratory infrastructure and technical expertise, and are costly.

In recent years, IGRAs have become widely endorsed in high-income countries for diagnosis of LTBI and several guidelines (albeit equivocal) on their use have been issued. Currently, there are no guidelines for their use in high TB- and HIV-burden settings, typically found in low-and middle-income countries, where IGRA use are being marketed and promoted, especially in the private sector. While strong evidence has emerged that IGRAs are unaffected by BCG vaccination, systematic reviews have suggested that IGRA performance differs in high- versus low TB and HIV incidence settings, with relatively lower sensitivity in high-burden settings. The WHO Stop TB department has therefore commissioned systematic reviews on the use of IGRAs in low- and middle-income settings, in pre-defined target groups, with funding support from the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and TREAT-TB/The Union. The target groups and the rationale for their selection are briefly summarized below:

Use of IGRAs in diagnosis of active TB: IGRAs were explicitly designed to replace the TST in diagnosis of LTBI, and were not intended for active TB. Diagnosis and treatment of LTBI remains limited in scope in most low- and middle-income countries, where detection and management of active TB is the highest priority for national TB programmes. Because IGRAs (like the TST) cannot distinguish LTBI from active TB, these tests can be expected to have poor specificity for active TB in high-burden settings due to a high background prevalence of LTBI. Additional differences in patient spectrum, such as anergy due to advanced disease, malnutrition, and HIV-associated immune suppression, or characteristics of the setting, such as laboratory procedures and infrastructure, may also contribute to a lower performance of IGRAs observed in these settings. Yet, especially private sector laboratories in high-burden, countries increasingly employ IGRAs for active TB diagnosis, and many investigators continue to recommend the use of IGRAs either as individual or adjunct tests for diagnosis of active TB.

Use of IGRAs in children: Children carry an estimated 15% of the global burden of TB disease. More than 60% of children <5 years of age diagnosed with TB in high-burden countries have documented household exposure, while community exposure increases with age. Children therefore constitute an increasing TB infection reservoir that are at high risk of primary disease progression in the absence of isoniazid preventive therapy (IPT) and who may also develop subsequent adult reactivation disease. In addition, young children have a disproportionately high risk of early progression to primary disease and developing severe forms of disease (eg. TB meningitis or miliary TB), often exacerbated by HIV infection (with increased mortality), especially in Sub-Saharan Africa. Limited public health resources are available to identify and manage the increasingly large pool of TB-infected children. In addition, the diagnosis of paucibacillary disease in children is complicated by the difficulty of bacteriological confirmation and often relies on a composite of risk factors, clinical and radiological findings, all of which are rather unspecific. Diagnostic algorithms for pediatric disease often include use of the TST, with a positive TST considered supportive of the diagnosis. Possible improved performance of IGRAs over TST in this context therefore needs to be explored.

Use of IGRAs in HIV-infected individuals: TB has become the leading cause of death in persons with HIV and HIV is the most potent risk factor for progression from latent to active TB. Preventative therapy with isoniazid reduces the risk of active TB by up to 60%; however, the optimal test to identify HIV-infected individuals who could benefit from IPT remains uncertain. Importantly, there is strong evidence that IPT reduces the risk of TB in persons with positive TST results (irrespective of HIV result); however, the TST is impaired in HIV infection, and severely compromised in individual with a low CD4 count. Data are urgently
Use of IGRAs in diagnosis of active TB

needed to evaluate the use of IGRAs to improve the identification of HIV-infected persons who could benefit from IPT, diagnosing LTBI rather than ruling out active TB (an important distinction in HIV-infected persons initiating IPT).

**Use of IGRAs in health care worker (HCW) screening and contact investigations**: TB poses a significant occupational health problem and HCWs are at increased risk for exposure to TB and subsequent disease, especially if co-infected with HIV. In many high-income countries, periodic screening of HCWs and contacts of confirmed TB patients for LTBI is a routine component of TB control; however, contact and HCW screening is often neglected in low- and middle-income settings. Traditionally, prevalence of LTBI and incidence of new TB infection (i.e. conversion) among such individuals have been estimated using the TST. IGRAs have emerged as an alternative, being *ex-vivo* blood-based tests that, in contrast to the TST, can be repeated any number of times without sensitization or boosting. However, data are lacking on how to interpret repeated (serial) IGRA testing results and studies have documented conversions and reversions during serial testing. Several questions also remain about the usefulness of IGRAs to determine incidence of new infections among HCWs and contacts, an issue critical for understanding of TB transmission, nosocomial spread, and the impact of existing and new TB infection control interventions and strategies.

**Predictive value of IGRAs**: The clinical benefit of IGRAs, supported by data on the longitudinal predictive (prognostic) value of IGRA and their added value in the control of TB is currently unknown. In contrast, the predictive value of a positive TST has been well-defined, showing that TST reactivity is associated with an increased risk of active TB in subsequent years. Strong evidence from randomized trials has shown that IPT benefit is restricted to individuals with a positive TST (irrespective of HIV result), providing a relative risk reduction of around 60%. To demonstrate equivalent or superior clinical utility of IGRAS over TST, IGRAs would have to be subjected to similar evaluations and in various at-risk populations, especially in low- and middle-income countries with limited and often competing public health resources.

2. **EVIDENCE BASE**

2.1 **Evidence synthesis**

The systematic, structured, evidence-based process for TB diagnostic policy generation developed by WHO-STB was followed: The first step constituted systematic reviews and meta-analysis of available data (published and unpublished) using standard methods appropriate for diagnostic accuracy studies. The second step involved the convening of an Expert Group to a) evaluate the strength of the evidence base; b) recommend operational and logistical considerations for mainstreaming such methods/approaches into national TB control programmes; and c) identify gaps to be addressed in future research. Based on the Expert Group findings, the third and final step involves WHO policy guidance on the use of these tools/approaches, presented to the WHO Strategic and Technical Advisory Group for TB (STAG-TB) for consideration, and eventual dissemination to WHO Member States for implementation.

The Expert Group (Annex 1) consisted of researchers, clinicians, epidemiologists, end-users (programme and laboratory representatives), community representatives and evidence synthesis experts. The Expert Group meeting followed a structured agenda (Annex 1) and was co-chaired by WHO secretariat.

To comply with current standards for evidence assessment in formulation of policy recommendations, the GRADE system ([www.gradeworkinggroup.org](http://www.gradeworkinggroup.org)), adopted by WHO for all policy and guidelines development, was used. The GRADE approach, assessing both the quality of evidence and strength of recommendations, aims to provide a comprehensive and transparent approach for developing policy guidance. Started about 10 years ago to assess treatment interventions, the GRADE approach has recently been refined for
diagnostics; however, while the latter process shares the fundamental logic of recommendations for other interventions (notably treatment), it also presents unique challenges, most often due to study limitations related to a lack of data on patient-important outcomes and impact (see below).

Given the absence of studies evaluating patient-important outcomes among TB suspects randomized to treatment based on IGRA results, reviews were focused on the diagnostic accuracy of IGRAs versus TST in detecting LTBI or active TB.

2.2 Systematic reviews and meta-analyses

Systematic reviews were done following standard guidelines for systematic reviews and detailed protocols with predefined questions relevant to the individual topics. Summaries of methodology followed for each topic are given in the respective sections below. Detailed methodology is described in the individual systematic review reports available at http://www.who.int/tb/laboratory/policy_statements.

Hierarchy of reference standards: Studies evaluating the performance of IGRAs are hampered by the lack of an adequate gold standard to distinguish the presence or absence of LTBI. Since diagnostic accuracy for LTBI could not be directly assessed, a hierarchy of reference standards was developed and agreed beforehand with the systematic reviewers (Figure 1) that would evaluate the role of IGRAs depending on the individual topic (i.e. not all systematic reviews necessarily used the hierarchy).

Primary outcomes were predefined for each systematic review as relevant, e.g. the predictive value of IGRAs for development of active TB, the sensitivity of IGRAs in persons with culture-confirmed active TB (as a surrogate reference standard for TB infection), and the correlation between IGRA and TST results.

In addition to primary outcomes, specific characteristics of IGRAs that could influence their overall utility were evaluated where relevant, e.g. the proportion of indeterminate IGRA results (i.e. not interpretable either due to high IFN-γ response in the negative control or low IFN-γ response in the positive control), the impact of HIV-related immunosuppression (i.e. CD4+ cell count) on test performance where available, and correlation of IGRA results with an exposure gradient (typically used in contact and outbreak investigations).

Figure 1. Hierarchy of reference standards used to assess the evidence base

![Hierarchy of reference standards](image-url)
Search methods: All studies evaluating IGRAs published through May 2010 were reviewed using predefined data search strings. In addition to database searches, bibliographies of reviews and guidelines were reviewed, citations of all included studies were screened, and experts in the field as well as IGRA manufacturers were contacted to identify additional published, unpublished, and ongoing studies. Pertinent information not reported in the original publication, from the primary authors of all studies included, were requested and obtained by the systematic reviewers.

Study selection: Studies published in all languages and in all low- and middle-income countries that evaluated the performance of the newest commercial, RD1 antigen-based IGRAs were reviewed per individual topic. Excluded were: (1) studies that evaluated non-commercial (in-house) IGRAs, older generation IGRAs (i.e., purified protein derivative [PPD]-based IGRAs) and IGRAs performed in specimens other than blood; (2) studies focused on the effect of anti-TB treatment on IGRA response; (3) studies including < 10 individuals; (4) studies reporting insufficient data to determine diagnostic accuracy measures; and (5) conference abstracts, letters without original data, and reviews.

Assessment of study quality: Study quality was assessed by relevant standardised methods depending on the topic. For primary outcomes focused on test accuracy, a subset of relevant criteria from QUADAS, a validated tool for diagnostic accuracy studies, was used. For studies of the predictive value of IGRAs, quality was appraised with a modified version of the Newcastle-Ottawa Scale (NOS) for longitudinal/cohort studies.

<table>
<thead>
<tr>
<th>Table 1: QUADAS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item #</strong></td>
</tr>
<tr>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>3.</td>
</tr>
<tr>
<td>4.</td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td>6.</td>
</tr>
<tr>
<td>7.</td>
</tr>
<tr>
<td>8.</td>
</tr>
<tr>
<td>9.</td>
</tr>
<tr>
<td>10.</td>
</tr>
<tr>
<td>11.</td>
</tr>
<tr>
<td>12.</td>
</tr>
<tr>
<td>13.</td>
</tr>
<tr>
<td>14.</td>
</tr>
</tbody>
</table>

Conflicts of interest are a known concern in TB diagnostic studies; therefore, the systematic reviews added a quality item about involvement of commercial test manufacturers in published studies and reported whether IGRA manufacturers had any involvement with the design or conduct of each study, including donation of test materials, provision of monetary support, work/financial relationships with study authors, and participation in data analysis.

Outcome definitions: Explicit definitions for primary and secondary outcomes were defined in the original systematic review protocols, pre-specified per individual topic and described in the individual sections below.
Data synthesis and meta-analysis: A standardised overall approach was specified a priori for each systematic review to account for significant heterogeneity in results expected between studies. First, data were synthesised separately for each commercial IGRA and by the World Bank country income classification (low/middle income versus high income) as a surrogate for TB incidence. Second, heterogeneity was visually assessed using forest plots, the variation in study results attributable to heterogeneity was characterised (I-squared statistic), and statistically tested (chi-squared test). Third, pooled estimates were calculated using random effects modelling, which provides more conservative estimates than fixed effects modelling when heterogeneity is present.

For each individual study, all outcomes for which data were available were assessed. First, forest plots were generated to display the individual study estimates and their 95% confidence intervals. Pooled estimates were calculated when at least three studies were available in any sub-group and individual study results summarised when less than four studies were available. Standard statistical packages were used for analyses.

2.3 GRADE evaluation

Evaluation followed the GRADE system for grading quality of evidence and strength of recommendations for diagnostic tests. The quality of evidence was graded by six criteria:

- **Study design**
  
  Cross-sectional: Random or consecutive selection of patients/specimens at risk (preferred);
  Case-control: Selection of patients/specimens according to reference standard.

- **Risk of bias (as reflected by the QUADAS tool)**
  
  Compliance of studies with relevant independent quality assessment criteria (Table 1).

- **Directness**
  
  Presence of direct evidence of impact on patient-important outcomes and generalisability.

- **Inconsistency**
  
  Unexplained inconsistency in sensitivity or specificity estimates.

- **Imprecision**
  
  Wide confidence intervals for pooled sensitivity or specificity estimates.

- **Publication/reporting bias**
  
  Publications of research based on their nature and outcome, eg. studies showing poor performance not being published, language bias, etc.

As called for by GRADE, the Expert Group also considered for each method/approach the strength of the recommendation (strong or conditional/weak), based on a balance of effects (advantages weighed against disadvantages), patient values and preferences, and costs.
Use of IGRAs in diagnosis of active TB

Given the absence of relevant data from the studies reviewed, assumed patient values and preferences were assessed by test accuracy as a proxy measure, based on the relative importance/impact of false-positive and false-negative results:

**True positives:** Benefit to patients from earliest diagnosis and treatment;

**True negatives:** Patients spared unnecessary treatment; benefit of reassurance and alternative diagnosis;

**False positives:** Likely patient anxiety and morbidity from additional testing, unnecessary treatment; may halt further diagnostic evaluation;

**False negatives:** Increased risk of patient morbidity and mortality, and continued risk of community transmission of TB.

### 2.4 Meeting procedural issues

The systematic review reports were made available to the Expert Group for scrutiny before the meeting.

As agreed, interchange by Expert Group meeting participants was restricted to those who attended the Expert Group meeting in person, both for the discussion and follow-up dialogue.

WHO is committed to ensuring that the highest standards of evidence are used in formulation of recommendations and has therefore standardised the synthesis process based on the GRADE approach. The first paper specifically addressing the GRADE approach to diagnostic tests and strategies was published in 2008 (Schunemann. BMJ 2008; 336:1106-1110) and was made available to the Expert Group in the background documentation for the meeting.

It was explained that individuals were selected to be members of the Expert Group to carefully represent and balance important perspectives for the process of formulating recommendations. Therefore the Expert Group included technical experts, end-users, patient representatives and evidence synthesis methodologists.

Expert Group members were asked to submit completed Declaration of Interest (DOI) forms. These were reviewed by the WHO legal department prior to the Expert Group meeting. A summary is attached in Annex 2. DOI statements were summarised by the co-chair (WHO-STB) of the Expert Group meeting at the start of the meeting.

Selected individuals with intellectual and/or research involvement in the methods were invited as observers to provide technical input and answer technical questions. These individuals did not participate in the GRADE evaluation process and were excluded from the Expert Group discussions when recommendations were developed. They were also not involved in the development of the final Expert Group meeting reports, nor in preparation of the STAG-TB documentation or preparation of the final WHO policy statements.
3. RESULTS

3.1 Use of IGRAs in diagnosis of active TB

3.1.1 Objectives, reference standards and outcomes

Objective
To assess IGRA test performance in active pulmonary TB suspects and confirmed TB cases in low- and middle-income settings.

Reference standards
Well-designed diagnostic accuracy studies focus on a representative target population in whom genuine diagnostic uncertainty exists (i.e. patients in whom clinicians would apply the test in the course of regular clinical practice). Evidence suggests that diagnostic studies that include only known cases and healthy controls tend to overestimate test accuracy. Therefore, it was considered that studies simultaneously evaluating IGRA sensitivity and specificity among active TB suspects represent the highest quality evidence, while studies evaluating IGRA performance among patients with known active TB (for sensitivity) are of lesser quality.

Because of the focus on active TB diagnostic accuracy and the high prevalence of LTBI in high TB-burden settings, IGRA specificity was estimated exclusively among studies enrolling active TB suspects where the diagnostic workup ultimately showed no evidence of active disease.

Acceptable ‘gold standard’ tests are available to diagnose active TB disease (although these vary according to local resources). The hierarchy of reference standards (Figure 1.) for active TB was therefore used to judge the quality of each individual assessment of IGRA diagnostic accuracy. From most to least favourable, these reference standards included:

1) Culture-confirmation or sputum smear-positivity in high TB incidence settings (defined for this purpose as ≥50/100,000), where sputum smear microscopy has been shown to have high specificity;

2) Sputum smear-positivity without culture in low or intermediate TB incidence settings (defined for this purpose as <50/100,000);

3) Clinical diagnosis based on presenting symptoms, radiology and/or response to TB treatment without microbiological confirmation.

Because the TST remains in widespread use, and because indeterminate IGRA results may affect assay performance in low-income settings, the following were also evaluated:

1) Observed differences in sensitivity for active TB diagnosis between IGRA and TST;

2) The proportion of IGRA results among patients with active disease which are indeterminate.

Primary outcomes
Sensitivity: proportion of individuals with a positive IGRA result among those with culture-positive TB (indeterminate IGRA results included in the denominator if they occurred in individuals with culture-positive TB);
Specificity: proportion of individuals with a negative IGRA result among those ruled out for active TB disease (indeterminate IGRA results excluded from analysis).

Using the GRADE framework, sensitivity and specificity were interpreted as proxies for patient-important outcomes: Poor sensitivity would result in false-negative results where TB patients will be missed with consequences for morbidity, mortality and transmission of tuberculosis. Poor specificity would result in false-positive results where patients without TB disease will be prescribed unnecessary anti-TB therapy for six months, with possible consequences including adverse events such as hepatotoxicity and (rarely) deaths.

Secondary outcome
Incremental or added value of IGRAs to determine if these contributed to active TB diagnosis beyond that established through conventional tests (symptoms, chest radiograph, and sputum smears). Such studies would be expected to use multivariable analysis to determine the added value of IGRAs.

3.1.2 Search results
The initial search yielded 789 citations. After full-text review of 168 papers, 19 papers were determined to meet eligibility criteria for IGRA evaluation of active TB in low- and middle-income settings (Annex 2). Three unpublished reports were also included, for a total of 22 papers. Because some papers included more than one commercial IGRA, there were 33 unique evaluations (referred to as studies) – 21 of QFT-GIT, and 12 of T-SPOT – that included a total of 1,815 HIV-uninfected and 1,057 HIV-infected individuals.

3.1.3 Data analysis
Multiple sources of heterogeneity commonly exist when estimates from studies of diagnostic tests are summarised. Therefore, expected heterogeneity was approached as follows: First, when possible, data were synthesized separately for each commercial IGRA and by HIV status. The pre-specified sub-groups minimized heterogeneity related to differences in testing platform (ELISA vs. ELISPOT), antigens used to elicit IFN-γ release (ESAT-6/CFP-10 vs. ESAT-6/CFP-10/TB 7.7), and test performance related to HIV-associated host immunosuppression. Second, heterogeneity was visually assessed using forest plots, and variation in study results attributable to heterogeneity was characterised (I-squared value) and statistically tested (chi-squared test). Third, pooled estimates were calculated using random effects modelling, which provides more conservative estimates than fixed effects modelling when heterogeneity is a concern.

For each individual study, all outcomes were assessed where data were available. First, forest plots were generated to display the individual study estimates and their 95% confidence intervals. Second, bivariate random effects regression models were used when both sensitivity and specificity could be reported from the same TB suspect population. Pooling sensitivity and specificity separately can produce biased estimates of test accuracy; therefore pooled estimates were generated when both sensitivity and specificity were reported within a study and ranked as higher quality evidence. Third, hierarchical summary receiver operating characteristic (HSROC) curves were generated to summarize global test performance.

Because of the need to summarize two correlated measures (sensitivity and specificity), and because substantial inter-study heterogeneity is common, meta-analysis of diagnostic accuracy requires different and more complex methods than traditional meta-analytic techniques. Graphically illustrating the trade-off between sensitivity and specificity, HSROC curves differ from traditional Receiver Operating Characteristics (ROC) in allowing accuracy to vary by each individual study (ie. allowing for random effects, and thus allowing asymmetry in the plotted curve), and by discouraging extrapolation beyond the available data by plotting the curve only over the observed range of test characteristics. The HSROC approach is closely
related to the bivariate random effects regression model. These methods generally produce similar results and are both recommended by the Cochrane Diagnostic Test Accuracy Working Methods group.

Pooled estimates were calculated when at least four studies were available in any sub-group and summarized individual study results when fewer than four studies were available. All analyses were performed using Stata 11 (Stata Corporation, College Station, Texas, USA), including the user-written “metandi” programme for Stata for bivariate random effects regression and HSROC analyses.

### 3.1.4 Study characteristics

Of the total studies, 10 (30%) were from low income countries, and 23 (70%) were from middle income countries. Seventeen studies (52%) included HIV-infected individuals, and 27 (82%) studies involved ambulatory subjects (i.e. out-patients as well as hospitalized patients). IGRAs were performed in persons suspected of having active TB in 19 (58%) studies and in persons with known active TB in 14 (42%) studies.

### 3.1.5 Study quality

The majority of studies satisfied the QUADAS criteria assessed (Table 1), with the exception of patient spectrum (biased sampling) and blinding. Seventeen (52%) studies did not enrol a representative spectrum of patients, and nine (27%) studies did not clearly report that assessment of the reference standard was performed blinded to IGRA results. Industry involvement was unknown in six (18%) studies and acknowledged in nine (27%) studies, including donation of IGRA kits (6 studies) and work/financial relationships between authors and IGRA manufacturers (3 studies).

Figure 2. Assessment of study quality using the QUADAS tool

### 3.1.6 Sensitivity and specificity estimation among TB suspects

A total of 19 studies were identified that simultaneously estimated sensitivity and specificity in TB suspects, and test accuracy estimates were pooled using bivariate random effects/HSROC methods. Overall, studies enrolling active TB suspects demonstrated a sensitivity of 83% (95% CI 70-91%) and specificity of 58% (95% CI 42-73%) for T-SPOT (8 studies), and a sensitivity of 73% (95% CI 61-82%) and specificity of 49% (95% CI 40-58%) for QFT-GIT (11 studies).
Sensitivity

With the exception of two studies, the sensitivity of IGRAs was assessed based on a positive culture result (27 studies, 82%) or a positive AFB sputum-smear result in a high TB incidence setting (4 studies, 12%). Among studies performed in patients with known active TB, 6 (43%) included patients who had been treated for greater than one week.

**HIV-positive:** Eleven studies assessed IGRA sensitivity among HIV-positive active TB suspects. HSROC/bivariate pooled sensitivity estimates were higher for T-SPOT (78%, 95% CI 56-91%, 5 studies) than for QFT-GIT (62%, 95% CI 41-79%, 6 studies (Figure 3A-3C).

**Figure 3A - 3C.** Hierarchical summary receiver operating characteristics (HSROC) plot of studies that reported both sensitivity and specificity in active TB suspects

The summary curves from the HSROC model contain a summary operating point (red square) representing summarised sensitivity and specificity point estimates for individual study estimates (open circles). The 95% confidence region is delineated by the area within the orange dashed line.

Pooled sensitivity estimates did not appreciably change for either T-SPOT (70%, 95% CI 59-82%, 6 studies) or QFT-GIT (65%, 95% CI 56-73%, 9 studies) when studies evaluating patients with known active TB were included in the analysis (Figure 4). Pooled sensitivity estimates for both T-SPOT (I-squared 74%, p<0.01) and QFT-GIT (I-squared 69%, p=0.001) showed significant heterogeneity.
Figure 4. Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB in HIV-positive persons with confirmed active tuberculosis in low- and middle-income countries

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sensitivity (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>QFT-GIT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aabye 2009</td>
<td>Tanzania</td>
<td>65 (52, 76)</td>
<td>13</td>
</tr>
<tr>
<td>Kabeer 2009</td>
<td>India</td>
<td>66 (50, 80)</td>
<td>11</td>
</tr>
<tr>
<td>Leidl 2009</td>
<td>Uganda</td>
<td>74 (49, 91)</td>
<td>9</td>
</tr>
<tr>
<td>Ling 2010</td>
<td>South Africa</td>
<td>67 (50, 80)</td>
<td>11</td>
</tr>
<tr>
<td>Markova 2009</td>
<td>Bulgaria</td>
<td>92 (64, 100)</td>
<td>10</td>
</tr>
<tr>
<td>Raby 2008</td>
<td>Zambia</td>
<td>63 (49, 75)</td>
<td>12</td>
</tr>
<tr>
<td>Rangaka 2010</td>
<td>South Africa</td>
<td>63 (53, 72)</td>
<td>14</td>
</tr>
<tr>
<td>Tsiouris 2006</td>
<td>South Africa</td>
<td>65 (44, 83)</td>
<td>9</td>
</tr>
<tr>
<td>Veldsman 2009</td>
<td>South Africa</td>
<td>30 (15, 49)</td>
<td>10</td>
</tr>
<tr>
<td><strong>Subtotal (I-squared=69%, p=0.001)</strong></td>
<td></td>
<td>65 (56, 73)</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sensitivity (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TSPOT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattamanchi 2010</td>
<td>Uganda</td>
<td>54 (45, 64)</td>
<td>20</td>
</tr>
<tr>
<td>Jiang 2009</td>
<td>China</td>
<td>66 (47, 81)</td>
<td>15</td>
</tr>
<tr>
<td>Leidl 2009</td>
<td>Uganda</td>
<td>89 (67, 99)</td>
<td>16</td>
</tr>
<tr>
<td>Ling 2010</td>
<td>South Africa</td>
<td>82 (66, 92)</td>
<td>18</td>
</tr>
<tr>
<td>Markova 2009</td>
<td>Bulgaria</td>
<td>62 (32, 86)</td>
<td>10</td>
</tr>
<tr>
<td>Oni 2010</td>
<td>South Africa</td>
<td>68 (57, 78)</td>
<td>20</td>
</tr>
<tr>
<td><strong>Subtotal (I-squared=74%, p&lt;0.01)</strong></td>
<td></td>
<td>70 (59, 82)</td>
<td>100</td>
</tr>
</tbody>
</table>

The forest plots display the sensitivity estimates obtained from individual studies and pooled estimates derived from random effects (DerSimonian-Laird) modelling.

**HIV-negative**: Eight studies assessed IGRA sensitivity among HIV-negative active TB suspects. HSROC/bivariate pooled sensitivity estimates for were QFT-GIT were 82%, 95% CI 76-87%, 5 studies; data were insufficient to report HSROC/bivariate pooled sensitivity estimates for T-SPOT. Pooled sensitivity estimates were similar for T-SPOT (87%, 95% CI 82-91%, 5 studies) and QFT-GIT (85%, 95% CI 80-90%, 11 studies) when studies evaluating patients with known active TB were included in the analysis (Figure 5). Pooled sensitivity estimates showed significant heterogeneity for QFT-GIT (I-squared 53%, p=0.02), but not for T-SPOT (I-squared 6%, p=0.38).
Figure 5. Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB in HIV-negative persons with confirmed active tuberculosis in low- and middle-income countries

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sensitivity (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aabye 2009</td>
<td>Tanzania</td>
<td>81 (71, 88)</td>
<td>12</td>
</tr>
<tr>
<td>CDC TO20</td>
<td>Vietnam</td>
<td>88 (80, 93)</td>
<td>14</td>
</tr>
<tr>
<td>Chigou 2009</td>
<td>South Africa</td>
<td>96 (78, 100)</td>
<td>10</td>
</tr>
<tr>
<td>Chen 2009</td>
<td>China</td>
<td>85 (71, 94)</td>
<td>9</td>
</tr>
<tr>
<td>Dheda 2009</td>
<td>South Africa</td>
<td>73 (45, 92)</td>
<td>3</td>
</tr>
<tr>
<td>Katiyar 2008</td>
<td>India</td>
<td>95 (87, 99)</td>
<td>15</td>
</tr>
<tr>
<td>Ling 2010</td>
<td>South Africa</td>
<td>81 (71, 89)</td>
<td>11</td>
</tr>
<tr>
<td>Pai 2007</td>
<td>India</td>
<td>74 (60, 84)</td>
<td>9</td>
</tr>
<tr>
<td>Raby 2008</td>
<td>Zambia</td>
<td>84 (68, 94)</td>
<td>8</td>
</tr>
<tr>
<td>Tahereh 2010</td>
<td>Iran</td>
<td>77 (59, 90)</td>
<td>6</td>
</tr>
<tr>
<td>Tsiouris 2006</td>
<td>South Africa</td>
<td>77 (46, 95)</td>
<td>3</td>
</tr>
<tr>
<td>Pooled Estimate (I-squared 53%, p = 0.019)</td>
<td></td>
<td>85 (80, 90)</td>
<td>100</td>
</tr>
<tr>
<td>TSPOT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dheda 2009</td>
<td>South Africa</td>
<td>93 (68, 100)</td>
<td>9</td>
</tr>
<tr>
<td>Ling 2010</td>
<td>South Africa</td>
<td>85 (76, 92)</td>
<td>31</td>
</tr>
<tr>
<td>Ozekinci 2007</td>
<td>Turkey</td>
<td>93 (76, 99)</td>
<td>17</td>
</tr>
<tr>
<td>Shao-ping 2009</td>
<td>China</td>
<td>91 (71, 99)</td>
<td>11</td>
</tr>
<tr>
<td>Soysal 2008</td>
<td>Turkey</td>
<td>81 (72, 88)</td>
<td>31</td>
</tr>
<tr>
<td>Pooled Estimate (I-squared 6%, p = 0.375)</td>
<td></td>
<td>87 (82, 91)</td>
<td>100</td>
</tr>
</tbody>
</table>

The forest plots display the sensitivity estimates obtained from individual studies and pooled estimates derived from random effects (DerSimonian-Laird) modelling.

*Head-to-head comparisons of QFT and T-SPOT sensitivity:* Six studies (four involving HIV-positive subjects and two involving HIV-negative subjects) reported head-to-head comparisons of T-SPOT and QFT-GIT sensitivity. Overall, T-SPOT sensitivity was higher but not significantly different from QFT-GIT sensitivity (sensitivity difference 14%, 95% CI -4% to 33%, p=0.14) (Table 2). Results were similar when restricted to HIV-positive subjects (sensitivity difference 18%, 95% CI -16% to 51%, p=0.30). Among HIV-negative subjects, T-SPOT had higher sensitivity (4-20%) in both available studies.
### Table 2. Head-to-head comparison of sensitivity of T-SPOT.TB versus QuantiFERON-TB Gold In-Tube among active TB suspects

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Country</th>
<th>HIV Status</th>
<th>Active TB, n (%)</th>
<th>Positive T-SPOT result, n (%)</th>
<th>Positive QFT result, n (%)</th>
<th>Sensitivity difference* (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dheda, 2009†</td>
<td>South Africa</td>
<td>HIV-</td>
<td>15 (31), 15 (29)</td>
<td>14 (93)</td>
<td>11 (73)</td>
<td>20</td>
</tr>
<tr>
<td>Ling, 2010†</td>
<td>South Africa</td>
<td>HIV-</td>
<td>82 (35), 81 (34)</td>
<td>70 (85)</td>
<td>66 (82)</td>
<td>4</td>
</tr>
<tr>
<td>Dheda, 2009</td>
<td>South Africa</td>
<td>HIV+</td>
<td>5 (25)</td>
<td>5 (100)</td>
<td>1 (20)</td>
<td>80</td>
</tr>
<tr>
<td>Leidl, 2009</td>
<td>Uganda</td>
<td>HIV+</td>
<td>19 (15)</td>
<td>17 (89)</td>
<td>14 (74)</td>
<td>15</td>
</tr>
<tr>
<td>Markova, 2009</td>
<td>Bulgaria</td>
<td>HIV+</td>
<td>13 (14)</td>
<td>8 (62)</td>
<td>12 (92)</td>
<td>-31</td>
</tr>
<tr>
<td>Ling, 2010†</td>
<td>South Africa</td>
<td>HIV+</td>
<td>39 (39), 42 (39)</td>
<td>32 (82)</td>
<td>28 (67)</td>
<td>15</td>
</tr>
</tbody>
</table>

* Sensitivity difference (%) is T-SPOT sensitivity (%) minus QFT-GIT sensitivity (%).
† Total numbers of active TB suspects evaluated by each IGRA differed within some studies; these are listed in the order T-SPOT, QFT-GIT.
Head-to-head comparison of TST and IGRA sensitivity: TST sensitivity in the seven studies involving HIV-negative patients was higher (80%, 95% CI 74-85%) than in the six studies involving HIV-positive patients (58%, 95% CI 32-84%). Fifteen studies reported head-to-head comparisons of IGRAs (5 T-SPOT and 10 QFT-GIT) with TST. Overall, IGRA sensitivity was not statistically different than TST sensitivity for either T-SPOT (sensitivity difference 9%, 95% CI -10% to 28%, p=0.34) or QFT-GIT (sensitivity difference 1%, 95% CI -11% to 13%, p=0.89) (Figure 6). There was significant heterogeneity for both estimates (I-squared >80%, p<0.001). Results were unchanged for QFT-GIT when restricted to either HIV-positive or HIV-negative subjects (data not shown); there were insufficient studies to form HIV-stratified pooled sensitivity difference estimates for T-SPOT.

Figure 6. Percent sensitivity difference between IGRA and TST results

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sensitivity Difference</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDC T020</td>
<td>Vietnam</td>
<td>0.09 (-0.01, 0.19)</td>
<td>12</td>
</tr>
<tr>
<td>Kabeer 2009</td>
<td>India</td>
<td>0.41 (0.22, 0.60)</td>
<td>10</td>
</tr>
<tr>
<td>Katiyar 2008</td>
<td>India</td>
<td>0.18 (0.08, 0.29)</td>
<td>12</td>
</tr>
<tr>
<td>Ling 2010</td>
<td>South Africa</td>
<td>-0.07 (-0.21, 0.08)</td>
<td>11</td>
</tr>
<tr>
<td>Ling 2010*</td>
<td>South Africa</td>
<td>-0.33 (-0.51, -0.16)</td>
<td>10</td>
</tr>
<tr>
<td>Raby 2008*</td>
<td>Zambia</td>
<td>0.07 (-0.11, 0.26)</td>
<td>10</td>
</tr>
<tr>
<td>Raby 2008</td>
<td>Zambia</td>
<td>0.03 (-0.15, 0.21)</td>
<td>11</td>
</tr>
<tr>
<td>Rangaka 2010*</td>
<td>South Africa</td>
<td>-0.03 (-0.17, 0.11)</td>
<td>11</td>
</tr>
<tr>
<td>Tsirou 2006*</td>
<td>South Africa</td>
<td>-0.19 (-0.42, 0.04)</td>
<td>9</td>
</tr>
<tr>
<td>Tsirou 2006</td>
<td>South Africa</td>
<td>-0.16 (-0.43, 0.11)</td>
<td>9</td>
</tr>
<tr>
<td>Pooled Estimate (I-squared 82%, p&lt;0.001)</td>
<td>0.01 (-0.11, 0.13)</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

| TSPOT     |              |                        |          |
| Jiang 2009 | China        | 0.50 (0.29, 0.71)      | 18       |
| Ling 2010  | South Africa | -0.03 (-0.16, 0.10)   | 21       |
| Ling 2010* | South Africa | -0.18 (-0.34, -0.02)  | 20       |
| Ozbekini 2007 | Turkey     | 0.11 (-0.06, 0.28)    | 20       |
| Soysal 2008 | Turkey      | 0.11 (-0.01, 0.23)    | 21       |
| Pooled Estimate (I-squared 87%, p<0.001) | 0.09 (-0.10, 0.28) | 100       |

Specificity

All specificity estimates were determined in TB suspects using HSROC/bivariate techniques. Overall, pooled specificity was low for both T-SPOT (58%, 95% CI 42-73%, 8 studies) and QFT-GIT (49%, 95% CI 40-58%, 11 studies). When stratified by HIV-status, pooled specificity for QFT-GIT was higher among HIV-positive active TB suspects (51%, 95% CI 39-64%, 6 studies) than HIV-negative active TB suspects (42%, 95% CI 33-53%, 5 studies). When restricted to HIV-positive active TB suspects, pooled specificity for T-SPOT was 55% (95% CI
45-64%, 5 studies); an insufficient number of studies were available to estimate pooled specificity for T-SPOT among HIV-uninfected patients.

### 3.1.7 Proportion of indeterminate IGRA results

The proportion of indeterminate IGRA results among patients with suspected or confirmed active TB varied considerably (range 0-26% among studies enrolling 50 or more subjects). The proportion of indeterminate results was low (3%, 95% CI 1-5%) among HIV-negative patients, regardless of IGRA platform (Figure 7). However, the proportion of indeterminate results was considerably higher among HIV-positive subjects for both QFT-GIT (16%, 95% CI 10-21%, 10 studies) and T-SPOT (8%, 95% CI 1-15%, 7 studies) (Figure 8). Results were similar for HIV-positive subjects when stratified by TB suspects versus known TB cases.

**Figure 7.** Proportion of indeterminate IGRA results among HIV-negative subjects in low- and middle-income countries

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>% Indeterminate</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>QFT-GIT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aabye 2009</td>
<td>Tanzania</td>
<td>9 (4, 16)</td>
<td>10</td>
</tr>
<tr>
<td>CDC TO20</td>
<td>Vietnam</td>
<td>0 (0, 1)</td>
<td>19</td>
</tr>
<tr>
<td>Chegou 2009</td>
<td>South Africa</td>
<td>0 (0, 15)</td>
<td>8</td>
</tr>
<tr>
<td>Chen 2009</td>
<td>China</td>
<td>4 (4, 25)</td>
<td>5</td>
</tr>
<tr>
<td>Dheda 2009</td>
<td>South Africa</td>
<td>27 (16, 42)</td>
<td>4</td>
</tr>
<tr>
<td>Katyar 2008</td>
<td>India</td>
<td>0 (0, 5)</td>
<td>17</td>
</tr>
<tr>
<td>Ling 2010</td>
<td>South Africa</td>
<td>7 (4, 11)</td>
<td>15</td>
</tr>
<tr>
<td>Pai 2007</td>
<td>India</td>
<td>0 (0, 6)</td>
<td>15</td>
</tr>
<tr>
<td>Raby 2008</td>
<td>Zambia</td>
<td>14 (5, 20)</td>
<td>4</td>
</tr>
<tr>
<td>Tsioirus 2005</td>
<td>South Africa</td>
<td>0 (0, 25)</td>
<td>4</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 78%, p&lt;0.001)</strong></td>
<td>4 (1, 6)</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

| **TSPOT**      |                |  |          |
| Dheda 2009     | South Africa   | 4 (0, 14) | 6        |
| Ling 2010      | South Africa   | 2 (1, 5)  | 64       |
| Ozekinci 2007  | Turkey         | 0 (0, 12)  | 8        |
| Shao-ping 2009 | China          | 7 (3, 15)  | 7        |
| Soysal 2008    | Turkey         | 4 (1, 10)  | 15       |
| **Pooled Estimate (I-squared 0%, p=0.47)** | 3 (1, 4) | 100 |
3.1.8 Incremental value of IGRAs for active TB

Two completed but unpublished studies were identified (both from South Africa). These studies used multivariate methods to estimate the added value of IGRAs beyond conventional tests for active TB. The first focused on 164 smear-negative TB suspects in South Africa, examining changes in the area under the receiver operating curve (AUC) resulting from addition of chest radiograph and/or IGRA (QFT-GIT and T-SPOT) to a baseline clinical prediction model including age, male sex, previous TB, HIV, haemoptysis, weight loss, and loss of appetite (selected using stepwise regression). When added to the baseline model (AUC 0.75 (0.65-0.85), chest radiograph (AUC 0.80 (0.70-0.89), p=0.12) nor IGRA (T-SPOT, AUC 0.78 (0.68-0.88), p=0.28; QFT-GIT, AUC 0.78 (0.67-0.88), p=0.30) statistically contributed additional diagnostic information. Neither QFT-GIT nor T-SPOT had further added value when added to a baseline model including chest radiograph (Figure 9).
Use of IGRAs in diagnosis of active TB

Figure 9. Receiver operating characteristic (ROC) curves for baseline clinical prediction model with and without addition of IGRA (unpublished)

Abbreviations: AUC = Area under the receiver operating curve, the probability that a randomly selected case will have a higher test value than a randomly selected non-case; a perfect test has an area under the curve of 1.0, while a worthless test has an area of 0.5.

The second study examined a cohort in South Africa to report on the added value of QFT-GIT in risk stratifying 779 individuals (6% active TB prevalence) with HIV-infection prior to initiating isoniazid preventive therapy. A single positive culture was regarded as definitive evidence of active TB. A baseline clinical prediction model (AUC 0.72 (0.65-0.79)) including low weight (< 60 kg), prior history of TB, CD4+ count less than 250 cells/mm³, and any single sign or symptom of active TB was selected using p-values from univariate screening of covariates and stepwise logistic regression. Addition of QFT-GIT to the baseline model produced no significant increase in AUC (0.74 (0.64-0.82); p=0.41) (Figure 10).
Use of IGRAs in diagnosis of active TB

3.1.9 Summary of findings and GRADE evidence profiles

- In low- and middle-income countries, the sensitivity of IGRAs in detecting active TB among persons suspected of having TB ranged from 73-83% and specificity ranged from 49-58%; one in four patients, on average, with culture-confirmed active TB can therefore be expected to be IGRA-negative in low and middle income countries, with serious consequences for patients in terms of morbidity and mortality.
- There was no evidence that IGRAs have added value beyond conventional microbiological tests diagnosis of active TB. Among studies that enrolled active TB suspects (i.e. patients with diagnostic uncertainty), both IGRAs demonstrated suboptimal ‘rule-out’ values for active TB.
- Though data were limited, the sensitivity of both IGRAs was lower among HIV-positive patients (around 60-70%), suggesting that nearly one in three HIV-positive patients with active TB will be IGRA-negative.
- There was no consistent evidence that either IGRA was more sensitive than the TST for active TB diagnosis, although comparisons with pooled estimates of TST sensitivity were difficult to interpret due to substantial heterogeneity.
- The few available head-to-head comparisons between QFT-GIT and T-SPOT demonstrated higher sensitivity for the T-SPOT platform, though this difference did not reach statistical significance.
- The specificity of both IGRAs for active TB was low, regardless of HIV status, and suggested that one in two patients without active TB will be IGRA-positive, with consequences for patients because of unnecessary therapy for TB and a missed differential diagnosis.
- Two unpublished reports reported no incremental or added value of IGRA test results combined with important baseline patient characteristics (e.g. demographics, symptoms, or chest radiograph findings),

Abbreviations: AUC = Area under the receiver operating curve, the probability that a randomly selected case will have a higher test value than a randomly selected non-case; a perfect test has an area under the curve of 1.0, while a worthless test has an area of 0.5; TST = tuberculin skin test; QFT-GIT = QuantiFERON-Gold In-tube.
thus not supporting a meaningful contribution of IGRAs for diagnosis of active TB beyond readily available patient data and conventional tests.

- The systematic review focused on the use of IGRAs to diagnose active pulmonary TB, since data focusing exclusively on extra-pulmonary TB were rare; nevertheless, consensus by the Expert Group was that recommendations for pulmonary TB could reasonably be extrapolated to extra-pulmonary TB.

### 3.1.10 Strengths and limitations of the evidence base

Heterogeneity was substantial for the primary outcomes of sensitivity and specificity. Empirical random effects weighting, excluding studies contributing fewer than 10 eligible individuals and separately synthesizing data for currently manufactured IGRAs were performed in order to minimize heterogeneity.

No standard criteria exist for defining high TB incidence countries and the World Bank income classification is an imperfect surrogate for national TB incidence; nevertheless, results were fundamentally unchanged when restricted to countries with an arbitrarily chosen annual TB incidence of greater than or equal to 50/100,000.

It is likely that ongoing studies were missed despite systematic searching. It is also possible that studies that found poor IGRA performance were less likely to be published. Given the lack of statistical methods to account for publication bias in diagnostic meta-analyses, it would be prudent to assume some degree of overestimation of estimates due to publication bias.

The systematic review focused on test accuracy (i.e. sensitivity and specificity). None of the studies reviewed provided information on patient-important outcomes, such as showing that IGRAs used in a given situation resulted in a clinically relevant improvement in patient care and/or outcomes. In addition, no information was available on the values and preferences of patients.

### 3.1.11 Final recommendations

The GRADE evidence profiles are provided in Tables 3 and 4. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRAS in diagnosis of active TB was **low** and recommended that **these tests should not be used** as a replacement for conventional microbiological diagnosis of pulmonary and extra-pulmonary TB in low-and middle-income countries.

The Expert Group also noted that current evidence did not support the use of IGRAs as part of the diagnostic workup of adults suspected of active TB in low-and middle-income countries, irrespective of HIV status. This recommendation places a high value on avoiding the consequences of unnecessary treatment (high false-positives) given the low specificity of IGRAs in these settings.

<table>
<thead>
<tr>
<th>OVERALL QUALITY OF EVIDENCE</th>
<th>LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRENGTH OF RECOMMENDATION</td>
<td>STRONG</td>
</tr>
</tbody>
</table>
Table 3. GRADE Evidence Profile: Diagnostic accuracy of currently available commercial interferon-gamma release assays (QuantiFERON-TB Gold In-Tube [QFT-GIT], Cellestis, Australia and T-SPOT.TB [T-SPOT], Oxford Immunotec, United Kingdom) for evaluation of patients with pulmonary TB in low- and middle-income countries

<table>
<thead>
<tr>
<th>No of Participants (studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Outcome: Diagnostic accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Positives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19)^A1</td>
<td>Cross-sectional</td>
<td>No Serious Limitation^A2</td>
<td>No Serious Indirectness^A3</td>
<td>Serious^A4 (-1)</td>
<td>Serious^A5 (-1)</td>
<td>Likely^A6</td>
<td>Low ◦ ◦ ◦ ◦</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>True Negatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19)^A1</td>
<td>Cross-sectional</td>
<td>No Serious Limitation^A2</td>
<td>No Serious Indirectness^A3</td>
<td>Serious^A4 (-1)</td>
<td>Serious^A5 (-1)</td>
<td>Likely^A6</td>
<td>Low ◦ ◦ ◦ ◦</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>False Positives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19)^A1</td>
<td>Cross-sectional</td>
<td>No Serious Limitation^A2</td>
<td>No Serious Indirectness^A3</td>
<td>Serious^A4 (-1)</td>
<td>Serious^A5 (-1)</td>
<td>Likely^A6</td>
<td>Low ◦ ◦ ◦ ◦</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>False Negatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19)^A1</td>
<td>Cross-sectional</td>
<td>No Serious Limitation^A2</td>
<td>No Serious Indirectness^A3</td>
<td>Serious^A4 (-1)</td>
<td>Serious^A5 (-1)</td>
<td>Likely^A6</td>
<td>Low ◦ ◦ ◦ ◦</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>B. Outcome: Proportion indeterminate tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2872 (33)^B1</td>
<td>Cross-sectional</td>
<td>Serious^B2 (-1)</td>
<td>No Serious Indirectness^B3</td>
<td>Serious^B4 (-1)</td>
<td>No Serious Imprecision^B5</td>
<td>Likely^B6</td>
<td>Low ◦ ◦ ◦ ◦</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>C. Outcome: Incremental value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high quality observational studies (cross-sectional or cohort studies enrolling patients with diagnostic uncertainty) and at moderate when these types of studies were absent. One point was then subtracted when there was a serious issue identified or two points when there was a very serious issue identified in any of the criteria used to judge the quality of evidence. The evidence rankings were considered to be the same for consideration of true positives, false positives, false negatives, and true negatives.

Footnotes:

1 Sensitivity and specificity were determined exclusively among active TB suspects. 19 studies (11 of QFT-GIT and 8 of T-Spot) were included that assessed the specificity of IGRAs in patients with suspected active TB.

A2 Study limitations were assessed using the QUADAS tool. Three (16%) studies did not enrol a representative spectrum of patients. Five (26%) studies did not clearly report that assessment of the reference standard was performed blinded to IGRA results.

A3 Diagnostic accuracy was considered as a surrogate for patient-important outcomes. No studies measured the impact of IGRAs on patient-important outcomes among TB suspects randomized to treatment based on IGRA results; however, the Expert Group members voted not to downgrade for this factor, in part due to the low likelihood of such studies being undertaken.

A4 Heterogeneity of studies is visually apparent in the Hierarchical Summary Receiver Operating Characteristics (HSROC) Plots.

A5 Pooled sensitivity derived from the highest quality data (studies enrolling active TB suspects) had relatively wide confidence intervals for T-Spot.TB (sensitivity 83% (95% CI 70-91%)) and QFT-GIT (sensitivity 73% (95% CI 61-82%)). Pooled specificity had wide confidence intervals for T-Spot.TB (specificity 58% (95% CI 42-73%)) and acceptable confidence intervals for QFT-GIT (specificity 49% (95% CI 40-58%).

A6 Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests, and publication bias cannot be ruled out. Although points were not deducted, a degree of publication bias is likely because: 1) literature on IGRAs is expanding rapidly; 2) anecdotal examples of unpublished negative studies on IGRAs exist; and 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

B1 33 studies were identified (21 of QFT-GIT and 12 of T-Spot) from which proportions of indeterminate IGRA results could be derived.

B2 Study limitations were assessed using the QUADAS tool. Seventeen (52%) studies did not enrol a representative spectrum of patients.

B3 Please see footnote A3.

B4 Pooled proportions of indeterminate results showed substantial heterogeneity for HIV-uninfected subjects evaluated with QFT-GIT (range 0-27%, I² 78%, p<0.001), and HIV-infected subjects evaluated with both QFT-GIT (range 3-40%, I² 72%, p<0.001) and T-Spot (range 0-25%, I² 88%, p<0.001).

B5 Precision was acceptable for both IGRAs in both HIV-infected (+/-7%) and HIV-uninfected (+/-3%) subjects.

B6 Please see footnote A6.

C1 Two completed but unpublished studies were identified (1 QFT-GIT and TSPOT, 1 QFT-GIT) that used multivariate methods to estimate the added value of IGRAs beyond conventional tests for active TB diagnosis.

C2 As assessed by QUADAS criteria, one (50%) study did not enrol a representative spectrum of patients. Model specification was undertaken for both studies using traditional parametric statistical methods.
Use of IGRAs in diagnosis of active TB

C3 See footnote A1. In addition, area under the receiver-operating-characteristic curve (AUC) may be a less clinically interpretable measure of risk assessment than risk-reclassification statistics.
C4 Only two studies were available; effect estimates for both studies were in the same direction and consistent.
C5 Imprecision, as evaluated by 95% confidence intervals of the area under the receiver-operating-characteristic curves (AUC), was reasonable for both studies.
C6 Because of the relative novelty of these methods, at this time it is unlikely that studies of IGRA incremental value have been unpublished due to publication bias.
**Table 4. GRADE summary of findings – Role of IGRAs for evaluation of patients with pulmonary TB in low- and middle-income countries**

**Review question:** What is the diagnostic accuracy of commercial IGRAs for pulmonary tuberculosis?

**Patients/population:** Adult pulmonary TB suspects and confirmed TB cases in low- and middle-income countries

**Setting:** Outpatients and inpatients

**Index test:** Commercial interferon-gamma release assays (QuantiFERON-TB Gold In-Tube [QFT-GIT], Cellestis, Australia and T-SPOT.TB [T-SPOT], Oxford Immunotec, United Kingdom)

**Importance:** Rapid, accurate, simple test could supplement conventional microbiology and expand testing to peripheral health centers

**Reference standard:** Microbiologic (culture or smear-microscopy) or clinical diagnosis of pulmonary TB

**Studies:** Cross-sectional or cohort

<table>
<thead>
<tr>
<th>Outcomes: TP, TN, FP, FN</th>
<th>Effect % (95% CI)</th>
<th>No. of participants (studies)</th>
<th>What do these results mean given 10% prevalence among suspects being screened for TB?</th>
<th>What do these results mean given 30% prevalence among suspects being screened for TB?</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subgroups</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T-SPOT.TB, HIV-infected</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity 78% (56, 91)</td>
<td>549 (5)</td>
<td>With a prevalence of 10%, 100/1000 will have TB. Of these, 78 (TP) will be identified; 22 (FN) will be missed by T-SPOT.TB. Of the 900 patients without TB, 495 (TN) will not be treated; 405 (FP) will be unnecessarily treated.</td>
<td>With a prevalence of 30%, 300/1000 will have TB. Of these, 234 (TP) will be identified; 66 (FN) will be missed by T-SPOT.TB. Of the 700 patients without TB, 385 (TN) will not be treated; 315 (FP) will be unnecessarily treated.</td>
<td>Low @@@@</td>
<td></td>
</tr>
<tr>
<td>Specificity 55% (45, 64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T-SPOT.TB, HIV-uninfected</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insufficient data for pooled estimates</td>
<td>364 (3)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>QuantiFERON-TB Gold In-Tube, HIV-infected</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity 62% (41,79)</td>
<td>469 (6)</td>
<td>With a prevalence of 10%, 100/1000 will have TB. Of these, 62 (TP) will be identified; 38 (FN) will be missed by QFT-GIT. Of</td>
<td>With a prevalence of 30%, 300/1000 will have TB. Of these, 186 (TP) will be identified; 114 (FN) will be missed by QFT-GIT. Of the</td>
<td>Low @@@@</td>
<td></td>
</tr>
</tbody>
</table>
Use of IGRAs in diagnosis of active TB

**Outcome** | **Subgroup** | **Effect % (95% CI)** | **No. of participants (studies)** | **What do these findings mean?** | **Quality of Evidence**
--- | --- | --- | --- | --- | ---
IGRA-TST sensitivity difference* | QuantiFERON-TB Gold In-Tube | 1% (-11 to 13%)* | 475 (10) | This evidence suggests that QFT-GIT is no more sensitive than TST for active TB diagnosis in low- and middle-income countries. | Low ⊗⊗⊗
| T-SPOT.TB | 9% (-10% to 28%)* | 206 (5) | This evidence suggests that TSPOT is slightly more sensitive than TST for active TB diagnosis in low- and middle-income countries. This evidence should be interpreted with caution given the low number of studies available. | Low ⊗⊗⊗
Proportion indeterminate tests | QuantiFERON-TB Gold In-Tube, HIV-uninfected Subjects | 4% (1-7%) | 1603 (11) | This evidence suggests that among HIV-uninfected subjects, the proportion of indeterminate QFT-GIT test results in low- and middle-income countries will be low and similar to high-income countries. | Low ⊗⊗⊗
| T-SPOT.TB, HIV-uninfected Subjects | 3% (1-4%) | 494 (5) | This evidence suggests that among HIV-uninfected subjects, the | Low ⊗⊗⊗

---

**QuantiFERON-TB Gold In-Tube, HIV-uninfected**

Sensitivity 82% (76, 87)  
Specificity 42% (33, 53)  
1304 (5)  
With a prevalence of 10%, 100/1000 will have TB. Of these, 82 (TP) will be identified; 18 (FN) will be missed by QFT-GIT. Of the 900 patients without TB, 378 (TN) will not be treated; 522 (FP) will be unnecessarily treated.

700 patients without TB, 357 (TN) will not be treated; 343 (FP) will be unnecessarily treated.

With a prevalence of 30%, 300/1000 will have TB. Of these, 246 (TP) will be identified; 54 (FN) will be missed by QFT-GIT. Of the 700 patients without TB, 294 (TN) will not be treated; 406 (FP) will be unnecessarily treated.

---

**Outcome** | **Effect % (95% CI)** | **No. of participants (studies)** | **What do these findings mean?** | **Quality of Evidence**
--- | --- | --- | --- | ---
IGRA-TST sensitivity difference* | QuantiFERON-TB Gold In-Tube | 1% (-11 to 13%)* | 475 (10) | This evidence suggests that QFT-GIT is no more sensitive than TST for active TB diagnosis in low- and middle-income countries. | Low ⊗⊗⊗
| T-SPOT.TB | 9% (-10% to 28%)* | 206 (5) | This evidence suggests that TSPOT is slightly more sensitive than TST for active TB diagnosis in low- and middle-income countries. This evidence should be interpreted with caution given the low number of studies available. | Low ⊗⊗⊗
Proportion indeterminate tests | QuantiFERON-TB Gold In-Tube, HIV-uninfected Subjects | 4% (1-7%) | 1603 (11) | This evidence suggests that among HIV-uninfected subjects, the proportion of indeterminate QFT-GIT test results in low- and middle-income countries will be low and similar to high-income countries. | Low ⊗⊗⊗
| T-SPOT.TB, HIV-uninfected Subjects | 3% (1-4%) | 494 (5) | This evidence suggests that among HIV-uninfected subjects, the | Low ⊗⊗⊗

---

**What do these findings mean?**

**Quality of Evidence**

- Low ⊗⊗⊗
- Medium ⊗⊗⊙
- High ⊗⊙⊙
- Very High ⊙⊙⊙

---

**IGRA-TST sensitivity difference**

**Subgroup** | **Effect % (95% CI)** | **No. of participants (studies)** | **What do these findings mean?** | **Quality of Evidence**
--- | --- | --- | --- | ---
QuantiFERON-TB Gold In-Tube | 1% (-11 to 13%)* | 475 (10) | This evidence suggests that QFT-GIT is no more sensitive than TST for active TB diagnosis in low- and middle-income countries. | Low ⊗⊗⊗
T-SPOT.TB | 9% (-10% to 28%)* | 206 (5) | This evidence suggests that TSPOT is slightly more sensitive than TST for active TB diagnosis in low- and middle-income countries. This evidence should be interpreted with caution given the low number of studies available. | Low ⊗⊗⊗
Proportion indeterminate tests | QuantiFERON-TB Gold In-Tube, HIV-uninfected Subjects | 4% (1-7%) | 1603 (11) | This evidence suggests that among HIV-uninfected subjects, the proportion of indeterminate QFT-GIT test results in low- and middle-income countries will be low and similar to high-income countries. | Low ⊗⊗⊗
| T-SPOT.TB, HIV-uninfected Subjects | 3% (1-4%) | 494 (5) | This evidence suggests that among HIV-uninfected subjects, the | Low ⊗⊗⊗
Use of IGRAs in diagnosis of active TB

<table>
<thead>
<tr>
<th>Study</th>
<th>Proportion of Indeterminate Results</th>
<th>Low (1-10%)</th>
<th>Medium (11-25%)</th>
<th>High (26-100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QuantiFERON-TB Gold In-Tube, HIV-infected Subjects</td>
<td>16% (10-21%)</td>
<td>Low 4-10%</td>
<td>Medium 11-25%</td>
<td>High 26-100%</td>
</tr>
<tr>
<td>T-SPOT.TB, HIV-infected Subjects</td>
<td>8% (1-15%)</td>
<td>Low 4-10%</td>
<td>Medium 11-25%</td>
<td>High 26-100%</td>
</tr>
</tbody>
</table>

**Incremental value**

Neither study demonstrated significant added value over conventional tests for active TB diagnosis, as measured by change in the area under receiver operating curve (AUC).

<table>
<thead>
<tr>
<th>Study</th>
<th>Effect</th>
<th>Low (1-10%)</th>
<th>Medium (11-25%)</th>
<th>High (26-100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low 4-10%</td>
<td>Medium 11-25%</td>
<td>High 26-100%</td>
</tr>
</tbody>
</table>

Note: Incremental value is IGRA minus TST.

*Value is IGRA minus TST.*
3.2 Use of IGRAs in children

3.2.1 Objectives, reference standards and outcomes

Objective 1: Assessment of IGRAs for the diagnosis of LTBI

Reference standards

a. Primary reference standard: Incident active TB disease in groups defined on the basis of results of IGRA tests and followed in prospective longitudinal studies

b. Secondary reference standard: TB exposure expressed as
   - Dichotomous exposure (e.g. exposed/unexposed)
   - Exposure gradient of three or four categories, based on microbiologic indicators (smear status) of the index case, or proximity to the index case or time of exposure to the index case.

In these studies, the groups had to be mutually exclusive as described by the study-specific definition of exposure, but not on other important characteristics nor on the basis of prior exposure.

Objective 2: Assessment of IGRAs for the diagnosis of active TB disease in children

Reference standards

Reference standards for TB disease:

a. Definite (Confirmed) TB disease
   Confirmed bacteriologic disease defined as the presence of at least 1 clinical specimen positive for M. tuberculosis on culture or AFB smear microscopy positive, or 1 histology sample positive for necrotising granulomas or nucleic acid amplification test positive for M. tuberculosis, with or without positive Ziehl-Neelsen stain. (Definition of disease must not include TST or IGRAs)

b. Probable TB Disease
   Clinical evidence may include 1) chest radiologic findings consistent with active TB (ideally confirmed by experienced reviewers), 2) typical symptoms, 3) other radiological evidence of TB, including extrapulmonary TB (eg. computed brain tomography with classical appearance of TB meningitis) in conjunction with symptoms 4) exposure to TB, 5) response to appropriate full anti-TB therapy and 6) must not include TST or IGRA as part of the case definition.

Possible TB was not accepted as a reference standard for active TB.

Reference standards for specificity (‘No TB’ group)

a. TB suspects with symptoms suggestive of active TB, where active TB was excluded;

b. Mixed study groups of children with either suspected active TB or TB contact, where active TB was excluded by the same investigations that led to the diagnosis of active TB in cases (mainly clinical and radiological investigations, mycobacterial culture was not necessarily required)

The following were not accepted as non-diseased reference standards:

- Healthy children without symptoms or risk factors for TB;
- Mixed groups of children with symptoms or TB contact or immigrants without other risk factors, where it was unclear if active TB was systematically excluded in all;
Objective 3: To review and summarize user-important characteristics of IGRAs, such as impact on test performance of BCG vaccination status, age, HIV status and TB burden of the setting where these tests have been used.

Subgroup analysis was performed to assess performance for the diagnosis of active TB across the following strata of interest:

- TST cut-offs at 5, 10, 15 mm;
- QFT-Gold versus QFT-Gold IT;
- Country where study conducted – dichotomized on the basis of World Bank income index into: low, lower and upper-middle income, and high income countries;
- Country where study conducted – dichotomized on the basis of WHO incidence of smear positive TB cases (2007), stratified by incidence below or above 25/100,000;
- Age of study subjects: stratified analysis using a cut-off of 5 years (mean age, median age if the mean age was not provided);
- BCG status of study subjects: stratified analysis, dichotomised at 50% BCG coverage;
- HIV infection dichotomised at 100% HIV-infected in the study population versus < 15% or ‘not reported’.

Objective 4: To assess operational aspects of IGRAs such as cost, feasibility in children and other aspects

Outcomes

Data were extracted on the number of positive, negative as well as indeterminate results (if provided) for each of the reference standards and for each test assessed (IGRAs and TST with different cut-offs, i.e. 5, 10 and 15 mm).

Outcomes were given a hierarchy ranking, depending on their importance for patients receiving the test. This hierarchy was used in the GRADE framework, when assessing the quality of evidence provided for each outcome.

For objective 1, the risk of progression to active TB as well as correlation of IGRAs with different gradients of exposure was ranked as critical. These outcomes, if favorable, will directly have an impact on the management of child TB contacts and children infected.

For objective 2, sensitivity and specificity of the assays in TB suspects as a surrogate for patient-important outcomes was rated as critical.

For objective 2, analysis was also completed to assess the impact of indeterminate results on IGRA test performance. Sensitivity and specificity estimates were calculated with indeterminate results considered as false negatives. Additional analysis was completed to assess the association between indeterminate IGRA results and specific factors including age, TB burden of the study setting, HIV co-infection and BCG vaccination rates in the study population.

Other important outcomes, relevant to all review questions, were the performance of the tests in high-risk groups such as HIV-infected and young children under five years of age as well as the rate of indeterminate test results.
3.2.2 Data analysis

Data analysis was performed using SAS version 9.2 (SAS Institute, Inc., Cary, NC) and STATA version 11 (Stata Corporation, College Station, Texas, USA).

Three independent analyses were conducted using three independent types of outcome measures to assess test performance. As described previously, the three types of outcome measures considered were incident TB and exposure (Objective 1), and prevalent TB (Objective 2). Since only two longitudinal studies were identified that estimated incidence of active TB, with very different methodologies, these studies were simply summarized in descriptive analysis.

Given the small number of pediatric studies available, particularly from LMIC, all analyses were done for LMIC and HIC combined but also separately for LMIC. GRADE assessment was limited to LMIC data to ensure consistency with the scope of the Expert Group meeting.

Objective #1 (to evaluate the accuracy and performance of IGRAs for the detection of LTBI) was addressed using two approaches: studies that measured test performance in children with and without exposure (i.e. dichotomous exposure) were analyzed collectively whilst studies that measured test performance in children with varying degrees of exposure (ie. a gradient of exposure) were analyzed collectively.

For the analysis of dichotomous exposure, the association between test result and TB exposure was assessed as an odds ratio. Then, results from all studies in this category were pooled together using both a fixed and a random effects approach.

For the analysis of categorical exposure (more than two categories), the dose-response relationship was assessed between test result (TST, QFT, or T-SPOT) and TB exposure. For the analyses of correlation between exposure gradients and prevalence of positive tests, the exposure was categorized in 3-4 categories, depending on the study. Initially the Spearman correlation between the categorical test result and the outcome was estimated, followed by a pooled correlation coefficient for each test with both a fixed and a random effects approach.

In a second approach, the odds ratio was used as the measure of effect to assess performance in studies that described TB exposure as a gradient. First, the odds ratio was calculated for each level of exposure relative to the reference group, and then an overall odds ratio was calculated for increasing exposure category. Both fixed and random effects approaches were used to estimate a pooled exposure effect across the studies, accounting for the correlation between the category-specific odds ratios estimated for each study (because they use a common reference category). Inter-study heterogeneity was also estimated via I-squared statistics.

To address objective #2 (to evaluate the accuracy and performance of IGRAs for the diagnosis of active TB), studies that measured test performance in children with active disease compared to children without active disease were analysed collectively. All studies included in this analysis provided data to estimate test sensitivity (proportion of positive test results among patients with definite and/or probable TB disease). Indeterminate test results were excluded from the primary estimate of sensitivity; sensitivity was re-estimated after adding indeterminates as false negative test results. For studies that included an appropriate group in whom disease was excluded using methods that met the pre-defined review criteria, the specificity (proportion of negative test results among patients without TB disease) was calculated. Because many of these studies did not have
non-diseased populations that were judged acceptable, only a portion of studies provided data for the estimation of test specificity.

A random effects meta-analysis was used to estimate the overall pooled estimates of sensitivity and specificity and 95% CI. The exact binomial likelihood approach was used, which uses a binomial distribution to approximate the distribution of the outcome of interest. This approach accounts for study size, and includes a random effect to account for inter-study heterogeneity. When proportions are the outcome measure, this approach has been demonstrated to produce less biased estimates of the pooled effect and the between-study variability. Heterogeneity was assessed of proportions of subjects with outcomes of interest, within sub-groups defined by covariates of interest, by estimating the I-squared statistic and associated 95% CIs. To calculate the I-square zero cells were corrected by 0.5.

Significant heterogeneity was expected due to significant variation in study designs and outcome measures; therefore, when there were sufficient studies, sub-group analyses were performed stratified by predefined covariates of interest, including World Bank economic status of the study setting (LMIC versus HIC), TB burden of the study setting, age, BCG vaccination, HIV infection, TST cut-point, and generation of QFT used. Descriptive statistics were used to examine the relationship between the frequency of IGRA indeterminate results and co-variates of interest.

Forest plots were generated to display sensitivity and specificity estimates for each study. Data used to generate forest plots was derived from analysis using all TB disease definitions including definitive and probable TB. The forest plots display the sensitivity and specificity estimates obtained for each study and pooled estimates for all studies and for studies stratified by World Bank income category (HIC vs LMIC). STATA version 11 was used to create the forest plot figures and display SAS generated data as described above.

### 3.2.3 Search results

240 titles were identified by the electronic database search and from searching reference lists of selected articles, additional existing databases, review articles and systematic reviews. From these, 68 studies were selected for full text review, of which 36 articles were excluded and 32 articles included for data analysis (Annex 3).

### 3.2.4 Study characteristics

One of the 32 articles included reported data from two settings and was counted as two separate studies, i.e. 33 studies.

Five studies were case-control studies and the remaining 28 were cross-sectional studies; 21 studies were performed in inpatient and/or outpatient settings, 9 studies were household contact studies or outbreak investigations, and the setting was not clearly defined in 3 studies.

The studies were performed in 18 different countries, of which nine (50%) were high-income (19 studies), four upper-middle (8 studies), two lower-middle (3 studies), and three low-income countries (3 studies). 42% (14/33) of the studies were performed in LMIC. The incidence of smear positive TB was <25/100 000 in eight of these countries in 2007, and >25/100 000 in the remaining ten countries. Studies performed in high-income countries included between 11% and 100% immigrant children from countries with higher burdens of TB.
A total of 5,922 children were assessed in the 33 studies but results from only 4,505 of these children were analysed (1,930 children from LMIC), as some sub-groups of children were excluded for not meeting the criteria of any reference standard. The mean (or median, if mean was not given) age of children included ranged from 1.9 to 14.6 years with an average of 7.6 years. Overall BCG coverage ranged between 8-100%, and between 77-100% in LMIC.

Nineteen studies reported on HIV-infection status; of these, only seven studies included any HIV-infected children (0.5-100%). Two studies, both from LMIC, were performed in HIV-infected children only.

Eleven studies reported on immune-suppression and of these, two studies from LMIC included children with immune-suppression (oncology patients, malnourished children), and one study reported on immune-suppression in HIV-infected children (CD4 count <200) in an upper-middle income country.

Reference standards

Authors of 21 studies were contacted for clarification of reported data or to request revised data that met the criteria for the reference standards. Only two studies reported data on incident TB; 18 studies reported data on exposure (10 on dichotomous exposure, 8 on exposure gradients), and 21 studies reported data on prevalent TB. Nine studies met the criteria for assessing specificity of IGRAs in TB disease as they included a well-defined group in whom TB had been excluded as specified in the reference standards.

Index tests used

Studies evaluated one or more index tests including T-SPOT (15 studies), QFT-G (10 studies) and QFT-GIT (21 studies). In three of the QFT-GIT studies, an early version of QFT-GIT kits was used that did not contain a mitogen-control tube. TST data was provided in 32 studies but only assessed in 30 due to incorporation bias or interpretation issues in two studies.

3.2.5 Study quality

Thirteen QUADAS quality items for the assessment of quality of diagnostic studies were used and an additional item on industry involvement added. QUADAS results are shown in Figures 11 and 12.
Results for relevant QUADAS items show that only 3/21 (14%) of active TB studies and 8/18 (45%) of exposure studies clearly reported on the sampling methods (consecutive, random or convenient) and representativeness of their patient spectrum. Blinding of clinicians to IGRA results was reported in only 6/21 (29%) of studies assessing active TB.

For studies assessing active TB, it remained unclear in 9/21 (43%) whether differential verification was avoided (whether all children received the reference standard). The execution of the reference standard (definition of active TB) was described in 17/21 (81%) of studies. However, there was still a wide variation amongst studies regarding the criteria used for the definition of confirmed or probable TB, and how detailed these disease categories were described.

Eleven of the 33 studies (33%) were supported by either or both manufacturers of commercial IGRAs, mainly through donation of test kits.

3.2.6 Test failure and indeterminate results across studies and populations

Test failure of IGRAs was defined as technical errors, failed phlebotomy or an insufficient amount of peripheral blood mononuclear cells (in the case of T-SPOT) to perform the assay. For TST, failure was
defined as patients not returning for test reading. Failure, especially for TST, was only infrequently reported. Reported failure rates ranged from 0-7% for QFT-GIT (19 studies), 0% for QFT-G (6 studies), 0-21% for T-SPOT (15 studies) and 0-11% for TST (20 studies).

Average indeterminate results reported across all studies (and including study populations that were not included in the remaining analysis) were 6.5% for QFT-GIT, 6.4% for QFT-G and 3.5% for T-SPOT with a wide range among individual studies. In individual studies, high indeterminate rates >10% were associated with young age (<4 years), co-morbid infections, particularly helminth infection, and immune-suppression, eg. in cancer patients or HIV-infected children; in some studies no explanation was found.

3.2.7 Studies assessing incident TB

Only two included studies longitudinally assessed development of TB, both were performed in HIC and results are therefore not reported here.

3.2.8 Studies assessing exposure

Dichotomous TB exposure

Table 5 describes the characteristics of the ‘exposed’ and ‘unexposed’ groups in the 10 studies included (including four studies from LMIC). The definition of the exposure groups varied highly between studies, either depending on the study setting or the inclusion criteria for each group.

Table 5: Studies with dichotomous expression of exposure included in analysis of test performance for LTBI, sorted by World Bank income index

<table>
<thead>
<tr>
<th>LMIC</th>
<th>Author, year</th>
<th>Description “Exposed”</th>
<th>Description “Unexposed”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hansted 2009</td>
<td>HH or school contact</td>
<td>No TB contact, no symptoms, chest radiography normal</td>
<td></td>
</tr>
<tr>
<td>Hesseling 2009</td>
<td>Known TB contact</td>
<td>No known TB contact</td>
<td></td>
</tr>
<tr>
<td>Mandalakas 2008</td>
<td>Known HH contact</td>
<td>No known HH contact</td>
<td></td>
</tr>
<tr>
<td>Stefan 2010</td>
<td>Known TB contact</td>
<td>No known TB contact</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIC</th>
<th>Author, year</th>
<th>Description “Exposed”</th>
<th>Description “Unexposed”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bianchi 2009</td>
<td>TB contacts, Italian and Immigrant children</td>
<td>Immigrant children without TB contact</td>
<td></td>
</tr>
<tr>
<td>Chun 2008</td>
<td>HH contacts = close contacts</td>
<td>Contact outside HH = casual contacts</td>
<td></td>
</tr>
<tr>
<td>Dominguez 2008</td>
<td>Children from contact investigations</td>
<td>TST + children detected during routine screening (school or pediatrician)*</td>
<td></td>
</tr>
<tr>
<td>Higuchi 2009</td>
<td>Same class as index case (contact ≥ 90 hours)</td>
<td>Same school, different classes as index case (contact &lt; 18 hours)</td>
<td></td>
</tr>
<tr>
<td>Lighter 2009</td>
<td>Close contact to TB index case</td>
<td>No risk factors for TB exposure</td>
<td></td>
</tr>
<tr>
<td>Lucas 2010</td>
<td>Immigrants with HH contact</td>
<td>Immigrants without HH contact</td>
<td></td>
</tr>
</tbody>
</table>

* TST was not used for analysis due to incorporation bias

For TST, 10mm was used as the cut-point to define a positive test, except in a few studies which only reported results of a cut-point of 5mm (2 studies), or 15mm (1 study). One study used a composite cut off of either 6 or 15 mm depending on risk factors. The overall pooled odds ratio for HIC and LMIC combined was 1.34 (95% CI: 0.66-2.72), 1.93 (95% CI: 0.98-3.77), and 1.83 (95% CI: 0.67-5.02) for TST when 5mm, 10mm, or 15 mm was used as the cut-off. For QuantiFERON, the pooled OR was 3.51 (95% CI: 1.85-6.66), and for T-SPOT it was 1.31 (95% CI: 0.76-2.27). All confidence intervals
overlapped, ie. the performance of all assays were similar. was all assays perthere was no statistically significant difference.

The odds remained positive for a positive QFT (1.30, 95% CI: 0.20-8.32) or T-SPOT (2.24, 95% CI: 0.88-5.64) result in exposed versus unexposed individuals when analysing data from LMIC separately, but was less clear for TST with a cut off at 5, 10 or 15 mm (1.04, 95% CI: 0.46-2.36; 0.81, 95% CI 0.38-1.74 and 0.48, respectively). However, the number of studies was small. As for the HIC/LMIC combined analysis, confidence intervals were overlapping for all three assays, ie. the performance was similar (Figure 13).

![Figure 13. Pooled odds ratios (95% CIs): Concordance of test with dichotomous exposure](image)

In the legend, the number of studies included for each test is indicated in parentheses.

**Exposure gradients**

Eight studies allowed analysis of the correlation between different grades of exposure (0=least to 3=most exposure) and test results; five of these were performed in LMIC (Table 6). The definitions of the different grades of exposure varied between these studies, as some studies measured smear status of an index case, while others measured duration of exposure and proximity to the index case.

The lowest level of exposure was assigned to grade 0 in the dataset, the next level to grade 1 etc. As a result, even studies using the same measure of exposure, e.g. sputum smear status of the index cases, did not have the same grades in each level; in some studies contact with index cases with smear-negative TB was the lowest level of exposure and in other studies the least exposed had contact with smear-positive TB.
Table 6. Studies with exposure gradients included in analysis of test performance for LTBI, sorted by World Bank income index

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Description of assignment to exposure gradients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LMIC</strong></td>
<td></td>
</tr>
<tr>
<td>Adetifa 2010</td>
<td>Grade 0: Different House</td>
</tr>
<tr>
<td></td>
<td>Grade 1: Different room</td>
</tr>
<tr>
<td></td>
<td>Grade 2: Same room</td>
</tr>
<tr>
<td></td>
<td>Grade 3: -</td>
</tr>
<tr>
<td>Nakaoka 2006</td>
<td>Grade 0: Community controls</td>
</tr>
<tr>
<td></td>
<td>Grade 1: Exposure to smear - TB</td>
</tr>
<tr>
<td></td>
<td>Grade 2: Exposure to smear + TB</td>
</tr>
<tr>
<td></td>
<td>Grade 3: -</td>
</tr>
<tr>
<td>Okada 2008</td>
<td>Grade 0: Exposure to smear - TB</td>
</tr>
<tr>
<td></td>
<td>Grade 1: Exposure to smear + TB</td>
</tr>
<tr>
<td></td>
<td>Grade 2: Exposure to smear ++ TB</td>
</tr>
<tr>
<td></td>
<td>Grade 3: Exposure to smear +++ TB</td>
</tr>
<tr>
<td>Petrucci 2008, Brazil</td>
<td>Grade 0: Exposure to scanty TB</td>
</tr>
<tr>
<td></td>
<td>Grade 1: Exposure to smear + TB</td>
</tr>
<tr>
<td></td>
<td>Grade 2: Exposure to smear ++ TB</td>
</tr>
<tr>
<td></td>
<td>Grade 3: Exposure to smear +++ TB</td>
</tr>
<tr>
<td>Petrucci 2008, Nepal</td>
<td>Grade 0: Exposure to scanty TB</td>
</tr>
<tr>
<td></td>
<td>Grade 1: Exposure to smear + TB</td>
</tr>
<tr>
<td></td>
<td>Grade 2: Exposure to smear ++ TB</td>
</tr>
<tr>
<td></td>
<td>Grade 3: Exposure to smear +++ TB</td>
</tr>
<tr>
<td><strong>HIC</strong></td>
<td></td>
</tr>
<tr>
<td>Bergamini 2009</td>
<td>Grade 0: Exposed to probable TB</td>
</tr>
<tr>
<td></td>
<td>Grade 1: Exposed to smear -/culture + TB</td>
</tr>
<tr>
<td></td>
<td>Grade 2: Exposed to smear + TB</td>
</tr>
<tr>
<td></td>
<td>Grade 3: -</td>
</tr>
<tr>
<td>Diel 2008</td>
<td>Grade 0: 40-59 hrs exposure</td>
</tr>
<tr>
<td></td>
<td>Grade 1: 60 -99 hrs exposure</td>
</tr>
<tr>
<td></td>
<td>Grade 2: 100 -199 hrs exposure</td>
</tr>
<tr>
<td></td>
<td>Grade 3: ≥ 200 hrs exposure</td>
</tr>
<tr>
<td>Girardi 2007</td>
<td>Grade 0: Low risk, other students</td>
</tr>
<tr>
<td></td>
<td>Grade 1: Intermediate risk, sharing some activities with index case</td>
</tr>
<tr>
<td></td>
<td>Grade 2: High risk, attending same class with index case</td>
</tr>
<tr>
<td></td>
<td>Grade 3: -</td>
</tr>
</tbody>
</table>

In LMIC, the pooled correlation coefficient, calculated using a random effects model, was 0.28 (95% CI: 0.06-0.86, I-squared=0.90) for the association between QuantiFERON result (QFT-G and QFT-GIT combined) and the gradient of TB exposure. For T-SPOT, the pooled correlation coefficient using a fixed model (only one study included) was 0.15 (95% CI: 0.02-0.37). For TST, the pooled correlation coefficient (random effects) was 0.19 (95% CI: 0.02-0.61, I-squared=0.80), 0.22 (95% CI: 0.11-0.39, I-squared=0.65) when 5 mm and 10 mm were used as the cut-off respectively (no studies used a 15mm cut-off). Confidence intervals overlapped for all three assays, i.e. the performance was similar (Figure 14). A similar trend was seen if analysis was performed for HIC and LMIC combined (data not shown).
The slopes of the effect (exposure gradient) were then estimated by regressing the logs of the odds ratio between each successive higher exposure (Figures 15). With this analytic method, a steeper slope is associated with a greater change in odds across exposure categories. Hence, the steeper the slope the better the test is able to distinguish infection across exposure categories. Using either fixed or random effects models, the slopes estimated for QFT-G and QFT-GIT combined, T-SPOT and TST employing a 10 mm cut-point were very similar, both in LMIC as well as HIC and LMIC combined. Estimates were associated with wide and overlapping confidence intervals and as a result no test can be declared superior.

There was no data for TST 15 mm cut-off in LMIC. In the legend, the number of studies included for each test is indicated in parentheses.
3.2.9 Studies assessing active TB

Test performance was assessed in active child TB suspects and cases (sensitivity), and children categorized as having ‘no TB’ (specificity) according to the predefined reference standards used for the systematic review. The exact definition for each group of TB suspects/patients varied between studies and for the ‘no TB’ categories. Because many of these studies did not have non-diseased populations that were consistent with criteria used in the systematic review, only a portion of studies provided data for the estimation of test specificity (Table 7).
Table 7. Studies included in analysis of test performance in TB disease, sorted by World Bank income index

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Active TB: disease categories*</th>
<th>Comparison group: description of “no TB”***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dogra 2007</td>
<td>Definite/probable TB combined</td>
<td>Hospitalized children with clinical suspicion of TB or TB contact, TB disease was ruled out</td>
</tr>
<tr>
<td>Warier 2009</td>
<td>Definite TB, Probable TB</td>
<td>Hospitalized children with other diagnosis, no TB contact</td>
</tr>
<tr>
<td>Hansted 2009</td>
<td>Definite TB</td>
<td>Reported group not used***</td>
</tr>
<tr>
<td>Moyo, unpublished</td>
<td>Definite/probable TB combined</td>
<td>Children with clinically suspected TB or HH contact, active TB ruled out</td>
</tr>
<tr>
<td>Nicol 2009</td>
<td>Definite/probable TB combined</td>
<td>Children admitted for either clinically suspected TB or TB contact, active TB ruled out by chest radiography and culture.</td>
</tr>
<tr>
<td>Stavri 2009</td>
<td>Definite TB</td>
<td>No group reported****</td>
</tr>
<tr>
<td>Bamford 2009</td>
<td>Definite TB, probable TB</td>
<td>No group reported</td>
</tr>
<tr>
<td>Bergamini 2009</td>
<td>Definite/probable TB combined</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Bianchi 2009</td>
<td>Definite/probable TB combined</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Chun 2008</td>
<td>Probable TB</td>
<td>TB ruled out, other diagnosis</td>
</tr>
<tr>
<td>Connell 2006</td>
<td>Definite/probable TB combined</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Connell 2008</td>
<td>Probable TB</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Detjen 2007</td>
<td>Definite TB</td>
<td>Children with other respiratory illness, low risk for TB</td>
</tr>
<tr>
<td>Dominguez 2008</td>
<td>Definite/probable TB combined</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Grare 2010</td>
<td>Definite/probable TB combined</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Haustein 2009</td>
<td>Definite TB, probable TB</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Herrmann 2009</td>
<td>Definite/probable TB combined</td>
<td>Children hospitalized for any other disease, no TB contact</td>
</tr>
<tr>
<td>Higuchi 2008</td>
<td>Definite/probable TB combined (detected in school outbreak)</td>
<td>School outbreak investigation, active TB excluded by chest radiography</td>
</tr>
<tr>
<td>Higuchi 2009</td>
<td>Definite/probable TB combined</td>
<td>Reported group not used</td>
</tr>
<tr>
<td>Kampmann 2009</td>
<td>Definite TB, probable TB</td>
<td>Children with risk factors for TB but disease ruled out, other diagnosis made</td>
</tr>
<tr>
<td>Lighter 2009</td>
<td>Definite/probable TB combined</td>
<td>Reported group not used</td>
</tr>
</tbody>
</table>

* Definite TB = Culture confirmed disease, Probable TB = Diagnosis made on the basis of symptoms and radiologic findings, no culture result. Definite/probable combined = some cases confirmed but others diagnosed on clinical and radiological criteria only, and results not stratified by method of diagnosis.

** No TB = TB suspects with symptoms suggestive of active TB, or TB contacts where active TB was excluded.

*** Reported control group did not meet review criteria for an appropriate control group

**** Study assessed active TB group only.

Sensitivity and specificity was calculated for each test. For sensitivity calculations, different disease categories were considered separately: some studies used an outcome of 1) bacteriologically
confirmed (definite) TB while others used a combined outcome of 2) definite and probable (clinically and radiologically confirmed) TB combined. These were analyzed separately, and then results from both types of studies were combined (Table 8 A-C). Analysis counting indeterminate results as negative was also performed for both IGRAs.

For each category, stratified analysis for test type (TST with different cut-offs and QFT-GIT versus QFT-G), World Bank income, TB burden based on 2007 WHO TB smear incidence data, BCG vaccination status, age and HIV infection was performed, excluding indeterminate results. Confidence intervals for almost all analyses were wide and overlapping, i.e. there were no significant differences between tests or sub-strata (Table 7). Stratified analysis for LMIC alone provided very small numbers in the subgroups and did not allow further analysis.

Forest plots for study-specific and pooled estimates of sensitivity and specificity by WB income index are shown in (Figures 16 - 18). There was a significant amount of heterogeneity among studies; therefore the results should be interpreted with caution.

Indeterminate results in QFT and T-SPOT lowered the sensitivity of both assays slightly when these were added to the ‘false-negative’ results, with overlapping confidence intervals (Figure 7). IGRA indeterminate results were not significantly correlated with any factors including age, HIV status, TB burden of the setting or BCG vaccination. In addition, there was no difference in the frequency of IGRA indeterminate results in stratified analysis of these same factors.

Table 8. Diagnostic accuracy of TST, QFT and T-SPOT in active disease in children, including all acceptable definitions of TB and all countries (HIC and LMIC)

<table>
<thead>
<tr>
<th>TST*</th>
<th>True positives (sensitivity)</th>
<th>False positives (specificity)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N studies</td>
<td>N Pos/Tested</td>
</tr>
<tr>
<td>Overall</td>
<td>19</td>
<td>386/575</td>
</tr>
<tr>
<td>By TST size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 mm</td>
<td>13</td>
<td>220/265</td>
</tr>
<tr>
<td>10 mm</td>
<td>16</td>
<td>261/362</td>
</tr>
<tr>
<td>15 mm</td>
<td>11</td>
<td>246/389</td>
</tr>
<tr>
<td>By World Bank category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIC</td>
<td>14</td>
<td>299/407</td>
</tr>
<tr>
<td>LMIC</td>
<td>5</td>
<td>87/168</td>
</tr>
<tr>
<td>By TB incidence/100,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>13</td>
<td>296/402</td>
</tr>
<tr>
<td>&gt;25</td>
<td>6</td>
<td>90/173</td>
</tr>
<tr>
<td>By BCG vaccination coverage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50%</td>
<td>6</td>
<td>57/66</td>
</tr>
<tr>
<td>&gt; 50%</td>
<td>13</td>
<td>329/509</td>
</tr>
<tr>
<td>By age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5 yrs</td>
<td>6</td>
<td>90/154</td>
</tr>
<tr>
<td>≥ 5 yrs</td>
<td>13</td>
<td>296/421</td>
</tr>
</tbody>
</table>
### Use of IGRAs in children

#### By HIV status**

<table>
<thead>
<tr>
<th>HIV status</th>
<th>N studies</th>
<th>N Pos/Tested</th>
<th>Sensitivity % (95% CI)</th>
<th>N studies</th>
<th>N Pos/Tested</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;15% or not reported</td>
<td>18</td>
<td>372/539</td>
<td>79 (69-90)</td>
<td>7</td>
<td>158/707</td>
<td>84 (66-100)</td>
</tr>
<tr>
<td>&gt;100%</td>
<td>1</td>
<td>14/36</td>
<td>39 (0-92)</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* For overall and stratified analysis a TST cut-off of 10mm was preferentially used; for 2 studies TST 5 mm data was used and for 1 study TST 15 mm as this was the only data available.

** Only one study conducted in HIV-infected children.

NR = Not reported

#### B. QFT-G and QFT-GIT

<table>
<thead>
<tr>
<th>QFT-G/ QFT-GIT</th>
<th>True positives (sensitivity)</th>
<th>False positives (specificity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N studies</td>
<td>N Pos/Tested</td>
<td>Sensitivity % (95% CI)</td>
</tr>
<tr>
<td>Overall</td>
<td>18</td>
<td>335/467</td>
</tr>
<tr>
<td>Overall *</td>
<td>18</td>
<td>335/500</td>
</tr>
<tr>
<td>By QFT type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QFT-G</td>
<td>6</td>
<td>68/79</td>
</tr>
<tr>
<td>QFT-GIT</td>
<td>13</td>
<td>272/393</td>
</tr>
<tr>
<td>By World Bank category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIC</td>
<td>15</td>
<td>298/394</td>
</tr>
<tr>
<td>LMIC</td>
<td>3</td>
<td>37/73</td>
</tr>
<tr>
<td>By TB incidence/100,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>14</td>
<td>294/389</td>
</tr>
<tr>
<td>&gt;25</td>
<td>4</td>
<td>41/78</td>
</tr>
<tr>
<td>By BCG vaccination coverage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50%</td>
<td>6</td>
<td>57/62</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>12</td>
<td>278/405</td>
</tr>
<tr>
<td>By age of child</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5 yrs</td>
<td>5</td>
<td>69/94</td>
</tr>
<tr>
<td>≥5 yrs</td>
<td>13</td>
<td>266/373</td>
</tr>
<tr>
<td>By HIV status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 15% or NR</td>
<td>17</td>
<td>318/440</td>
</tr>
<tr>
<td>&gt;100%</td>
<td>1</td>
<td>17/27</td>
</tr>
</tbody>
</table>

*Indeterminate results included as false-negative.

NR = Not reported
### Use of IGRAs in children

#### C. T-SPOT

<table>
<thead>
<tr>
<th>T-SPOT</th>
<th>False positives (specificity)</th>
<th>True positives (sensitivity)</th>
<th>N studies</th>
<th>N Pos/Tested</th>
<th>Sensitivity % (95% CI)</th>
<th>N studies</th>
<th>N Pos/Tested</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>4</td>
<td>12/143</td>
<td>9</td>
<td>194/336</td>
<td>84 (63-100)</td>
<td>9</td>
<td>194/347</td>
<td>81 (59-100)</td>
</tr>
<tr>
<td>Overall *</td>
<td>-</td>
<td>-</td>
<td>9</td>
<td>194/336</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>By World Bank category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIC</td>
<td>2</td>
<td>3/46</td>
<td>6</td>
<td>126/202</td>
<td>86 (67-100)</td>
<td>6</td>
<td>126/202</td>
<td>86 (67-100)</td>
</tr>
<tr>
<td>LMIC</td>
<td>2</td>
<td>9/97</td>
<td>3</td>
<td>68/134</td>
<td>77 (23-100)</td>
<td>3</td>
<td>68/134</td>
<td>73 (29-100)</td>
</tr>
<tr>
<td>By TB incidence/100,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>2</td>
<td>3/46</td>
<td>6</td>
<td>126/202</td>
<td>87 (68-100)</td>
<td>6</td>
<td>126/202</td>
<td>87 (68-100)</td>
</tr>
<tr>
<td>≥25</td>
<td>2</td>
<td>9/97</td>
<td>3</td>
<td>68/134</td>
<td>73 (29-100)</td>
<td>3</td>
<td>68/134</td>
<td>73 (29-100)</td>
</tr>
<tr>
<td>By BCG vaccination coverage**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50%</td>
<td>1</td>
<td>0/21</td>
<td>3</td>
<td>42/44</td>
<td>97 (92-100)</td>
<td>1</td>
<td>0/21</td>
<td>100 (84-100)</td>
</tr>
<tr>
<td>&gt; 50%**</td>
<td>3</td>
<td>11/75</td>
<td>5</td>
<td>130/239</td>
<td>69 (45-93)</td>
<td>3</td>
<td>11/75</td>
<td>87 (53-100)</td>
</tr>
<tr>
<td>By age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5 yrs</td>
<td>2</td>
<td>8/71</td>
<td>2</td>
<td>49/86</td>
<td>74 (23-100)</td>
<td>2</td>
<td>49/86</td>
<td>74 (23-100)</td>
</tr>
<tr>
<td>≥5 yrs</td>
<td>2</td>
<td>4/72</td>
<td>7</td>
<td>145/250</td>
<td>86 (65-100)</td>
<td>2</td>
<td>145/250</td>
<td>85 (66-100)</td>
</tr>
<tr>
<td>By HIV status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 15% or NR***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 100%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>94 (87-100)</td>
<td>0</td>
<td>0</td>
<td>94 (87-100)</td>
</tr>
</tbody>
</table>

* Indeterminate results included as false-negative.

** BCG vaccination information not provided in one study; study was categorized as >50% BCG vaccinated as national guidelines recommend neonatal vaccination

NR = Not reported
The measure of sensitivity and specificity is based on TST results using a 10 mm cut-off. Some studies are included that gave only 5 mm or 15 mm results, while one study used a cut-off of either 6 or 15 mm according to risk factors.
Figure 17. Forest plots for Quantiferon sensitivity and specificity in active TB, separated by World Bank income index

NOTE: Weights are from random effects analysis.

Overall (I-squared = 0.77; 0.64-0.85)

**Quantiferon Sensitivity**

**Study**

<table>
<thead>
<tr>
<th>High Income Countries</th>
<th>Sensitivity</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bamford, 2009, UK</td>
<td>64 (56, 72)</td>
<td>7.5</td>
</tr>
<tr>
<td>Bianchi, 2009, Italy</td>
<td>94 (70, 100)</td>
<td>6.4</td>
</tr>
<tr>
<td>Chun, 2008, Korea</td>
<td>80 (28, 99)</td>
<td>3.4</td>
</tr>
<tr>
<td>Connell, 2006, Australia</td>
<td>100 (66, 100)</td>
<td>6.1</td>
</tr>
<tr>
<td>Connell, 2008, Australia</td>
<td>89 (52, 100)</td>
<td>5.0</td>
</tr>
<tr>
<td>Detjen, 2007, Germany</td>
<td>93 (77, 99)</td>
<td>7.0</td>
</tr>
<tr>
<td>Dominguez, 2007, Spain</td>
<td>67 (30, 93)</td>
<td>3.9</td>
</tr>
<tr>
<td>Grare, 2010, France</td>
<td>60 (15, 95)</td>
<td>3.0</td>
</tr>
<tr>
<td>Haustein, 2009, UK</td>
<td>88 (68, 97)</td>
<td>6.4</td>
</tr>
<tr>
<td>Herrmann, 2009, France</td>
<td>78 (60, 91)</td>
<td>6.4</td>
</tr>
<tr>
<td>Higuchi, 2008, Japan</td>
<td>96 (75, 100)</td>
<td>6.8</td>
</tr>
<tr>
<td>Higuchi, 2009, Japan</td>
<td>100 (69, 100)</td>
<td>6.3</td>
</tr>
<tr>
<td>Kampmann, 2009, UK</td>
<td>68 (54, 79)</td>
<td>6.8</td>
</tr>
<tr>
<td>Lighter, 2009, US</td>
<td>100 (29, 100)</td>
<td>3.4</td>
</tr>
<tr>
<td><strong>Subtotal (I-squared = 0.69; 0.48-0.82)</strong></td>
<td>86 (78, 94)</td>
<td>84.2</td>
</tr>
</tbody>
</table>

**Low Middle Income Countries**

| Dogra, 2006, India   | 50 (19, 81) | 3.9    |
| Dogra, 2010, South Africa | 42 (26, 59) | 6.1    |
| Stavri, 2009, Romania| 63 (42, 81) | 5.7    |
| **Subtotal (I-squared = 0.32; 0.03-0.81)** | 51 (38, 63)| 15.8   |

**Overall (I-squared = 0.77; 0.64-0.85)**

| 82 (72, 91) | 100 |

**Quantiferon Specificity**

**Study**

<table>
<thead>
<tr>
<th>High Income Countries</th>
<th>Specificity</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chun, 2008, Korea</td>
<td>99 (92, 100)</td>
<td>15.5</td>
</tr>
<tr>
<td>Detjen, 2007, Germany</td>
<td>100 (84, 100)</td>
<td>14.7</td>
</tr>
<tr>
<td>Herrmann, 2009, France</td>
<td>70 (51, 85)</td>
<td>11.8</td>
</tr>
<tr>
<td>Higuchi, 2008, Japan</td>
<td>35 (25, 47)</td>
<td>14.0</td>
</tr>
<tr>
<td>Kampmann, 2009, UK</td>
<td>92 (73, 99)</td>
<td>13.3</td>
</tr>
<tr>
<td><strong>Subtotal (I-squared = 0.90; 0.85-0.93)</strong></td>
<td>91 (74, 100)</td>
<td>69.2</td>
</tr>
</tbody>
</table>

**Low Income Countries**

| Dogra, 2006, India   | 94 (87, 98) | 15.3   |
| Moyo, 2010, South Africa | 85 (81, 89)| 15.5   |
| **Subtotal (I-squared = 0.71; 0.02-0.92)** | 90 (83, 95)| 30.8   |

**Overall (I-squared = 0.89; 0.84-0.92)**

| 90 (78, 100) | 100 |

**NOTE:** Weights are from random effects analysis.
Figure 18. Forest plots for T-SPOT sensitivity and specificity in active TB, separated by World Bank income index.

### TSPOT SENSITIVITY

<table>
<thead>
<tr>
<th>Study</th>
<th>Sensitivity</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Income Countries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bamford, 2009, UK</td>
<td>50 (40, 60)</td>
<td>11.63</td>
</tr>
<tr>
<td>Bergamini, 2009, Italy</td>
<td>100 (99, 100)</td>
<td>10.54</td>
</tr>
<tr>
<td>Connell, 2008, Australia</td>
<td>100 (66, 100)</td>
<td>10.99</td>
</tr>
<tr>
<td>Detjen, 2007, Germany</td>
<td>93 (77, 99)</td>
<td>11.56</td>
</tr>
<tr>
<td>Dominguez, 2007, Spain</td>
<td>86 (42, 100)</td>
<td>9.38</td>
</tr>
<tr>
<td>Kampmann, 2009, UK</td>
<td>54 (41, 68)</td>
<td>11.35</td>
</tr>
<tr>
<td>Subtotal (I-squared = 0.89; 0.79-0.94)</td>
<td>86 (67, 100)</td>
<td>65.45</td>
</tr>
<tr>
<td>Low Income Countries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hansted, 2008, Lithuania</td>
<td>100 (85, 100)</td>
<td>11.84</td>
</tr>
<tr>
<td>Nicol, 2009, South Africa</td>
<td>40 (27, 53)</td>
<td>11.39</td>
</tr>
<tr>
<td>Warrier, 2009, India</td>
<td>42 (28, 56)</td>
<td>11.32</td>
</tr>
<tr>
<td>Subtotal (I-squared = 0, 0-0.73)</td>
<td>77 (23, 100)</td>
<td>34.55</td>
</tr>
<tr>
<td>Overall (I-squared = 0.89; 0.81-0.93)</td>
<td>84 (63, 100)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

### TSPOT SPECIFICITY

<table>
<thead>
<tr>
<th>Study</th>
<th>Specificity</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Income Countries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detjen, 2007, Germany</td>
<td>100 (84, 100)</td>
<td>28.2</td>
</tr>
<tr>
<td>Kampmann, 2009, UK</td>
<td>88 (69, 97)</td>
<td>15.6</td>
</tr>
<tr>
<td>Subtotal (I-squared = 0; 0-0.70)</td>
<td>95 (84, 100)</td>
<td>43.8</td>
</tr>
<tr>
<td>Low Income Countries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nicol, 2009, South Africa</td>
<td>84 (71, 93)</td>
<td>21.4</td>
</tr>
<tr>
<td>Warrier, 2009, India</td>
<td>98 (89, 100)</td>
<td>34.9</td>
</tr>
<tr>
<td>Subtotal (I-squared = 0.67; 0-0.91)</td>
<td>93 (83,100)</td>
<td>56.2</td>
</tr>
<tr>
<td>Overall (I-squared = 0; 0-0.62)</td>
<td>94 (87, 100)</td>
<td>100</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis.
3.2.10 Summary of findings and GRADE evidence profiles

- The majority of IGRA studies in children have been performed in high-income countries and extrapolation to low- and middle-income settings with high background TB infection rates is not appropriate. However, based on available data, IGRA and the TST have very similar accuracy for diagnosis of LTBI and active TB in children in LMIC settings;

- Major methodological inconsistencies between studies had a negative effect on the comparability of studies and results. A key constraint is the lack of appropriate reference standards for diagnosis of paediatric TB, limiting the interpretation of estimates of test accuracy in children other than those with definite TB;

- A clear advantage of IGRA over TST in detecting LTBI in exposed or unexposed individuals or in a gradient of exposure was not detected;

- Lower sensitivity of both IGRA and TST was found in study populations with >50% BCG coverage. The reasons are not clear; however, BCG coverage may capture populations from settings with higher burden of TB, hence with different backgrounds and underlying conditions that may impair test accuracy, such as co-infections with helminths and malnutrition;

- Both IGRA and TST showed lower sensitivity in HIV-infected children in one study assessed.

- Overall, the ability of TST and IGRA were suboptimal to ‘rule out’ active TB. The main limitation for assessment of the specificity of the diagnostic assays among ‘no-TB’ groups was the small number of studies that described adequate methodology to exclude and diagnose active TB;

- Indeterminate IGRA results varied across all studies, but higher rates were associated with young age, immune-suppression or helminth co-infection in individual studies on TB exposure;

- In studies on active TB no correlation was found between indeterminate results and age, HIV status, TB burden or BCG vaccination status;

- Studies rarely addressed the operational aspects and implementation feasibility of IGRA. Cost was noted as an important and limiting factor. Aspects inherent to the use of IGRA in children, such as the difficulty of phlebotomy and the amount of blood needed in young children, are relevant implementation considerations.

3.2.11 Strengths and limitations of the evidence base

Studies included assessed very different populations in diverse settings, with the biggest challenge and limitation related to major differences in methodological approaches across studies and non-standardised definitions of reference standards, TB exposure and TB disease.

Sample sizes in the different studies varied greatly and were less than ten in some of the subgroups analysed, which adversely impact on generalisability of the findings.

Empirical random effects weighting and separately synthesizing data for currently manufactured IGRA were performed in order to minimize heterogeneity; however, heterogeneity remained substantial for the primary outcomes of sensitivity and specificity.
No standard criteria exist for defining high TB-incidence countries and the World Bank income classification is an imperfect surrogate for national TB incidence; nevertheless, results were fundamentally unchanged when restricted to countries with an arbitrarily chosen annual TB incidence of greater than or equal to 25/100,000.

It is likely that ongoing studies were missed despite systematic searching. It is also possible that studies that found poor IGRA performance were less likely to be published. Given the lack of statistical methods to account for publication bias in diagnostic meta-analyses, it would be prudent to assume some degree of overestimation of estimates due to publication bias.

The systematic review focused on test accuracy (ie. sensitivity and specificity) for the diagnosis of active TB and TB exposure as surrogate for LBTI. None of the studies reviewed provided information on patient-important outcomes, ie. showing that IGRA s used in a given situation resulted in a clinically relevant improvement in patient care and/or outcomes. In addition, no information was available on the values and preferences of patients.

3.2.12 Final recommendations

The GRADE evidence profiles are provided in Tables 9 to 12. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRA s in children was very low and recommended that these tests should not be used as an alternative to TST in paediatric TB in low and middle-income countries for the diagnosis of latent TB infection, or as an alternative to TST in the workup of a diagnosis of active TB disease in children, irrespective of HIV status (strong recommendation).

The Expert Group also notes that there may be additional harms associated with blood collection in children and that issues such as acceptability and cost have not been adequately addressed in any studies.

<table>
<thead>
<tr>
<th>OVERALL QUALITY OF EVIDENCE</th>
<th>VERY LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRENGTH OF RECOMMENDATION</td>
<td>STRONG</td>
</tr>
</tbody>
</table>
Table 9.  
GRADE evidence profile: The performance of IGRAs for the diagnosis of latent tuberculosis infection in children in LMIC

<table>
<thead>
<tr>
<th>No of participants (studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Risk of progression to active TB</td>
<td>No studies in LMIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>B. Outcome: Performance of IGRAs in studies using a dichotomous measure of exposure as reference standard for LTBI (exposed/unexposed)</td>
<td>229 (4)&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Mainly cross-sectional</td>
<td>Not serious&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Not serious&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Serious&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Very serious&lt;sup&gt;5&lt;/sup&gt;</td>
<td>Likely&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Very Low ⊕⊕⊕⊕</td>
</tr>
<tr>
<td>C. Outcome: Performance of IGRAs in studies assessing different gradients of TB exposure as reference standard for LTBI</td>
<td>1057 (5)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Cross sectional</td>
<td>Not serious&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Not serious&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Serious&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Very serious&lt;sup&gt;5&lt;/sup&gt;</td>
<td>Likely&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Very Low ⊕⊕⊕⊕</td>
</tr>
</tbody>
</table>

The proportion of indeterminate results as well as the influence of HIV-status and young age on IGRA performance were rated as important outcomes (4-6 points) for patients with suspected LTBI. However, due to the small number of studies no subgroup analysis for these outcomes was performed.

Active TB was used as a surrogate measure for LTBI. Tables 10 and 11 describe the evidence profile and summary of findings for studies assessing IGRAs in active TB suspects.

Footnotes

1. The quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high, when there were randomized controlled trials or high quality observational studies (cross-sectional or cohort studies enrolling patients with diagnostic uncertainty) and at moderate, when these types of studies were absent. One point was then subtracted when there was a serious issue identified or two points, when there was a very serious issue identified in any of the criteria used to judge the quality of evidence.

2. Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias cannot be ruled out. Although no points were deducted, a degree of publication bias is likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future; 2) there are anecdotal
examples of unpublished negative studies on IGRAs; and 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

B1 Four studies identified: One evaluated TSPOT, two evaluated T-SPOT and QFT-G, one evaluated T-SPOT and QFT-GIT. In total, QFT-G or QFT-GIT was evaluated in 59 children, T-SPOT in 170 children.

B2 Study limitations were assessed using the QUADAS tool. Tw (50%) studies did not clearly enroll a representative spectrum (patient selection - random, consecutive or convenient - was not reported). Blinding of laboratory personnel was reported in 3/4 studies. Differential verification and execution of the reference standard were not considered important issues for exposure studies since all children were assessed for exposure.

B3 a) All four studies were performed in upper middle-income countries; the data are not necessarily representative for low-income countries.

b) TB exposure is a surrogate measure for patient important outcomes and does not necessarily classify the target condition (LTBI) correctly. Exposure increases the risk of infection and correctly identified children with infection will highly benefit from preventive chemotherapy. (No points subtracted)

B4 Heterogeneity was assessed by looking at the variation between odds ratios for the different studies. For QFT-G/QFT-GIT the ORs varied between 0.43 and 5, for T-SPOT between 1.5 and 24. Differences in the definition of exposure groups between the studies may be responsible for the heterogeneity of the results. Two studies were performed in immune-compromised children, one in 100% HIV-infected children, the other in oncology patients. (1 point subtracted)

B5 The 95% CIs for the odds of detecting exposed versus unexposed children were very wide for both QFT-G/QFT-GIT (1.30, 95%-CI 0.2-8.3) and T-SPOT (2.24, 95%-CI 0.88-5.64). The data available from LMIC was very limited and the sample size for exposure groups 3/4 studies was <50, some subgroups analyzed had a sample size of n=2, which highly increases the risk of imprecision. (2 points subtracted)

C1 Five studies identified: two evaluated QFT-GIT (one without using a mitogen control), one evaluated QFT-G and one evaluated T-SPOT and QFT-GIT. In total, QFT-G or QFT-GIT was evaluated in 773 and T-SPOT in 225 children.

C2 Study limitations were assessed using the QUADAS tool. One study assessed a representative spectrum of children and recruitment was performed in a consecutive manner. Blinding of laboratory technicians was reported in one study. Like for dichotomous exposure studies, differential verification and execution of the reference standard were not considered important issues for exposure studies since all children were assessed for exposure.

C3 a) Three studies were performed in low-income countries, one in a lower-middle, one in an upper middle income country.

b) TB exposure is a surrogate measure and does not necessarily classify the target condition (LTBI) correctly. Exposure increases the risk of infection and correctly identified children with infection will highly benefit from preventive chemotherapy. (No points subtracted)

C4 Heterogeneity for T-SPOT could not be assessed since there was only one study. Heterogeneity for QFT was assessed using I-squared statistics and considered to be high (90%). Four studies used microbiological indicators (smear status), one used proximity to the index case as measure of exposure. (1 point subtracted)

C5 The 95% CIs for the pooled random correlation between QFT-studies assessing exposure gradients were wide (QFT-G/QFT-GIT 0.28, 95%-CI 0.06-0.86) For T-SPOT, the fixed correlation was 0.15, 95%-CI 0.02-0.37. Similar, when calculating regression slopes for exposure gradients, confidence intervals were wide and overlapping for all tests assessed. The data available from LMIC was limited, and the sample sizes assessed small (2 points subtracted)
**Table 10. GRADE summary of findings – IGRAs for LTBI in children**

**Review question:** What is the performance of IGRAs for the detection of LTBI in children in LMIC?

**Patients/population:** Children <18 years old in low, lower-middle and upper-middle income countries being screened for LTBI

**Index test:** QuantiFERON-TB Gold [QFT-G], QuantiFERON-TB Gold In-Tube [QFT-GIT], and T-SPOT.TB [T-SPOT].

**Importance:** Children have a high risk of progression to active TB after infection. Correctly identified children with LTBI benefit from preventive therapy.

**Reference standards:** Incident TB, Exposure (dichotomous and gradient), prevalent TB

**Studies:** Observational studies (cohort, cross-sectional, case-control)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. Participants</th>
<th>Principal Findings</th>
<th>What do these findings mean?</th>
<th>Quality of Evidence</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive value for active TB</td>
<td></td>
<td>No studies in LMIC</td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
</tbody>
</table>
| Performance of IGRAs against dichotomous measure of exposure | QFT-G/QFT-GIT: 59 (3) | Pooled Odds ratios  
  - QFT-G/QFT-GIT: OR 1.30 (95% CI 0.20-8.32)  
  - T-SPOT: OR 2.24 (95% CI 0.88-5.64)  
  - TST (10mm): OR 0.81 (95% CI 0.38-1.74) | Children exposed to TB have a higher risk of LTBI, expressed by a higher probability of a positive test for LTBI (QFT, T-SPOT or TST) than in unexposed children. Wide and overlapping confidence intervals indicate similar performance of all three tests. | Very Low ☠️☠️☠️ | Critical (7-9) |
| | T-SPOT: 170 (4) | | | | |
| | TST (10mm): 159 (3) | | | | |
| Performance of IGRAs against exposure gradient | QFT-G/QFT-GIT: 773 (5) | 1. Pooled correlation between test and exposure gradient:  
  - QFT-G/QFT-GIT: 0.28 (95%CI 0.06-0.86, I² 0.90)  
  - T-SPOT (not pooled, 1 study): 0.15 (95% CI 0.02-0.37)  
  - TST (10 mm): 0.22 (95% CI 0.11-0.39, I² 0.65)  
  2. Regression slopes  
  - QFT-G/QFT-GIT: 1.84 (95%CI 1.38-2.44, I² 0.66)  
  - T-SPOT: 1.63 (95%CI 1.12-2.39)  
  - TST (10 mm): 1.73 (95% CI 1.36-2.20, I² 0.59) | A higher level of exposure to TB indicates a higher risk for LTBI, expressed by a positive correlation between LTBI test and exposure gradients. IGRAs and TST show a similar correlation with exposure gradients (wide and overlapping confidence intervals). | Very Low ☠️☠️ | Critical (7-9) |
| | T-SPOT: 225 (1) | | | | |
| | TST (10 mm) 871 (5) | | | | |
Table 11. GRADE evidence profile: The diagnostic accuracy of IGRAs for the diagnosis of active tuberculosis in children in LMIC

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: What is the sensitivity of IGRAs in children with active TB?</td>
<td>207 (6)¹</td>
<td>Mainly cross-sectional</td>
<td>Serious⁴ (-1)</td>
<td>Not serious²</td>
<td>Not serious³</td>
<td>Very serious⁶ (-2)</td>
<td>Likely⁵</td>
<td>Very Low ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>B: What is the specificity of IGRAs in children without TB?</td>
<td>519 (4)¹</td>
<td>Mainly cross-sectional</td>
<td>Serious⁴ (-1)</td>
<td>Not serious³</td>
<td>Serious⁴</td>
<td>Serious⁶ (-1)</td>
<td>Likely⁵</td>
<td>Very Low ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>C: What is the proportion of indeterminate IGRA results among children assessed for active TB?</td>
<td>656 (5)¹</td>
<td>Mainly cross sectional</td>
<td>Serious⁴ (-1)</td>
<td>Not serious³</td>
<td>Not serious⁴</td>
<td>Serious⁵</td>
<td>Likely⁵</td>
<td>Very Low ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>D: What is the diagnostic accuracy of IGRAs in HIV-infected children?</td>
<td>36 (1)¹</td>
<td>Cross-sectional</td>
<td>Serious⁴ (-1)</td>
<td>Not serious³</td>
<td>Not applicable⁴</td>
<td>Very serious⁵</td>
<td>Likely⁵</td>
<td>Very Low ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>E: What is the diagnostic accuracy of IGRAs in children &lt; 5 years?</td>
<td>471 (2)¹</td>
<td>Cross-sectional</td>
<td>Serious⁴ (-1)</td>
<td>Not serious³</td>
<td>Not applicable⁴</td>
<td>Very serious⁵</td>
<td>Likely⁵</td>
<td>Very Low ☐ ☐ ☐ ☐</td>
</tr>
</tbody>
</table>

Footnotes

¹ The quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high, when there were randomized controlled trials or high quality observational studies (cross-sectional or cohort studies enrolling patients with diagnostic uncertainty) and at moderate, when these types of studies were absent. One point was then subtracted when there was a serious issue identified or two points, when there was a very serious issue identified in any of the criteria used to judge the quality of evidence.

² Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias cannot be ruled out. Although no points were deducted a degree of publication bias is likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future; 2) there are anecdotal examples of unpublished negative studies on IGRAs; and 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

¹¹ 6 studies identified for the assessment of sensitivity (TP and FN) of commercial IGRAs in children with suspected TB or active TB: 3 evaluated T-SPOT, 2 evaluated QFT-GIT, and 1 evaluated QFT-G. In total, 73 children were evaluated with QFT-G or QFT-GIT and 134 with T-SPOT.
Study limitations were assessed using QUADAS. One study described a representative spectrum with consecutive patient selection. In 2 studies it remained unclear whether differential verification was avoided. The execution of the reference standard (definition of active TB) was described in 5/6 studies but definition of the reference standard still varied between different studies and was described more clearly in some than others. Blinding of both laboratory technicians and clinicians remained unclear in the majority of studies. (1 point subtracted)

Four studies were performed in upper middle, 2 in lower middle-income countries and none in low income countries. Hence, the findings may not be generalisable to low-income countries. Diagnostic accuracy of IGRAs is only a surrogate for patient important outcomes. False negative tests result in children not being diagnosed and started on treatment, which will result in progression of disease, and potentially death. (No points subtracted).

The $I^2$ statistics showed low to moderate heterogeneity among studies assessing QFT-G/QFT-GIT (32%) with sensitivities ranging from 50 to 63%. Sensitivities for three studies assessing T-SPOT ranged between 42 and 100%; $I^2$-squared was 0%, which may be due to the small number of studies included in the analysis. Indeterminate results, if added to false negative results, lowered the pooled sensitivity for both assays. It can be assumed that the heterogeneity among the studies is caused by factors such as differences in the study populations, number of confirmed versus probable TB cases included in the studies, disease severity, age groups and others.

The 95% confidence interval for pooled sensitivity was wide for both QFT-G/QFT-GIT (51%, 95% CI 38-63%) and T-SPOT (77%, 95% CI 23-100%). The data available from LMIC was very limited and sample sizes in the individual studies small. (2 points subtracted)

Four studies assessed specificity in children where active TB was excluded: 2 evaluated T-SPOT, and 2 QFT-GIT. In total, 422 children were evaluated with QFT-G or QFT-GIT, and 97 children with T-SPOT.

Study limitations were assessed using QUADAS. One study described recruitment of a representative spectrum of children in a consecutive manner. Differential verification was avoided in all studies, and the execution of the reference standard was described in the majority, even though with differing quality. Blinding of laboratory technicians and clinicians remained unclear in the majority of studies. (1 point subtracted)

None of the studies was performed in low-income countries, two in lower, and two in upper middle-income countries. Diagnostic accuracy of IGRAs is a surrogate for patient-important outcomes. False positive results can lead to a delay in making a correct diagnosis. IGRAs cannot differentiate between disease and infection and positive results may just reflect underlying TB infection. (No points subtracted)

Specificity for QFT-GIT ranged between 85 and 94%, the $I^2$ statistics of 71% indicates that there is a considerable amount of heterogeneity and suggests that results should be interpreted with caution. For T-SPOT, specificity ranged between 84% and 98%, $I^2$ statistics was 0 (again, this is likely due to the small number of studies included in this analysis).

The 95% CI for pooled specificity for QFT-G/QFT-GIT (90%, 95%CI 83-95) and T-SPOT (93%, 95%CI 83-100) were relatively narrow. However, the data available for LMIC was limited and the sample sizes of included studies small. (1 point subtracted)

5 studies assessed commercial IGRAs in children with suspected TB, active TB or ‘no TB’ and included indeterminate results: indeterminate results for QFT-G or QFT-GIT were reported in 3 studies among 524 children, indeterminate results for T-SPOT were reported in 2 studies among 132 children.

Study limitations were assessed using QUADAS. One study described recruitment of a representative spectrum of children in a consecutive manner. Differential verification was avoided and the execution of the reference standard (definition of active TB) was described in the majority. Blinding of both laboratory technicians and clinicians remained unclear in the majority of studies. (1 point subtracted)
Three studies were performed in upper middle, 2 in lower middle-income countries and none in low income countries. Hence, the findings may not be generalizable to low-income countries. Diagnostic accuracy of IGRAs is only a surrogate for patient important outcomes. False negative or indeterminate tests result in children not being diagnosed and started on treatment, which will result in progression of disease, and potentially death. (No points subtracted)

Heterogeneity was assessed by looking at the range of indeterminate results across studies. The overall proportion of indeterminates was 25% for QFT-G, 4.1 for QFT-GIT studies (range 0-5% in individual studies) and 6.8% for T-SPOT (range 0-8% in individual studies). The QFT-G study showing 25% indeterminates was performed in 100% HIV-infected children with active TB and classifies a high-risk patient group that should be assessed separately for indeterminate results. (No points subtracted)

The number of studies from LMI assessing indeterminate results was limited and the sample size of study populations used for this analysis was small, accounting for serious imprecision. (1 point subtracted)

One study assessed QuantiFERON-TB Gold in 36 HIV-infected children with active TB in Romania (an upper middle-income country).

Study limitations were assessed using QUADAS. The spectrum of patients included in the study was not representative, patient selection was unclear. It also remained unclear whether laboratory technicians and clinicians were blinded. (1 point subtracted)

The study was performed among HIV-infected children with a diagnosis of TB in Romania, an upper middle-income country. The results may not be generalizable to low-income countries. Sensitivity of IGRAs is only a surrogate for patient-important outcomes. False negative results, particularly in HIV-infected children, may result in under-diagnosis of disease and, possibly in death. If indeterminate results were added to false negative results the sensitivity was lowered from 63% (indeterminates excluded) to 47% (95%CI 0-100). (No points subtracted)

Only one study – inconsistency therefore cannot be assessed.

The 95% CI for sensitivity of QFT-G in 36 HIV-infected children was very wide (63%, 95%CI 16-100). (2 points subtracted)

In 2 studies evaluating IGRAs for the diagnosis of active TB the mean or median age of children was below five years. One evaluated T-SPOT, and one QFT-GIT. QFT-GIT was assessed in 363 children (36 with active TB, 327 in ‘no TB’ group) and T-SPOT in 108 children (58 with active TB and 50 in ‘no TB’ group).

Study limitations were assessed using QUADAS. The spectrum and patient selection as well as blinding of laboratory technicians was unclear in both studies. Also, studies for this stratum were selected according to mean or median age since only few studies reported data stratified to age groups. (1 point subtracted)

Both studies were performed in upper middle-income countries, none in lower middle or low-income countries. Hence, the data may not be generalizable to low-income countries. Test accuracy is only a surrogate for patient-important outcomes. Children under 5 have the highest risk of severe disease and false negative results can result in fatal outcomes. At the same time, false positive results can result in misdiagnosis and prolong the time to correct diagnosis. (No points subtracted)

Heterogeneity could not be assessed since each test was only assessed in one study.

The confidence intervals for sensitivity and specificity of QFT-GIT were small, but wide for T-SPOT. The data from LMIC to address this objective was extremely limited. (2 points subtracted).
Table 12. GRADE Summary of Findings – IGRAs for the diagnosis of active TB in children

<table>
<thead>
<tr>
<th>Review question:</th>
<th>What is the diagnostic accuracy of IGRAs for the diagnosis of active TB in children in LMIC?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients/population</td>
<td>TB suspects or active TB patients and control group with ‘no TB’ in low and middle income countries</td>
</tr>
<tr>
<td>Setting:</td>
<td>Mainly mixed, in- and outpatients</td>
</tr>
<tr>
<td>Index test:</td>
<td>Quantiferon-TB gold (QFT-G), Quantiferon-TB gold in-tube (QFT-GIT), and T-spot.TB (T-SPOT).</td>
</tr>
<tr>
<td>Importance:</td>
<td>Diagnosis of childhood TB is often a composite of risk factors, clinical signs and symptoms and radiological imaging, since culture confirmation proves difficult. Highly sensitive assays would support a diagnosis of active TB.</td>
</tr>
<tr>
<td>Reference standard:</td>
<td>Culture confirmed TB and probable TB versus ‘no TB’</td>
</tr>
<tr>
<td>Studies:</td>
<td>Cross-sectional or case-control studies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Index test</th>
<th>No. of Participants (Studies)</th>
<th>Effect % (95% CI) Main findings</th>
<th>What do these results mean given a 10% prevalence among suspects being screened for TB?</th>
<th>What do these results mean given a 30% prevalence among suspects being screened for TB?</th>
<th>Quality of Evidence</th>
</tr>
</thead>
</table>
| What is the diagnostic accuracy of IGRAs for active TB? | T-SPOT | Sensitivity 143 (3)  
Specificity 97 (2) | Pooled sensitivity 77% (23-100)  
- Not considerably lower if indeterminate results counted as false negative 76% (18-100)  
Pooled specificity 93% (83-100)  
- Lower in population with >50% BCG coverage 85% (15-100) | With a prevalence of 10%, 100/1000 children will have TB. Of these, 77 will be correctly identified with T-SPOT, 23 will be missed. Of 900 children without TB, 837 will not be treated, 63 will be unnecessarily treated. | With a prevalence of 30%, 300/1000 will have TB. 231 will be correctly identified with T-SPOT, 69 will be missed. Of 700 children without TB, 651 will not be treated, 49 will be unnecessarily treated. | Very Low ●●●● |
| | QFT-G/  
QFT-GIT | Sensitivity 84 (3)  
Specificity 422 (2) | Pooled sensitivity QFT-G, 1 study: 65% (47-82)  
QFT-GIT, 2 studies: 36% (29-44)  
Combined: 51% (38-63)  
- Pooled sensitivity including indeterminates for QFT-G and QFT-GIT 36% (23-49)  
Pooled specificity | With a prevalence of 10%, 100/1000 will have TB. Of these, 65 will be correctly identified by QFT-G, 35 will be missed. 36 will be identified by QFT-GIT, 64 will be missed. Indeterminate results lead to slightly more missed cases. Of 900 children without TB, 90 children will be unnecessarily treated based on QFT-GIT results. | With a prevalence of 30%, 300/1000 will have TB. Of these, 195 will be correctly identified with QFT-G, 105 will be missed. 108 will be identified by QFT-GIT, 192 will be missed. Of 700 children without TB, 70 will be unnecessarily treated based on QFT-GIT results. | Very Low ●●●● |
<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Pooled Sensitivity</th>
<th>Pooled Specificity</th>
<th>IGRAs do not perform significantly different from TST</th>
</tr>
</thead>
<tbody>
<tr>
<td>TST</td>
<td>168 (5)</td>
<td>490 (3)</td>
<td>Pooled sensitivity</td>
<td>65% (31-99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pooled specificity</td>
<td>90% (82-98)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the proportion of indeterminate IGRAs among children assessed for active TB?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-SPOT</td>
<td>132 (2)</td>
<td></td>
<td>Indeterminates/total number of tests</td>
<td>9/132 = 6.82 %</td>
<td>Range of % indeterminates across studies 0-8%</td>
</tr>
<tr>
<td>QFT-G/ QFT-GIT</td>
<td>QFT-G 36 (1)</td>
<td>QFT-GIT 488 (2)</td>
<td>Indeterminates/total number of tests</td>
<td>QFT-G: 9/36 = 25%</td>
<td>QFT-GIT: 20/488 = 4.1%</td>
</tr>
</tbody>
</table>

Very Low

Rating: ⬜⬜⬜⬜
<table>
<thead>
<tr>
<th>Performance of IGRAs in HIV-infected children</th>
<th><strong>T-SPOT</strong></th>
<th><strong>No studies</strong></th>
<th><strong>No studies</strong></th>
<th><strong>No studies</strong></th>
<th><strong>Very Low</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>QFT-G</strong></td>
<td>Sensitivity</td>
<td>36 (1)</td>
<td>Sensitivity</td>
<td>QFT-G: 63% (16-100)</td>
<td>With a prevalence of 10%, 100/1000 HIV-infected children will have TB. Of these, 63 will be correctly identified with QFT-G, 37 will be missed.</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>No studies</td>
<td>Specificity</td>
<td>47% (0-100) if indeterminates counted as FN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled specificity</td>
<td>No studies</td>
<td>Pooled specificity</td>
<td>No studies</td>
<td></td>
</tr>
<tr>
<td><strong>TST</strong></td>
<td>Sensitivity</td>
<td>36 (1)</td>
<td>Sensitivity</td>
<td>39% (0-100)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>No studies</td>
<td>Specificity</td>
<td>No studies</td>
<td></td>
</tr>
<tr>
<td><strong>Performance in children &lt;5yrs</strong></td>
<td><strong>T-SPOT</strong></td>
<td>Sensitivity</td>
<td>134 (3)</td>
<td>Sensitivity</td>
<td>With a prevalence of 10%, 100/1000 will have TB. Of these, 77 will be correctly identified by T-SPOT, 23 will be missed. Of 900 children without TB, 837 will not be treated and 63 will be unnecessarily treated.</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>97 (2)</td>
<td>Specificity</td>
<td>93% (83-100)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled specificity</td>
<td>No studies</td>
<td>Pooled specificity</td>
<td>No studies</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Pooled sensitivity</td>
<td>Pooled specificity</td>
<td>With a prevalence of 10%, 100/1000 will have TB. Of these, 35 will be correctly identified by QFT-GIT, 65 will be missed. Of 900 children without TB, 765 will not be treated, 135 will be unnecessarily treated.</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>QFT-GIT</td>
<td>36 (1)</td>
<td>327 (1)</td>
<td>Pooled sensitivity</td>
<td>Pooled specificity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>35% (30-40)</td>
<td>85% (81-89)</td>
<td></td>
</tr>
<tr>
<td>TST</td>
<td>99 (2)</td>
<td>395 (2)</td>
<td>Pooled sensitivity</td>
<td>Pooled specificity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>41% (0-85)</td>
<td>83% (81-86)</td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity and specificity of T-SPOT are higher than of QFT-GIT or TST, but the difference is not significant (overlapping confidence intervals).
Use of IGRAs for the diagnosis of LTBI in HIV-infected individuals

3.3 Use of IGRAs for the diagnosis of LTBI in HIV-infected individuals

3.3.1 Objectives, reference standards and outcomes

Objectives

To assess IGRA test performance in diagnosing LTBI in HIV-infected individuals living in low/middle income countries, with the aim to identify those who would benefit from isoniazid preventive therapy (IPT).

Reference standards and primary outcomes

A major challenge for studies evaluating the performance of IGRAs is the lack of gold standard for LTBI. Consequently, studies were evaluated on three primary outcomes:

- Predictive value of IGRAs for development of active TB;
- Sensitivity of IGRAs in patients with culture-confirmed active TB (as a surrogate reference standard for TB infection);
- Correlation between IGRA and TST results.

In addition to the primary outcomes two characteristics that could influence the overall utility of IGRAs were evaluated:

- Proportion of indeterminate IGRA results (i.e. not interpretable either due to high IFN-\(\gamma\) response in the negative control or low IFN-\(\gamma\) response in the positive control);
- Impact of HIV-related immunosuppression (i.e., CD4+ cell count) on test performance (i.e. proportion of positive and indeterminate IGRA results, or positive TST results).

Outcome definitions

Incidence of active TB: the number of active TB cases that developed over a specified median duration of follow-up divided by the number of persons at risk (cumulative incidence), or the number of new active TB cases divided by the total person-time of follow-up (incidence rate);

Sensitivity: the proportion of individuals with a positive IGRA result among those with culture-positive TB (including indeterminate IGRA results in the denominator if they occurred in individuals with culture-positive TB);

Concordance (i.e. agreement): the proportion of individuals for whom IGRA and TST results were both positive or both negative among all individuals tested. In addition, the proportion of positive and indeterminate IGRA results was determined in the following CD4+ cell count strata: <200 cells/\(\mu\)l and \(\geq\)200 cells/\(\mu\)l. To assess the impact of HIV-related immunosuppression, the difference in the proportion of positive results and the difference in the proportion of indeterminate results was calculated between the higher and lower CD4+ cell count strata.

3.3.2 Search results

The initial search yielded 791 citations. After full-text review of 129 papers evaluating IGRAs in immunocompromised individuals, 29 were determined to meet eligibility criteria. Because some papers included more than one commercial IGRA, there were 37 unique evaluations (hereafter referred to as...
Use of IGRAs for the diagnosis of LTBI in HIV-infected individuals

studies) – 19 of QFT-GIT and 18 of T-SPOT – that included a total of 5,736 HIV-infected individuals (Annex 4). TST was concurrently performed in 23 (62%) studies.

### 3.3.3 Study characteristics

Of the 37 included studies, 22 (59%) were conducted in low/middle income countries. Among these 22 studies, there was a high degree of variation in study design and study population. 15 (68%) studies included only ambulatory HIV-positive individuals. IGRAs were performed in persons with or suspected of having active TB in 12 studies, asymptomatic HIV-positive persons being evaluated for LTBI in 6 studies, and both types of individuals in 4 studies. Seven (32%) studies had some industry involvement, including donation of IGRA kits (5 studies) and work/financial relationships between authors and IGRA manufacturers (2 studies).

#### 3.3.4 Study quality

A quality assessment was conducted separately for each of the three primary outcomes.

1. **Predictive value of IGRAs**: was appraised for quality with a modified version of the Newcastle-Ottawa Scale (NOS) for longitudinal/cohort studies. This tool assesses quality in three domains: selection of cohorts (representativeness of sample and absence of outcome at baseline), comparability of cohorts (adjustment for potential confounders), and assessment of outcome (blinding and complete verification);

2. **The sensitivity of IGRAs** in culture-confirmed active TB was evaluated using a subset of relevant items from the QUADAS tool.

3. **The agreement between IGRA and TST results** was evaluated using a subset of relevant items from the QUADAS tool.

#### 3.3.5 Risk of progression to active TB

One longitudinal study that evaluated the ability of IGRAs to predict future development of active TB among HIV-infected individuals living in a low/middle income country was identified (Table 13). Given the very limited data, two additional studies that were conducted in high income countries were identified and included in the analysis of risk of progression to active TB. Based on the Newcastle-Ottawa scale, all three studies enrolled a representative sample of patients. However, only one study had an adequate duration of follow-up (≥1 year) and all three studies scored poorly on outcome assessment (ie, did not adequately rule-out active TB at baseline or did not adequately evaluate all participants for active TB during follow-up). In addition, all studies had very few incident cases of active TB.

All three studies reported a higher risk of active TB in individuals with a positive IGRA result than in individuals with a negative IGRA result. However, in the two studies conducted in high income countries, the absolute difference in cumulative incidence of active TB was not statistically different between persons with positive and negative QFT-GIT results (8% vs. 0%, risk difference 8%, 95% CI -0.7% to 17%, median follow-up 19 months) or TSPOT results (10% vs. 0%, risk difference 10%, 95% CI -3% to +23%, median follow-up 12 months for positive TSPOT results and 3 months for negative TSPOT results). The only study from a low/middle income country reported that QFT-GIT results stratified HIV-positive individuals into low risk (1%) and high risk (12%) groups for development of active TB within 6 months of ART initiation. However, QFT-GIT results were adjusted for CD4+ T-lymphocyte and interpreted using a non-standard cut-point of 0.00625 IU/mL.
### Table 13. Risk of active tuberculosis in HIV-infected individuals, stratified by IGRA result

<table>
<thead>
<tr>
<th>IGRA result</th>
<th>Active TB (N)</th>
<th>Cumulative incidence (%, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aichelburg 2009 (Austria)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QFT-GIT positive (N=36)</td>
<td>3</td>
<td>8%, 2-22%</td>
</tr>
<tr>
<td>QFT-GIT negative (N=705)</td>
<td>0</td>
<td>0%, 0-0.5%</td>
</tr>
<tr>
<td>QFT-GIT indeterminate (N=44)</td>
<td>0</td>
<td>0%, 0-9%</td>
</tr>
<tr>
<td><strong>Clark 2009 (UK)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSPOT positive (N=20)</td>
<td>2</td>
<td>10%, 1-32%</td>
</tr>
<tr>
<td>TSPOT negative (N=114)</td>
<td>0</td>
<td>0%, 0-3%</td>
</tr>
<tr>
<td>TSPOT indeterminate</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Elliot 2009 (Cambodia)</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QFT-GIT positive</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>QFT-GIT negative</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>QFT-GIT indeterminate</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

* Cumulative incidence could not be calculated as the number of persons with active TB in each IGRA result category was not reported

Abbreviations: IGRA - interferon-gamma release assay; CI - confidence interval; TSPOT, T-SPOT.TB, QFT-GIT - QuantiFERON-Gold In-tube; NR - Not reported

#### 3.3.6 Sensitivity in culture-confirmed active TB patients

18 studies evaluating the sensitivity of IGRAs in HIV-infected adults with active TB were identified, of which 16 were conducted in low/middle income countries. Eleven (69%) studies did not enrol a representative spectrum of patients (consecutive, ambulatory HIV-positive patients suspected of having active TB. The majority of studies satisfied the remaining QUADAS criteria assessed.

Pooled sensitivity estimates were higher for TSPOT (72%, 95% CI 62-81%, 8 studies) than for QFT-GIT (60%, 95% CI 47-75%, 8 studies) (Figure 19). However, results varied widely across studies, resulting in significant heterogeneity in the pooled estimates for both IGRAs (I-squared >70% and p<0.001).
Sensitivity of IGRAs in HIV-positive individuals with confirmed active TB

### A. Low/Middle Income Countries

**TSPOT**

- Cattamanchi 2010, Uganda: 54% (45, 64)
- Dheda (a) 2009, South Africa: 100% (48, 100)
- Jiang 2009, China: 66% (47, 81)
- Leidl (a) 2009, Uganda: 89% (67, 99)
- Ling (a) 2010, South Africa: 82% (66, 92)
- Markova (a) 2009, Bulgaria: 62% (32, 86)
- Oni 2010, South Africa: 68% (57, 78)
- Rangaka 2010, South Africa: 63% (53, 72)

**Pooled Estimate (I-squared 73%, p<0.001)**: 72% (62, 81)

**QFT-GIT**

- Aabye 2009, Tanzania: 65% (52, 76)
- Baba 2008, South Africa: 58% (28, 85)
- Dheda (b) 2009, South Africa: 20% (1, 72)
- Kabeer 2009, India: 66% (50, 80)
- Leidl (b) 2009, Uganda: 68% (49, 91)
- Ling (b) 2010, South Africa: 67% (50, 80)
- Markova (b) 2009, Bulgaria: 92% (64, 100)
- Veldsman 2009, South Africa: 30% (15, 49)

**Pooled Estimate (I-squared 77%, p<0.001)**: 61% (47, 75)

### B. High Income Countries

**TSPOT**

- Clark 2007, United Kingdom: 94% (73, 100)

**QFT-GIT**

- Sauzullo 2010, Italy: 67% (47, 83)

The forest plots display the sensitivity estimates obtained from individual studies and pooled estimates derived from random effects modeling. Pooled estimates are shown only for sub-groups in which 4 or more studies were available. Abbreviations: IGRA, interferon-gamma release assay; CI, confidence interval; TSPOT, T-SPOT.TB, QFT-GIT, QuantiFERON-Gold In-tube.

Five studies compared head-to-head the sensitivity of IGRAs and TST for diagnosis of active TB in HIV-infected adults. TSPOT was more sensitive than TST in one study (absolute difference 50%, 95% CI 29-71%), less sensitive than TST in one study (absolute difference 18%, 95% CI 2-34%), and as sensitive as TST in one study (absolute difference -3%, 95% CI -17% to +11%). Similarly, QFT-GIT was more sensitive than TST in one study (absolute difference 41%, 95% CI 22-60%) and less sensitive than TST in another study (absolute difference 33%, 85% CI 16-51%).

### 3.3.7 Agreement between IGRA and TST results

Data on agreement (concordance) between TST and IGRA results in HIV-positive individuals being evaluated for LTBI were available for 15 studies, of which five were done in low- or middle-income countries. Three (60%) studies did not enrol a representative spectrum of patients (consecutive, ambulatory HIV-positive patients being screened for LTBI). The majority of studies satisfied the remaining QUADAS criteria assessed.
Of three studies that reported test agreement using kappa values, two reported poor or moderate agreement (kappa 0.4-0.6) and one reported strong agreement (kappa >0.6). When pooled across studies, TSPOT and TST results were concordant in 77% (95% CI 67-88%) of cases but there was significant heterogeneity among individual studies ($I^2$ 63%, p=0.04) (Figure 20). There were insufficient studies to calculate pooled estimates for QFT-GIT.

### Figure 20. Percent concordance between IGRA and TST results

#### A. Low/Middle Income Countries

<table>
<thead>
<tr>
<th>IGRA</th>
<th>Country</th>
<th>TST</th>
<th>Concordance</th>
<th>CI</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSPOT</td>
<td>Hoffmann (b) 2007</td>
<td>sub-Saharan Africa</td>
<td>90 (76, 97)</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jiang 2009</td>
<td>China</td>
<td>71 (58, 81)</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandalakas 2008</td>
<td>South Africa</td>
<td>79 (49, 95)</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oni 2010</td>
<td>South Africa</td>
<td>70 (59, 80)</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 63%, p=0.04)</strong></td>
<td></td>
<td></td>
<td>77 (67, 88)</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>QFT-GIT</td>
<td>Balcels 2008</td>
<td>Chile</td>
<td>91 (84, 96)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### B. High Income Countries

<table>
<thead>
<tr>
<th>IGRA</th>
<th>Country</th>
<th>TST</th>
<th>Concordance</th>
<th>CI</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSPOT</td>
<td>Hoffmann (a) 2007</td>
<td>Switzerland</td>
<td>100 (91, 100)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Richeldi (a) 2009</td>
<td>Italy</td>
<td>93 (86, 97)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rivas (a) 2009</td>
<td>Spain</td>
<td>78 (56, 93)</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stephan 2008</td>
<td>Germany</td>
<td>76 (70, 81)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Talati (a) 2009</td>
<td>USA</td>
<td>94 (90, 97)</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 92%, p&lt;0.001)</strong></td>
<td></td>
<td></td>
<td>89 (81, 98)</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>QFT-GIT</td>
<td>Jones 2007</td>
<td>USA</td>
<td>93 (88, 96)</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Luetkemeyer 2007</td>
<td>USA</td>
<td>89 (84, 93)</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Richeldi (b) 2009</td>
<td>Italy</td>
<td>95 (90, 98)</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rivas (b) 2009</td>
<td>Spain</td>
<td>91 (72, 99)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Talati (b) 2009</td>
<td>USA</td>
<td>96 (92, 98)</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 38%, p=0.17)</strong></td>
<td></td>
<td></td>
<td>94 (91, 96)</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

### 3.3.8 Indeterminate IGRA results

Data on the proportion of indeterminate IGRA results among HIV-positive individuals being evaluated for LTBI were available for 23 studies, of which nine (39%) were done in low- or middle-income countries. The proportion of indeterminate was ≤5% in 4 of 6 studies evaluating T-SPOT (range 0-10%) and two of three studies evaluating QFT-GIT (range 0-11%).

For TSPOT, the pooled proportion of indeterminate results was 2% (95% CI 0-3%) and results were consistent across studies ($I^2$ 0%, p=0.42). There were insufficient studies to calculate pooled estimates for QFT-GIT.
3.3.9 Impact of immunosuppression

In 21 studies, IGRA results were available for at least five HIV-infected adults in each of the following CD4+ cell count strata: <200 and ≥200 cells/μl. Among the seven (33%) studies conducted in low/middle income countries, the proportion of positive IGRA results was higher in persons with CD4+ cell count ≥200 cells/μl than in persons with CD4+ cell count <200 cells/μl in four of five studies evaluating TSPOT and both studies evaluating QFT-GIT. In contrast, the proportion of indeterminate IGRA results was higher in persons with CD4+ cell count <200 cells/μl than in persons with CD4+ cell count ≥200 cells/μl in four of 12 studies evaluating TSPOT and six of nine studies evaluating QFT-GIT.

For TSPOT, the pooled proportion of positive test results was significantly lower in individuals with CD4+ cell count <200 cells/μl compared with individuals with CD4+ cell count ≥200 cells/μl (difference -18%, 95% CI -34% to -2%) (Figure 22A). However, the pooled proportion of individuals with indeterminate test results was similar among individuals in the two CD4+ cell count strata (4%, -3% to 10%) (Figure 22B). There were insufficient studies to calculate pooled estimates for QFT-GIT.
## Use of IGRAs for the diagnosis of LTBI in HIV-infected individuals

### Impact of immunosuppression on IGRA results

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Difference in % Positive (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Low/Middle Income Countries</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TSPOT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoffmann (b) 2007</td>
<td>sub-Saharan Africa</td>
<td>-26 (-53, 0)</td>
<td>21</td>
</tr>
<tr>
<td>Jiang 2009</td>
<td>China</td>
<td>-27 (-61, 8)</td>
<td>15</td>
</tr>
<tr>
<td>Leidl (a) 2009</td>
<td>Uganda</td>
<td>5 (-15, 25)</td>
<td>27</td>
</tr>
<tr>
<td>Mandalakas 2008</td>
<td>South Africa</td>
<td>-18 (-62, 26)</td>
<td>11</td>
</tr>
<tr>
<td>Oni 2010</td>
<td>South Africa</td>
<td>-31 (-53, -9)</td>
<td>26</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 44%, p=0.13)</strong></td>
<td></td>
<td>-18 (-34, -2)</td>
<td>100</td>
</tr>
<tr>
<td><strong>QFT-GIT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balcells 2008</td>
<td>Chile</td>
<td>-2 (-20, 17)</td>
<td>51</td>
</tr>
<tr>
<td>Leidl (b) 2009</td>
<td>Uganda</td>
<td>-23 (-43, -4)</td>
<td>49</td>
</tr>
<tr>
<td><strong>TST</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jiang 2009</td>
<td>China</td>
<td>-35 (-59, -11)</td>
<td>50</td>
</tr>
<tr>
<td>Oni 2010</td>
<td>South Africa</td>
<td>15 (-11, 41)</td>
<td>50</td>
</tr>
<tr>
<td><strong>B. High Income Countries</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TSPOT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clark 2007</td>
<td>United Kingdom</td>
<td>-18 (-45, 9)</td>
<td>2</td>
</tr>
<tr>
<td>Dheda 2005</td>
<td>United Kingdom</td>
<td>-4 (-34, 26)</td>
<td>1</td>
</tr>
<tr>
<td>Hoffmann (a) 2007</td>
<td>Switzerland</td>
<td>0 (-19, 19)</td>
<td>3</td>
</tr>
<tr>
<td>Richeldi (a) 2009</td>
<td>Italy</td>
<td>1 (-9, 11)</td>
<td>12</td>
</tr>
<tr>
<td>Rivas (a) 2009</td>
<td>Spain</td>
<td>-11 (-45, 22)</td>
<td>1</td>
</tr>
<tr>
<td>Stephan 2008</td>
<td>Germany</td>
<td>-8 (-21, 5)</td>
<td>7</td>
</tr>
<tr>
<td>Talati (a) 2009</td>
<td>USA</td>
<td>-3 (-7, 1)</td>
<td>75</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 0%, p=0.70)</strong></td>
<td></td>
<td>-3 (-7, 0)</td>
<td>100</td>
</tr>
<tr>
<td><strong>QFT-GIT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aichelburg 2009</td>
<td>Austria</td>
<td>-4 (-6, -1)</td>
<td>25</td>
</tr>
<tr>
<td>Brock 2006</td>
<td>Denmark</td>
<td>-3 (-7, 1)</td>
<td>18</td>
</tr>
<tr>
<td>Jones 2007</td>
<td>USA</td>
<td>-7 (-12, -2)</td>
<td>14</td>
</tr>
<tr>
<td>Luetkemeyer 2007</td>
<td>USA</td>
<td>-9 (-14, -4)</td>
<td>15</td>
</tr>
<tr>
<td>Richeldi (b) 2009</td>
<td>Italy</td>
<td>-5 (-12, 3)</td>
<td>9</td>
</tr>
<tr>
<td>Rivas (b) 2009</td>
<td>Spain</td>
<td>-18 (-52, 16)</td>
<td>1</td>
</tr>
<tr>
<td>Talati (b) 2009</td>
<td>USA</td>
<td>0 (-3, 4)</td>
<td>19</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 48%, p=0.07)</strong></td>
<td></td>
<td>-4 (-7, -2)</td>
<td>100</td>
</tr>
<tr>
<td><strong>TST</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jones 2007</td>
<td>USA</td>
<td>-8 (-14, -3)</td>
<td>39</td>
</tr>
<tr>
<td>Luetkemeyer 2007</td>
<td>USA</td>
<td>-6 (-12, -0)</td>
<td>35</td>
</tr>
<tr>
<td>Richeldi 2009</td>
<td>Italy</td>
<td>-6 (-15, 2)</td>
<td>16</td>
</tr>
<tr>
<td>Stephan 2008</td>
<td>Germany</td>
<td>-2 (-12, 9)</td>
<td>10</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared %, p=0.67)</strong></td>
<td></td>
<td>-7 (-10, -3)</td>
<td>100</td>
</tr>
</tbody>
</table>

* Difference = (% positive CD4 <200 cells/ul) – (% positive CD4 ≥200 cells/ul)
Use of IGRAs for the diagnosis of LTBI in HIV-infected individuals

B. Proportion of indeterminate IGRA results

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Difference in % Indeterminate (95% CI)*</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Low/Middle Income Countries</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSPOT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoffmann (b) 2007</td>
<td>sub-Saharan Africa</td>
<td>-6 (-30, 18)</td>
<td>7</td>
</tr>
<tr>
<td>Jiang 2009</td>
<td>China</td>
<td>0 (-14, 14)</td>
<td>20</td>
</tr>
<tr>
<td>Leidl (a) 2009</td>
<td>Uganda</td>
<td>3 (-5, 12)</td>
<td>48</td>
</tr>
<tr>
<td>Mandalakas 2008</td>
<td>South Africa</td>
<td>14 (15, 43)</td>
<td>5</td>
</tr>
<tr>
<td>Oni 2010</td>
<td>South Africa</td>
<td>8 (-5, 0.22)</td>
<td>21</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 0%, p=0.75)</strong></td>
<td></td>
<td>4 (-3, 10)</td>
<td>100</td>
</tr>
<tr>
<td><strong>QFT-GIT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balcells 2008</td>
<td>Chile</td>
<td>0 (-9, 9)</td>
<td></td>
</tr>
<tr>
<td>Leidl (b) 2009</td>
<td>Uganda</td>
<td>-1 (-8, 6)</td>
<td></td>
</tr>
<tr>
<td><strong>B. High Income Countries</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSPOT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clark 2007</td>
<td>United Kingdom</td>
<td>-3 (-14, 8)</td>
<td>12</td>
</tr>
<tr>
<td>Dheda 2005</td>
<td>United Kingdom</td>
<td>-6 (-22, 11)</td>
<td>6</td>
</tr>
<tr>
<td>Hoffmann (a) 2007</td>
<td>Switzerland</td>
<td>4 (-27, 36)</td>
<td>2</td>
</tr>
<tr>
<td>Richeldi (a) 2009</td>
<td>Italy</td>
<td>-2 (-9, 5)</td>
<td>29</td>
</tr>
<tr>
<td>Rivas (a) 2009</td>
<td>Spain</td>
<td>0 (-20, 20)</td>
<td>4</td>
</tr>
<tr>
<td>Stephan 2008</td>
<td>Germany</td>
<td>2 (-5, 9)</td>
<td>29</td>
</tr>
<tr>
<td>Talati (a) 2009</td>
<td>USA</td>
<td>8 (-1, 17)</td>
<td>19</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 8%, p=0.37)</strong></td>
<td></td>
<td>1 (-3, 5)</td>
<td>100</td>
</tr>
<tr>
<td><strong>QFT-GIT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aichelburg 2009</td>
<td>Austria</td>
<td>16 (9, 22)</td>
<td>18</td>
</tr>
<tr>
<td>Brock 2006</td>
<td>Denmark</td>
<td>6 (-1, 14)</td>
<td>17</td>
</tr>
<tr>
<td>Jones 2007</td>
<td>USA</td>
<td>23 (10, 35)</td>
<td>11</td>
</tr>
<tr>
<td>Luetkemeyer 2007</td>
<td>USA</td>
<td>9 (1, 17)</td>
<td>17</td>
</tr>
<tr>
<td>Richeldi (b) 2009</td>
<td>Italy</td>
<td>4 (-9, 17)</td>
<td>10</td>
</tr>
<tr>
<td>Rivas (b) 2009</td>
<td>Spain</td>
<td>-3 (-23, 17)</td>
<td>6</td>
</tr>
<tr>
<td>Talati (b) 2009</td>
<td>USA</td>
<td>4 (0, 9)</td>
<td>21</td>
</tr>
<tr>
<td><strong>Pooled Estimate (I-squared 67%, p=0.006)</strong></td>
<td></td>
<td>9 (3, 15)</td>
<td>100</td>
</tr>
</tbody>
</table>

* Difference = (% indeterminate CD4 <200 cells/ul) – (% indeterminate CD4 ≥200 cells/ul)

Data on the impact of immunosuppression on TST results were available for two studies. One study found that the proportion of positive TST results was 35% (95% CI 11-59%) lower in persons with CD4+ cell count <200cells/μl compared to persons with CD4+ cell count ≥200 cells/μl. Data from the second study trended in the opposite direction: the proportion of positive TST results was 15% (95% CI -11% to +41%) higher in persons with CD4+ cell count <200cells/μl compared to persons with CD4+ cell count ≥200cells/μl.

**3.3.10 Strengths and limitations of the evidence base**

The major limitation was the lack of an adequate reference standard to evaluate the accuracy of IGRAs for diagnosis of LTBI. The majority of studies were small (< 100 patients in 12 of 22 studies), only five studies performed a head-to-head comparison of IGRA and TST results to a reference standard, and there were insufficient studies to perform meta-analysis in many sub-groups.
Given that both TST and IGRAs have suboptimal sensitivity and that discordant results are common, it would be relevant to evaluate outcomes when both tests are used, either simultaneously or sequentially, for diagnosing LTBI in HIV-infected persons.

3.3.11 Research gaps
Important key questions remain unanswered despite the substantial body of literature on IGRAs. HIV-infected individuals with a negative IGRA result may have a low risk of progression to active TB, but this result should be confirmed in larger studies that simultaneously perform TST and include a longer duration of follow-up. IGRAs (particularly TSPOT) may be more sensitive than TST in HIV-infected individuals and less affected by advanced immunosuppression. However, these results were not observed consistently in head-to-head comparisons. Future studies should focus on treatment outcomes in HIV-infected individuals randomized to receive IPT based on IGRA results and evaluate the incidence of TB in HIV-infected individuals with discordant IGRA and TST results.

3.3.12 Summary of findings and GRADE evidence profiles
- Although WHO recently endorsed IPT as one of three key public health strategies to reduce the impact of TB on persons living with HIV the optimal test for identifying HIV-infected persons who could benefit from IPT remains an unanswered question;
- The majority of persons latently infected with TB, including persons co-infected with HIV, do not develop active TB. The clinical utility of any diagnostic test for LTBI is therefore dependent on its ability to identify which persons are truly at increased risk for progression to active TB and could benefit from IPT;
- All three studies of the predictive value of IGRAs in HIV-infected individuals showed that IGRAs have poor positive predictive value but high negative predictive value for active TB. While these results suggest that a negative IGRA result is reassuring (no person with a negative IGRA result developed culture-positive TB), the studies had serious limitations, including small sample sizes with short-duration of follow-up and differential evaluation and/or follow-up of persons with positive and negative IGRA results;
- Large prospective cohort studies have established that persons with a positive TST have a 1.4 to 1.7-fold higher rate of active TB within one year compared to persons with a negative TST result. Randomised controlled trials in HIV-infected persons demonstrated that IPT confers a 20-60% reduction in the risk of active TB and that this reduction occurs only in persons with positive TST results;
- In spite of limited data on predictive value, it has been suggested that IGRAs may have a role for identifying TB infection in HIV-infected individuals given the known decreased performance of TST in immunosuppressed persons. However, neither IGRA was consistently more sensitive than TST in head-to-head comparisons. Data on the impact of immunosuppression on IGRA validity remains unclear.

3.3.13 Final Recommendations
The GRADE evidence profiles are provided in Tables 14 and 15. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRAS in individuals living with HIV infection was very low and recommended that these tests should not be used as a replacement for TST for the assessment of LTBI (strong recommendation).

- This recommendation also applies to HIV-positive children based on the generalisation of data from adults;
<table>
<thead>
<tr>
<th>OVERALL QUALITY OF EVIDENCE</th>
<th>VERY LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRENGTH OF RECOMMENDATION</td>
<td>STRONG</td>
</tr>
</tbody>
</table>
Table 14. GRADE evidence profile: The role of IGRAs in the diagnosis of latent tuberculosis infection in HIV-infected individuals

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Outcome: Predictive value of IGRAs for active TB</td>
<td>1100 (3)&lt;sup&gt;a1&lt;/sup&gt; LMIC: 306 (1)</td>
<td>Prospective cohort</td>
<td>Serious&lt;sup&gt;a2&lt;/sup&gt; (-1)</td>
<td>Serious&lt;sup&gt;a3&lt;/sup&gt; (-1)</td>
<td>None&lt;sup&gt;a4&lt;/sup&gt;</td>
<td>Likely&lt;sup&gt;a6&lt;/sup&gt;</td>
<td>Very Low @@@@@</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>B. Outcome: Sensitivity for active TB (as a surrogate reference standard for LTBI)</td>
<td>1523 (18)&lt;sup&gt;b1&lt;/sup&gt; LMIC: 1056 (16)</td>
<td>Mainly cross-sectional</td>
<td>No serious limitations&lt;sup&gt;b2&lt;/sup&gt; (-1)</td>
<td>Very Serious&lt;sup&gt;b4&lt;/sup&gt; (-2)</td>
<td>Serious&lt;sup&gt;b5&lt;/sup&gt; (-1)</td>
<td>Likely&lt;sup&gt;b6&lt;/sup&gt;</td>
<td>Very Low @@@@@</td>
<td>Important (4-6)</td>
</tr>
<tr>
<td>C. Outcome: Concordance with TST</td>
<td>2158 (15)&lt;sup&gt;c1&lt;/sup&gt; LMIC: 401 (5)</td>
<td>Cross-sectional</td>
<td>No serious limitations&lt;sup&gt;c2&lt;/sup&gt; (-2)</td>
<td>Serious&lt;sup&gt;c4&lt;/sup&gt; (-1)</td>
<td>None&lt;sup&gt;c5&lt;/sup&gt; (-1)</td>
<td>Likely&lt;sup&gt;c6&lt;/sup&gt;</td>
<td>Very Low @@@@@</td>
<td>Important (4-6)</td>
</tr>
</tbody>
</table>

Footnotes

*Quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high quality observational studies (cross-sectional or cohort studies enrolling patients with diagnostic uncertainty) and at moderate when these types of studies were absent. One point was subtracted when there was a serious issue identified or two points when there was a very serious issue identified in any of the criteria used to judge the quality of evidence.

<sup>a1</sup>Three longitudinal studies that evaluated the ability of IGRAs to predict future development of active TB were identified. Two were conducted in high income countries (Austria and UK) and one in a low/middle income country (Cambodia).

<sup>a2</sup>Based on the Newcastle-Ottawa scale, the study samples were considered to be representative. However, only one study had an adequate duration of follow-up (≥1 year), all three studies scored poorly on outcome assessment did not adequately rule-out active TB at baseline or did not adequately evaluate all participants for active TB during follow-up, and all three studies had very few incident TB cases.

<sup>a3</sup>Two studies were carried out in high income countries; hence the findings may not be generalizable to low/middle income countries.

<sup>a4</sup>All three studies found that the risk of active TB was higher in IGRA positive compared to IGRA negative patients; but risk of progression to active TB was low in all groups.
The number of incident TB cases was small in all studies, leading to wide confidence intervals for risk estimates. In the two studies that reported cumulative incidence of TB, the difference in cumulative incidence of TB between IGRA positive and IGRA negative persons was not statistically significant.

Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. Some degree of publication bias was assumed likely because: 1) literature on IGRA is rapidly exploding and currently unpublished studies may come out in future (despite an attempt to be comprehensive and include unpublished studies); 2) there are anecdotal examples of unpublished negative studies on IGRA; and 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial. However, we did not deduct points for this factor.

18 studies were identified: 9 evaluated TSPOT and 9 evaluated QFT-GIT.

Study limitations were evaluated using the QUADAS tool. 12 (67%) studies did not enroll a representative spectrum of patients (ambulatory HIV-infected patients suspected of having active TB). The majority of studies satisfied the remaining QUADAS criteria assessed.

16 (89%) studies were conducted in low/middle income countries. However, sensitivity for active TB may not reflect performance for LTBI and diagnostic accuracy is only a surrogate for patient-important outcomes.

There was significant heterogeneity in sensitivity estimates for both TSPOT (range 54-100%, I^2 73%, p<0.002) and QFT-GIT (range 20-92%, I^2 78%, p<0.001) in low/middle income countries.

The 95% confidence interval for pooled sensitivity was within +/10% in most sub-groups.

Data included in our review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. However, no points were deducted as additional negative studies were unlikely to bias the principal finding (sub-optimal IGRA sensitivity).

15 studies were identified: 9 evaluated TSPOT and 6 evaluated QFT-GIT.

Study limitations were evaluated using the QUADAS scale. A majority of studies satisfied all QUADAS criteria assessed.

Only 5 of 9 studies for TSPOT and 1 of 6 studies for QFT-GIT were conducted in low/middle income countries. In addition, concordance between IGRA and TST is a poor surrogate for patient-important outcomes.

Among studies conducted in low/middle income countries, there was significant heterogeneity in estimates of percent concordance between IGRA and TST for TSPOT (range 70-90%, I^2 63%, p<0.04). There was only 1 study of QFT-GIT (concordance 91%).

The 95% confidence interval for pooled concordance was within +/10% in most sub-groups.

Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could be ruled out. Some degree of publication bias was assumed likely because: 1) literature on IGRA is rapidly exploding and currently unpublished studies may come out in future (despite an attempt to be comprehensive and include unpublished studies); 2) there are anecdotal examples of unpublished negative studies on IGRA; and 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial. However, no points were deducted for this factor.
**Table 15. GRADE Summary of Findings – Role of IGRAs for diagnosis of LTBI in HIV infected individuals**

**Review question:** What is the role of IGRAs in the diagnosis of latent tuberculosis infection (LTBI) in HIV-infected individuals?  
**Patients/population:** HIV-infected active TB suspects or HIV-infected persons being screened for LTBI; all ages, all countries (data specific to low- and middle-income countries presented when available).  
**Setting:** Outpatients and inpatients.  
**Index test:** QuantiFERON-Gold in-tube [QFT-GIT] and T-SPOT.TB [TSPOT].  
**Importance:** The performance IGRAs in diagnosing LTBI among HIV-infected individuals is uncertain; it is unclear if IGRAs should be used to identify HIV-infected persons with LTBI who could benefit from preventive therapy.  
**Reference standard:** See hierarchy of reference standards (Fig 1)  
**Studies:** Randomized controlled trials, observational studies (cohort, cross-sectional, case-control)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>N</th>
<th>Principal Findings</th>
<th>What do these findings mean?</th>
<th>Quality of Evidence</th>
<th>Importance</th>
</tr>
</thead>
</table>
| Predictive value for active TB | 1100 (3 studies) | 1) TSPOT: Cumulative incidence of active TB higher in IGRA+ compared to IGRA- individuals, but difference not statistically significant (10% vs. 0%, risk difference 10%, 95% CI -3% to +23%).  
2) QGT-GIT: Cumulative incidence of active TB higher in IGRA+ compared to IGRA- individuals, but difference not statistically significant (8% vs. 0%, risk difference 8%, 95% CI -0.7% to 17%). | IGRA+ individuals may have a higher risk of progression to active TB than IGRA- individuals, but the risk of progression is low in both groups. | Very Low  
©©©©© | Critical  
(7-9) |
| Sensitivity for active TB (a surrogate reference standard for LTBI) | 1523 (18 studies) | 1) TSPOT: Pooled sensitivity 72% (95% CI 62-81%); TSPOT more sensitive than TST in 1 study, less sensitive in 1 study, and as sensitive in 1 study.  
2) QFT-GIT: Pooled sensitivity was 61% (95% CI 47-75%). Compared to TST, QFT-GIT more sensitive in 1 study and less sensitive in 1 study. | In low- and middle-income countries, IGRAs have suboptimal sensitivity for active TB and do not consistently have higher sensitivity than TST. | Very Low  
©©©© | Important  
(4-6) |
| Concordance with TST           | 1822 (14 studies) | 1) TSPOT: Pooled concordance 77% (95% CI 67-88%).  
2) QFT-GIT: 1 study; concordance 91%. | In low- and middle-income countries, IGRAs have moderate concordance with TST. | Very Low  
©©©© | Important  
(4-6) |
Use of IGRAs for screening of health care workers

3.4 Use of IGRAs for screening of health care workers

3.4.1 Objectives, reference standards and outcomes

Objectives

To assess IGRA test performance in screening HCWs using cross-sectional, longitudinal and serial testing studies. Secondary objectives were to:

a) determine if IGRAs are better correlated than the TST with occupational exposure to TB in cross sectional studies;

b) estimate the rates of IGRA conversions and reversions, and assess whether IGRA conversions are more closely associated with recent occupational exposure than TST conversions;

c) to summarise evidence produced by cost-effectiveness analyses and programmatic studies.

Reference standards

A major challenge for studies evaluating the performance of IGRAs is the lack of gold standard for LTBI. Consequently, studies were evaluated using prevalence and incidence of LTBI, correlation between IGRA results and an exposure gradient, agreement with TST results.

Primary outcomes

In cross-sectional studies:

a) Prevalence of positive TST (ie. LTBI prevalence) versus positive IGRA and associated risk factors;

b) Concordance (ie. agreement) between TST and IGRA results and factors associated with concordance and discordance;

In longitudinal studies:

a) Incidence of TST and IGRA conversions and risk factors for conversions,

b) Incidence of TST and IGRA reversions and risk factors for reversions.

3.4.2 Search results

Forty-two cross-sectional, longitudinal and serial testing designs using a commercial IGRA assay: QuantiFERON-TB Gold® or In-Tube version and the T-SPOT.TB, for TB screening in HCWs in any setting were evaluated (Annex 5). Overall, only five (12%) were done in high incidence settings, while 37 (88%) were done in intermediate or low-incidence countries.

The following studies were excluded: 1) case reports and case series, 2) studies with 10 or fewer participants; 3) reviews and commentaries; 4) letters that did not report original data; 5) studies evaluating the use of IGRAs for treatment monitoring in HCWs (ie: not for diagnostic purposes); 6) short-term serial testing studies (serial testing within one month) and reproducibility studies (which have been systematically reviewed recently; 7) non-commercial/in-house assays; and 8) IGRA testing in the context of known nosocomial outbreaks or point-source exposure (as it was expected that these studies would report higher rates of conversions and reversions by both tests, and that results may not reflect the typical level of LTBI prevalence or conversions in an occupational environment.

Page | 73
3.4.3 Study characteristics

Studies varied greatly in their design, execution and outcomes. No meta-analyses were performed as methods are not defined for heterogeneous diagnostic studies with no gold standard. IGRA performance varies across populations, therefore, all results were stratified by TB incidence in the countries where the studies were done (high vs intermediate & low incidence). High incidence countries were defined as countries with more than 100 estimated incident TB cases per year/per 100,000 population as reported to WHO. However, due to the variety of study designs and HCW screening guidelines, even within the strata, study populations included HCWs with varying risk of TB exposure.

Studies were assessed by set criteria which included study design, participants, country, period of recruitment, proportion BCG vaccinated, IGRA methods, TST methods and outcome data, including: baseline TST and IGRA positivity rates, indeterminate rates, concordance between TST and IGRA (agreement and kappa value), predominant type of discordance and correlations found between risk factors and test results.

3.4.4 Study quality

Because IGRA studies in HCWs do not use the conventional diagnostic study design for sensitivity and specificity estimation, the QUADAS checklist was not used to evaluate quality. Selected study features were chosen as quality indicators. These included study design (cross-sectional vs longitudinal), use of standardised, commercial assays, use of standardised tuberculin material (PPD-S in North America, RT23 in Europe) for TST, proportion of indeterminate results and the duration of follow-up in longitudinal studies.

3.4.5 IGRA vs. TST positivity rates in high-incidence countries

Three cross-sectional studies were examined comparing IGRA and TST performance in HCWs in India, Russia, and Vietnam, although TST was not performed in the Russian study. TST and IGRA positivity rates were high in HCWs, ranging from 40% to 66% (Figure 23). IGRA positivity was slightly lower than TST positivity in the two studies comparing TST and IGRA; however, the difference in estimated prevalence was significant only in the study from Vietnam. The Vietnam study also reported the lowest rate of BCG vaccination among participants at 37.3%, compared to 71% vaccinated in the India study.
3.4.6 IGRA conversion and reversion rates in longitudinal, serial testing

Two serial testing studies were identified. One study conducted repeat testing at 0, 6, 12 months and the other tested at 0 and 18 months. Rates of IGRA conversions from these studies ranged from 11.6% to 21%. One of these studies calculated the TST conversion rate, and found 4% conversion after 18 months. One study found that conversion rates varied for both the TST and the IGRA when different cut-offs were used. Neither study reported data to suggest that IGRA conversions were better associated with TB exposure than TST conversions.

Reversion rates in one study ranged from 27% in the first 6 month period to 40% in the second 6 month period. A second study reported 18-month IGRA reversion rates around 7% among baseline concordant positives, but up to 70% among those with discordant baseline results (ie: TST negative/IGRA positive).

Overall, serial testing data from low-incidence countries suggest that IGRA results vary greatly during serial testing, and that rates of conversions may vary depending on the test used and the cut-off used to define conversions. When simple negative/positive changes are used as cut-offs, IGRA had a higher conversion rate than the TST.

There are no data to show that IGRA performs better than TST in identifying incidence of new TB infections.
3.4.7 Association between occupational risk factors and test results in HCWs

One study showed a stronger association between occupational risk factors and IGRA rather than the TST although confidence intervals overlapped. In studies done in high-income countries, being of foreign birth or having lived in a high TB incidence country was correlated with either IGRA or TST positivity.

3.4.8 Cost-effectiveness of IGRAs in HCW screening

Limited data exist on cost-effectiveness of IGRAs when used for HCW screening and all studies have been conducted in low-incidence settings and none accounted for serial testing of HCWs with IGRAs.

3.4.9 Strengths and limitations of the evidence base

The systematic review used a comprehensive search strategy using multiple sources and databases to retrieve relevant studies, including unpublished studies and conference proceedings. Only two studies in low- or middle-income countries were identified. Serial testing data, evidence on the predictive value of IGRAs in HCWs, as well as reproducibility data are still limited even in low-incidence settings.

3.4.10 Research gaps

Given the repeated nature of routine screening of HCWs (eg. annual testing), there are particular issues which may not be relevant in routine practice or contact investigations but become very important during repeated screening. These issue need to be explored in properly designed prospective studies and include test reproducibility, performance of IGRAs when repeated frequently, interpretation of discordant TST and IGRA results, and the IFN-g thresholds (cut-off values) which most accurately distinguish new TB infection (ie. conversion) from random variation.

3.4.11 Summary of findings and GRADE evidence profiles

- Prevalence of LTBI in HCWs depends on the test used and the particular TB incidence setting;

- Both the TST and IGRAs appear to be associated with markers of TB exposure, but the magnitude of associations varies; TST performance is associated with BCG vaccination, while IGRA performance seems to be unaffected.

- IFN-g responses seem to have natural variation and tend to fluctuate around the cut-off, causing apparent conversions and reversions. The exact cause of the conversions and reversions remains unclear, and might indicate spontaneous clearance of TB infection, or dynamic changes within the spectrum of latent TB infection.

- The use of IGRAs for serial testing is complicated by lack of data on optimum cut-offs for serial testing, and unclear interpretation and prognosis of conversions and reversions.

- Conversion rates are highest when a simple negative to positive change is used to define a conversion. This is true in both high and low incidence settings and has implications for deciding on criteria (cut-offs) for conversions and reversions.

- There are no data to show that IGRAs perform better at identifying incidence of new TB infections among HCWs than the TST, irrespective of HIV status.
3.4.12 Final Recommendations

The GRADE evidence profiles are provided in Tables 16 and 17. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRAS for screening in health care workers in low- and middle-income countries was very low and recommended that these tests should not be used in health care worker screening programmes.

The Expert Group also noted the lack of WHO policy on using the TST in health care worker screening programmes.

<table>
<thead>
<tr>
<th>OVERALL QUALITY OF EVIDENCE</th>
<th>VERY LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRENGTH OF RECOMMENDATION</td>
<td>STRONG</td>
</tr>
</tbody>
</table>
Table 16. GRADE evidence profile: Interferon-γ release assays for tuberculosis screening of healthcare workers in low and middle income countries

<table>
<thead>
<tr>
<th>No of participants (studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A. Efficacy of preventive therapy based on IGRA test results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td><strong>B. Predictive value of IGRA for active TB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td><strong>C. Outcome: Correlation of IGRA results with occupational TB exposure</strong></td>
<td>991 (2) A1</td>
<td>Cross-sectional</td>
<td>No serious limitations A2</td>
<td>No serious Indirectness A3</td>
<td>Serious A4 (-1)</td>
<td>Serious A5 (-1)</td>
<td>Likely A6</td>
<td>Low ⊕⊕⊕⊕</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>D. Outcome: Correlation between IGRA conversions and occupational TB exposure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td><strong>E. Outcome: Sensitivity for active TB (as a surrogate reference standard for LTBI)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Important (4-6)</td>
</tr>
<tr>
<td><strong>F. Outcome: Concordance between IGRAs and TST (cross-sectional)</strong></td>
<td>1,357 (4) B1</td>
<td>Cross-sectional</td>
<td>No serious limitations B2</td>
<td>Serious B3 (-1)</td>
<td>Serious B4 (-1)</td>
<td>Serious B5 (-1)</td>
<td>Likely B6</td>
<td>Very Low ⊕⊕⊕⊕</td>
</tr>
</tbody>
</table>
G. Outcome: concordance between IGRA and TST conversions (longitudinal)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Longitudinal</th>
<th>No serious limitations</th>
<th>Serious (-1)</th>
<th>No serious inconsistency</th>
<th>Very Serious (-2)</th>
<th>Likely</th>
<th>Very Low</th>
<th>Important (4-6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>216 (1)</td>
<td>Longitudinal</td>
<td>No serious limitations</td>
<td>Serious (-1)</td>
<td>No serious inconsistency</td>
<td>Very Serious (-2)</td>
<td>Likely</td>
<td>Very Low</td>
<td>Important (4-6)</td>
</tr>
</tbody>
</table>

Footnotes:

1. Quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: limitations, indirectness, inconsistency, imprecision, and publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high quality observational studies (cross-sectional or cohort studies with diagnostic uncertainty and direct comparison of test results with culture) and at moderate when these types of studies were absent. One point was subtracted when there was a serious issue identified or two points when there was a very serious issue identified in any of the criteria used to judge the quality of evidence.

A1 2 studies were identified evaluating an association between test positivity and occupational exposure to TB. These studies compared only QFT and the TST.

A2 Study limitations were assessed using select quality indicators. Studies satisfied majority of selected quality indicators.

A3 Some indirectness in the choice of reference standard was recognised although the studies were not downgraded for indirectness.

A4 Two studies evaluated the association between 5 variables of occupational exposure to TB and test positivity, estimates ranged from OR=1.28-5.09.

A5 Only 50% of estimates of association of test positivity and exposure reached statistical significance, 95% confidence intervals ranged from: 0.68-9.33. With only two studies, imprecision may be a concern.

A6 Data included in this review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. Although no points were deducted, some degree of publication bias was considered likely because: 1) literature on IGRA is rapidly exploding and currently unpublished studies may come out in future (despite an attempt to be comprehensive and include unpublished studies); 2) there are anecdotal examples of unpublished negative studies on IGRA; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

B1 4 cross-sectional studies were identified: 3 evaluated a previous version of the QFT, 1 study evaluated only the TSPOT.TB.

B2 Study limitations were assessed using select quality indicators as the QUADAS scale was not appropriate for concordance studies. Majority of studies satisfied selected quality indicators.

B3 Concordance between IGRA and the TST is a poor surrogate for patient important outcomes.

B4 Among studies conducted in low- and middle-income countries, there was moderate heterogeneity in estimates of percent agreement between TST and IGRA (Range: 50-81%).
Due to heterogeneity in effect estimates we could not pool concordance. However, confidence intervals for estimates of concordance for individual studies were wide, and with only 4 studies, imprecision may be a concern.

Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. Although no points were deducted, some degree of publication bias was considered likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future (despite an attempt to be comprehensive and include unpublished studies); 2) there are anecdotal examples of unpublished negative studies on IGRAs; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

1 longitudinal study was included which assessed concordance between TST and IGRA conversions, using the QFT test.

Study limitations were assessed using select quality indicators as the QUADAS scale was not appropriate for concordance studies. Both studies satisfied the majority of selected quality indicators.

This study was conducted in a low middle income country. Concordance between IGRA and the TST conversions is a poor surrogate for patient important outcomes, and may not be an appropriate reference standard.

This study estimated fair concordance between QFT and TST conversions (96%).

Only 1 study was identified with a small number of participants (n=216).

Data included in this review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. Although no points were deducted, some degree of publication bias was considered likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future (despite an attempt to be comprehensive and include unpublished studies); 2) there are anecdotal examples of unpublished negative studies on IGRAs; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.
Use of IGRAs for screening of healthcare workers

**Table 17. GRADE summary of findings – IGRAs for tuberculosis screening of healthcare workers in low and middle income countries**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. Participants</th>
<th>Principal Findings</th>
<th>What do these findings mean?</th>
<th>Quality of Evidence</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficacy of preventive therapy based on IGRA test results</td>
<td>No Studies in HCWs</td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>Predictive value of IGRA for active TB</td>
<td>No Studies in HCWs</td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>Correlation between IGRA positivity and occupational TB exposure</td>
<td>991 (2 studies)</td>
<td>1) TSPOT.TB: No studies evaluated TSPOT.TB 2) QFT: All 5 comparisons gave positive estimates for the association between test positivity and occupational exposure (OR=1.28-4.15), 3/5 reached statistical significance. 3) TST: All 5 comparisons gave positive effect estimates (OR=1.33-5.09), 2/5 reached statistical significance.</td>
<td>Data were limited on TSPOT.TB and from low and middle income settings. Occupational exposure was associated with positivity for both tests, although this was not always significant. There is no strong evidence that IGRAs are more strongly correlated with occupational TB exposure than TST.</td>
<td>Low</td>
<td>Critical (7-9)</td>
</tr>
</tbody>
</table>
## Use of IGRAs for screening of healthcare workers

<table>
<thead>
<tr>
<th>Correlation between IGRA conversions and occupational TB exposure</th>
<th>No Studies in HCWs</th>
<th></th>
<th>Critical (7-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity for active TB (as a surrogate reference standard for LTBI)</td>
<td>No Studies in HCWs</td>
<td></td>
<td>Important (4-6)</td>
</tr>
<tr>
<td>Concordance between TST and IGRAs</td>
<td>1,357 (4 studies)</td>
<td>In low and middle income studies, agreement between IGRA and TST results ranged from 50.2%–81.4%. While IGRA consistently estimated a lower rate, this difference was significant in only 2/4 cases.</td>
<td>Concordance was fair to poor in low and middle income settings. Both tests provide similar estimates of prevalence in low and middle income countries.</td>
</tr>
<tr>
<td>Concordance between IGRA and TST conversions</td>
<td>216 (1 study)</td>
<td>This study found 96% agreement between test conversions (QFT &amp; TST).</td>
<td>IGRA and TST conversions show moderate concordance. Data are limited in all settings.</td>
</tr>
</tbody>
</table>
3.5 Use of IGRAs in contact screening and outbreak investigations

3.5.1 Objectives, reference standards and outcomes

Objectives
The primary objective was to assess the accuracy and performance of IGRAs for the detection of latent TB infection (LTBI) in contacts and outbreak settings in low and middle-income settings. A secondary objective was to evaluate the factors (specifically TB exposure), associated with IGRA positivity in these settings.

Reference standards
TB exposure was defined as exposure to an infectious index case. TB exposure as a reference standard may be captured in a variety of ways, either as a dichotomous variable (exposed v. unexposed) or as a gradient (less exposed to more exposed using contact scores, measures of proximity or duration, household membership, sharing of a room or bed, etc.), based on a wide range of potential classifications.

Studies employing exposure gradients were classified broadly into two categories:

a) Exposure gradients based on microbiologic characteristics of the index case (eg. smear status, graded as 3+, 2+, and 1+)

b) Exposure gradients based on proximity/duration with the index case, including:
   - Close versus casual contact
   - Household member versus not
   - The so-called Senegal/Gambia gradient (same bed, same room, same house)
   - Hours spent with the index case

Studies employing exposure as dichotomous variables were grouped into two categories:

a) Smear-positive versus smear-negative index case
b) Exposed versus non-exposed

Outcomes
For studies using a dichotomous measure of exposure, both exposed and unexposed participants must have been represented in the sample population, and the TST and or IGRA tests could not have been used to classify exposure status of participants.

Apart from exposure gradient analysis, concordance between TST and IGRA results among contacts were evaluated, as well as the rates of conversions and reversions among contacts using both tests.

3.5.2 Search results
After a full text review of 99 studies, 65 studies conducted in high income countries were excluded, as were 18 studies using pre-commercial and in-house IGRAs. 16 studies (14 original manuscripts and 2 unpublished studies) involving 4,854 study participants were determined to meet the eligibility criteria (Annex 6).
3.5.3 Data analysis

Given the potential heterogeneity across populations and tests, and expected variations in how various studies constructed exposure gradients, a meta-analysis was not considered appropriate. A narrative synthesis was primarily used, involving detailed tables and plots, stratified by type of assay and type of study design (cross-sectional or longitudinal) as well as exposure category (dichotomous or exposure gradient).

3.5.4 Study characteristics

75% of studies were carried out in middle-income countries (South Africa, Lithuania, Brazil, Turkey, Taiwan, India, and Nigeria). The remaining studies were done in low-income countries (Cambodia, Gambia, Nepal and Zambia).

Most studies were small and ranged from 39-301 participants, the larger ones having been conducted in India, Gambia, South Africa and Brazil. However, the inclusion of a large unpublished contact study from Zambia doubled the total sample size (2,211 study participants). All studies included BCG vaccinated participants. HIV was frequently not reported, but when it was reported rates were low (0-1.5%) with the exception of the Zambia study which reported HIV infection rates around 38% in the (adult) study population, and one other study which reported a HIV infection rate of 5% (paediatric population).

Only one study did not include household contacts. This study evaluated health care workers exposed to a smear-positive TB case. The remaining 15 studies all included household contacts, while 3 studies also included school or work contacts. Nine (56%) of the included studies examined child contacts exclusively. A further 3 studies included both child and adult contacts while 4 studies include only adult contacts. Depending on the study design, most studies contained only known contacts of active TB cases; however 5 studies did recruit a comparison group with no known TB exposure.

3.5.5 Study quality

Because contact studies do not use conventional diagnostic study design for sensitivity and specificity estimation, the QUADAS checklist was not used to evaluate quality. Selected study features were chosen as quality indicators. These included study design (cross-sectional vs longitudinal), use of standardised, commercial assays, use of standardised tuberculin material for the TST (PPD-S in North America, RT23 in Europe), timing of IGRA blood draw and TST application, proportion of indeterminate results, reporting on inclusion/exclusion criteria, reporting on how participants were sampled, whether personnel performing test were blinded to other test results and duration of follow-up in longitudinal studies.

Quality was not assessed for the two unpublished studies as only preliminary reports were available. Among the 16 studies evaluated, 12 were cross-sectional in design. The included studies varied in quality, with several quality indicators frequently not reported. For example, only 3/14 studies reported whether study personnel had been blinded to other test results (or TB exposure) when performing and interpreting test results; 7/14 studies did not report the sequence of testing (eg. TST followed by IGRA); 5/15 studies (33%) reported some kind of industry involvement, most frequently the provision of test kits at no cost (n=4), while one study reported one of its authors having been a paid consultant of the manufacturer of the IGRA test kits evaluated (Table 18).

Study comparability: Only 4 studies presented adjusted odds ratios, and adjusted for different factors: All 4 adjusted for age, 3 also adjusted for sex, 2 adjusted for ethnicity as well as age and sex, and one study adjusted
for a number of additional factors. In the remaining studies which did not present adjusted estimates, residual confounding was a concern.

Ascertainment of test outcome: Two critical issues in performing and interpreting the tests (both IGRAs and the TST) were identified: Firstly, previous research has suggested a boosting effect on the IGRA if the TST is performed prior to blood being drawn. Seven studies did not report on the timing or sequence of testing, 6 studies reported that blood was drawn prior to the implementation of the TST (avoiding any chance of a boosting effect). Only one study reported performing the TST immediately prior to the blood draw and one study performed blood draw one month after the TST.

Secondly, only 3 studies reported that personnel had been blinded to previous test results during testing. Ascertainment bias could therefore not be ruled out for the remaining studies.

3.5.6 Agreement between IGRA and TST Results

Data on agreement or concordance between IGRA and the TST were available for all 16 studies. One study included data from two separate countries, Brazil and Nepal. Given inherent difference in the study populations these cohorts were analysed separately and are presented as separate studies.

The prevalence of positive tests varied greatly between studies and across assays. Prevalence of positive TST ranged from 22% in a study of children ≤5 years old to 84.6% in a cohort of adult HCWs exposed to a smear-positive TB case. Prevalence of positive QFT tests ranged from 10.3% to 63.2% and TSPOT.TB positivity ranged from 17.8% to 75% (Figure 24).
In total, 20 comparisons (TST with TSPOT.TB or TST with QFT) were evaluated. Only 4 studies reported a statistical difference in LTBI prevalence that was statistically significant. Three of these had study populations that were 100% BCG vaccinated while one study reported a vaccination rate of around 75%. This study evaluated both IGRAs against the TST, and found that the prevalence estimated by QFT was significantly lower than that estimated by TST and by TSPOT.TB.

Figure 25 shows summary differences in prevalence of positive tests, with proportion BCG vaccinated along the X-axis. Six comparisons resulted in a negative summary difference, while the remaining 13 comparisons (68%) showed a positive difference indicating that IGRAs estimated a lower prevalence of LTBI compared with the TST in these populations.
Concordance between IGRA positivity and TST positivity varied from 18% in a hospital contact study to 93% in a cohort of household contacts aged 0-2 years. Five studies reported poor or fair agreement (kappa<0.4), 9 studies reported moderate agreement (kappa0.4-0.6) and 5 studies reported strong agreement (kappa>0.6).

Among two studies that compared both TSPOT.TB and the QFT with TST results, one study reported very similar kappa values for both IGRA (defined as kappa=0.52-0.54) while the other study found poor agreement for the TSPOT.TB and TST results (defined as kappa=0.12), yet moderate agreement between the QFT and TST results (defined as kappa=0.45).

3.5.7. Correlation between test positivity and exposure

All 16 studies captured the contact’s degree of exposure to an active TB case. Studies used a variety of variables and constructs to classify participants into exposure categories. In order to compare the association between exposure and test positivity across studies, exposure variables were classified into two broad categories: dichotomous exposure variables and variables representing an exposure gradient.

Test positivity and dichotomous exposure

Eight studies employed a dichotomised exposure variable. Six studies compared dichotomous exposure with cross-sectional test positivity rate in TST and IGRA, one study employing up to 3 dichotomous exposure variables. The remaining two studies both had longitudinal designs and compared exposure (as a dichotomous variable) with TST or IGRA conversions post-exposure (Figure 26).
Only one study (using a dichotomised contact score) reported a significant association between exposure and test positivity for all 3 tests (adj OR: TST: 3.83, TSPOT: 38.4, QFT: 14.94).

The Zambia unpublished study did not find a statistically significant difference between contacts of smear-positive versus contacts of smear-negative TB cases (preliminary data). One study found weak associations for either TST (adj OR=1.1) or QFT (adj OR=1.3) and the dichotomous exposure variable, but neither reached statistical significance. The study in children did not find a statistically significant difference in positivity rates between exposure groups for any tests.

In the study evaluating the association between exposure variables and either concordant positives (i.e. TST+/IGRA+) or discordant results, a strong association between TST-/IGRA+ and >1 month exposure was found (adj OR=7.2; 95%CI 1.7% - 29.3%), but no statistically significant association between exposure and TST+/IGRA-. The same study also found a statistically significant association between exposure and smear-positive status graded as 3+ (adj OR=2.8; 95%CI 1.3% - 6.1%).
Test positivity and gradient of exposure

Nine studies investigated an association between test positivity (TST and/or IGRA) and an exposure gradient (Table 18). The most common exposure gradient (used by three studies) was the so-called “Senegal/Gambia exposure gradient” categories, including different house, same house but different room, and same room. Four studies used exposure gradients based on the index case smear status. Four of the studies used groups with no TB exposure as comparison (community controls). The rest used different designs as outlined in Table 18.

Only one study reported a comparison of all three tests against an exposure gradient. TST positivity was associated with the exposure gradient and significant at each level. The association between IGRA positivity and the exposure gradient was only significant in the highest category of exposure (QFT adj OR 4.0; 95CI 1.4% - 11.4%; TSPOT.TB adj OR 6.6; 95CI 1.7% - 25.2%). The Zambia unpublished study used the same exposure gradient but found no significant association at any level of the exposure gradient for either TST or QFT.

Figures 28A (TST) and 28B (IGRA) display the ORs estimated by each study for the association between test positivity and exposure gradient.
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Exposure level</th>
<th>QFT positivity</th>
<th>TSPOT.TB positivity</th>
<th>TST positivity</th>
<th>Other associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adetifa, 2007</td>
<td>Different House</td>
<td>adjOR=1</td>
<td>-</td>
<td>adjOR=1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Same house, different room</td>
<td>adjOR=1.5 95% CI:0.6-3.6</td>
<td>-</td>
<td>adjOR=2.4 95% CI:0.9-6.5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Same room</td>
<td>adjOR=3.8 95% CI:1.2-12.5</td>
<td>-</td>
<td>adjOR=4.8 95% CI:1.3-17.1</td>
<td>-</td>
</tr>
<tr>
<td>Adetifa, 2010</td>
<td>Different House</td>
<td>adjOR=1</td>
<td>adjOR=1</td>
<td>adjOR=1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Same house, different room</td>
<td>adjOR=1.5 95% CI:0.7-3.1</td>
<td>adjOR=2.6 95% CI:0.9-7.1</td>
<td>adjOR=2.9 95% CI:1.3-6.7</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Same room</td>
<td>adjOR=4.0 95% CI:1.4-11.4</td>
<td>adjOR=6.6 95% CI:1.7-25.2</td>
<td>adjOR=15 95% CI:4.7-47.2</td>
<td>-</td>
</tr>
<tr>
<td>Nakaoka, 2006</td>
<td>Community Controls</td>
<td>4/39 (10%) 95% CI:2.9-24.2</td>
<td>-</td>
<td>-</td>
<td>Association found between bacillary load of index case and the contact’s likelihood of testing QFT positive (p=0.03)</td>
</tr>
<tr>
<td></td>
<td>Contacts of smear negative TB cases</td>
<td>8/81 (10%) 95% CI:4.4-18.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Contacts of smear positive TB cases</td>
<td>53/72 (74%) 95% CI:61.9-83.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nicol, 2009</td>
<td>None: no contact</td>
<td>-</td>
<td>-</td>
<td>OR=2.4 95% CI:1.2-4.8</td>
<td>Increasing likelihood for positive TSPOT.TB and TST Results with increasing exposure TSPOT.TB p=0.003 TST p=0.0081</td>
</tr>
<tr>
<td></td>
<td>Any: any contact with a case of TB</td>
<td>-</td>
<td>OR=1.6 95% CI:0.7-3.3</td>
<td>OR=2.4 95% CI:1.2-4.8</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Adult: contact with an adult currently receiving TB treatment</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Household: contact with a patient with</td>
<td>-</td>
<td>OR=2.4 95% CI:1.3-4.6</td>
<td>OR=2.4 95% CI:1.4-4.2</td>
<td>-</td>
</tr>
<tr>
<td>Study</td>
<td>Group Description</td>
<td>Active TB in the same household</td>
<td>Exposure to smear negative TB</td>
<td>Exposure to smear grade 1+</td>
<td>Exposure to smear grade 2+</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td>---------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Okada, 2007</td>
<td></td>
<td></td>
<td>adjOR=1</td>
<td>adjOR=4.05 95%CI:1.04-15.75</td>
<td>adjOR=2.25 95%CI:0.81-6.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ozekinci, 2007</td>
<td>Group 4: Healthy &amp; non-exposed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group 3: Clinic and Lab Personnel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group 2: Household Contacts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruhwald, 2008</td>
<td>Community Controls</td>
<td>3/23 (13%) 95%CI: 2.8-33.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contacts of smear negative TB cases</td>
<td>3/38 (8%) 95%CI:1.7-21.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contacts of smear positive TB cases</td>
<td>42/59 (71%) 95%CI:57.9-82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAMSTAR, 2010</td>
<td>Different house</td>
<td>OR=1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Same bed</td>
<td>OR=1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Same room</td>
<td>Same house</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
<td>------------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR=1.21</td>
<td>OR=1.09</td>
<td>OR=1.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%CI:0.63-2.19</td>
<td>95%CI:0.53-2.77</td>
<td>95%CI:0.62-1.90</td>
<td>95%CI:0.62-1.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR=0.92</td>
<td>OR=0.77</td>
<td>OR=0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%CI:0.40-2.09</td>
<td>95%CI:0.42-1.41</td>
<td>95%CI:0.40-2.09</td>
<td>95%CI:0.39-1.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kasambira, 2010

<table>
<thead>
<tr>
<th></th>
<th>Index case: Smear Positive (REF)</th>
<th>Index case: Smear negative, culture pos TB</th>
<th>Index case: clinical TB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>adjOR=1</td>
<td>adjOR=0.84 (95%CI:0.09-7.8)</td>
<td>adjOR=3.9 (95%CI:0.67-23.5)</td>
</tr>
<tr>
<td></td>
<td>adjOR=1</td>
<td>adjOR=2.7 (95%CI:0.56-13)</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Smear negative (REF)</th>
<th>Smear scanty</th>
<th>Smear 1+</th>
<th>Smear 2+</th>
<th>Smear 3+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>adjOR=1</td>
<td>--</td>
<td>adjOR=5.5 (95%CI:0.89-34.7)</td>
<td>adjOR=8.7 (95%CI:1.2-62)</td>
<td>adjOR=11.4 (95%CI:1.8-72)</td>
</tr>
<tr>
<td></td>
<td>adjOR=1</td>
<td>--</td>
<td>adjOR=7.9 (95%CI: 1.5-41)</td>
<td>adjOR=15.7 (95%CI:2.6-92)</td>
<td>adjOR=11.7 (95%CI:2.2-62)</td>
</tr>
<tr>
<td></td>
<td>adjOR=1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Use of IGRAs in contact screening and outbreak investigation

**Figure 27A.** Association between TST positivity and exposure gradient

**Figure 27B.** Association between IGRA positivity and exposure gradient

(Effect size estimates displayed are odds ratios)
Use of IGRAs in contact screening and outbreak investigation
Two studies used the same exposure gradient, and included a non-exposed comparison group (community controls). In one of these studies, contacts of smear-positive cases were more likely to be QFT positive than either community controls or contacts of smear-negative cases. In addition, an association was found between the smear grading of the index case and the contact’s likelihood of testing QFT positive (p=0.03). The other study had similar results for QFT but not for TST. Studies that did not use non-exposed comparison groups were unable to show similar trends.

The remaining studies all used unique exposure gradients, making these difficult to compare. The study which included very young children (0-2 yrs), found identical estimates of association between TSPOT.TB positivity and TST positivity in household contacts. One study grouped participants into healthy unexposed, clinical and laboratory personnel and household contacts but found no difference in TST or TSPOT.TB positivity across different exposure groups.

### 3.5.8 Concordance between test results in longitudinal contact studies

Three longitudinal studies were conducted in low and middle income countries which followed contacts over time and repeated TST and IGRA. All three studies evaluated QFT and TST but not the TSPOT.TB test (Table 19).

One small study followed contacts (n=25) approximately 8 weeks after nosocomial exposure: QFT and TST conversion rates in the less exposed groups ranged from 0-8.3% and 0-6.9% respectively, depending on how the exposure groups were characterized. QFT and TST conversion rates in the high TB exposure groups ranged from 0-25% and 0-10% respectively. Intimate contact (OR=1.94; 95%CI 0.2% - 21.1%) and face-to-face contact for more than 1 hour (OR=9.2; 95CI 0.7% - 100%22.38) tended to be associated with a higher risk of QFT conversion, but this did not reach statistical significance.

The second study included an analysis of household contact in India over a 1-year period. QFT conversion rates of 21.2 % and TST conversion rates of 13.8% (6mm increase over baseline) and 7.5% (10mm increase over baseline) were observed. This study also assessed the incidence of QFT reversions, and found an overall rate of 6.4% over 1 year; this rate was higher (17-20%) when restricted to those with quantitative baseline QFT results (IFN-g <3.0 IU/ml). Associations between test conversions and exposure were not examined, however, 83% agreement between TST and QFT conversions (1 year post-exposure) was found, with a kappa=0.42 (0.17-0.68).

The third study (unpublished, South Africa) followed children exclusively. Conversion rates were calculated for a 6-month follow-up period. In baseline negatives, 18.8% (22/117) of QFT conversions were observed while TST conversions were found in 11% (13/118) of baseline TST negatives. This difference was statistically significant (p=0.03). While no exposure variables were associated with TST conversions, an association between exposure and QFT conversions (adjOR=0.06 for index case smear status 3+; 95CI 0.08-0.49) was seen.
## Table 19. TST and QFT conversions in contact studies in low- and middle-income countries

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Type of exposure</th>
<th>QFT conversion (%)</th>
<th>TST conversion (%)</th>
<th>QFT conversion OR (95CI)</th>
<th>TST conversion OR (95%CI)</th>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee, 2007</td>
<td>Intimate Exposure</td>
<td>12.5</td>
<td>8.3</td>
<td>1.94 (0.18-21.12)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>No intimate exposure</td>
<td>6.7</td>
<td>0</td>
<td>(Ref)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Contact time (&gt;8h)</td>
<td>15</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Contact time (&lt;8h)</td>
<td>0</td>
<td>0</td>
<td>9.2 (0.69-22.38)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Face to face contact (&gt;1h)</td>
<td>25</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Face to face contact (&lt;1h)</td>
<td>3.5</td>
<td>6.9</td>
<td>(Ref)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Same room (&gt;8h)</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Same room (&lt;8h)</td>
<td>8.3</td>
<td>5.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pai, 2009</td>
<td>Household contacts</td>
<td>21.2</td>
<td>(+6mm) 13.8 (+10mm)</td>
<td>7.5</td>
<td>-</td>
<td>kappa=0.42</td>
</tr>
<tr>
<td>Kasambira, 2010</td>
<td>Household contacts</td>
<td>22/117 (18.8%)</td>
<td>13/118 (11%)</td>
<td>p=0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Index case smear status 3+</td>
<td></td>
<td>adjOR=0.06 (0.008-0.49)</td>
<td>No association</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.5.9. **Indeterminate IGRA results**
Rates of indeterminate results varied across studies. 11/15 (80%) of studies reported indeterminate rates below 5%. In the two studies that evaluated both IGRA higher indeterminate rates were reported with the QFT test than the TSPOT.TB (data not shown).

3.5.10 **Strengths & limitations of the evidence base**
Due to heterogeneity in study designs and outcomes assessed in each study, it was not appropriate to pool the data.

The majority of studies were cross-sectional and looked at concordance between TST and IGRA. Studies that assessed associations between exposure and test positivity used different categorisation of exposure variables, making it difficult to compare results across studies.

3.5.11 **Summary of findings and GRADE evidence profiles**
- The majority of studies showed comparable LTBI prevalence by TST or IGRA in contacts;
- The most commonly observed discordance was of the TST+/IGRA- type;
- Both IGRA and the TST seem to show positive associations with higher levels of exposure in cross-sectional studies, but the strength of the association (effect) varied across studies;
- IGRA appear to be dynamic assays with frequent conversions and reversions;
- Both IGRA and TST seem to have similar and modest predictive value;

3.5.12 **Final Recommendations**
The GRADE evidence profiles are provided in Table 20. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRA for LTBI screening in contact and outbreak investigations was **very low** and recommended that **these tests should not be used** as a replacement for TST, neither in adults or children investigated as close contacts of patients with confirmed active TB.

<table>
<thead>
<tr>
<th>OVERALL QUALITY OF EVIDENCE</th>
<th>VERY LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRENGTH OF RECOMMENDATION</td>
<td>STRONG</td>
</tr>
</tbody>
</table>
### Table 20. GRADE evidence profile: Performance of IGRAs for the diagnosis of LTBI in contacts of active TB in low- and middle-income countries.

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Efficacy of preventive therapy based on IGRA test results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>No of Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**B. Predictive value of IGRA for active TB**

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 studies: Covered in Predictive SR: Rangaka et al</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
</tbody>
</table>

**C. Outcome: Correlation between IGRAs and different gradients of TB exposure (ordinal, continuous, etc.)**

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,868 (9)^A2</td>
<td>Cross-sectional</td>
<td>Serious^A2 (-1)</td>
<td>No Serious indirectness^A3</td>
<td>Serious^A4 (-1)</td>
<td>No serious imprecision^A5</td>
<td>Likely^A6</td>
<td>Low @@O@@O</td>
<td>Critical (7-9)</td>
</tr>
</tbody>
</table>

**D. Outcome: Correlation between IGRAs and TB exposure as a dichotomous variable**

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,145 (6)^B1</td>
<td>Mainly cross-sectional</td>
<td>Serious^B2 (-1)</td>
<td>No Serious indirectness^B3</td>
<td>Serious^B4 (-1)</td>
<td>Serious^B5 (-1)</td>
<td>Likely^B6</td>
<td>Very Low @O@O@O</td>
<td>Critical (7-9)</td>
</tr>
</tbody>
</table>

**E. Outcome: Correlation between IGRA conversions and TB exposure**

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>309 (2)^C1</td>
<td>Longitudinal</td>
<td>Serious^C2 (-1)</td>
<td>No Serious indirectness^C3</td>
<td>Very Serious^C4 (-2)</td>
<td>Serious^C5 (-1)</td>
<td>Likely^C6</td>
<td>Very Low @O@O@O</td>
<td>Critical (7-9)</td>
</tr>
</tbody>
</table>

**F. Outcome: Sensitivity for active TB (as a surrogate reference standard for LTBI)**

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Important (4-6)</td>
</tr>
</tbody>
</table>

**G. Outcome: Concordance with tuberculin skin test (TST)**

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,080 (16)^D1</td>
<td>Mainly cross-sectional</td>
<td>Serious^D2 (-1)</td>
<td>Very Serious^D3 (-2)</td>
<td>Very Serious^D4 (-2)</td>
<td>Serious^D5 (-1)</td>
<td>Likely^D6</td>
<td>Very Low @O@O@O</td>
<td>Important (4-6)</td>
</tr>
</tbody>
</table>
Footnotes:

1 Quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: imitations, indirectness, inconsistency, imprecision, and publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high quality observational studies (cross-sectional or cohort studies with diagnostic uncertainty and direct comparison of test results with culture) and at moderate when these types of studies were absent. One point was subtracted when there was a serious issue identified or two points when there was a very serious issue identified in any of the criteria used to judge the quality of evidence.

A1 9 studies were included: 1 study evaluated both TSPOT.TB and QFT-GIT, 2 studies evaluated TSPOT.TB, the 6 remaining studies evaluated QFT-GIT(n=5) or QFT-G(n=1).

A2 2 out of 9 studies were unpublished and quality indicators could not be assessed; remaining study populations were considered to be representative, however, only 1 of the remaining 7 studies reported that assessment of test results was performed blinded to other test results. Only 2/7 reported the blood draw had been performed prior to the TST.

A3 33% (3/9) studies were done in low-income settings and the remaining 6 studies were done in middle-income settings. Some indirectness in the choice of reference standard was observed.

A4 Serious heterogeneity in characterization of exposure gradient (some based on index case’s smear status, some based on sleeping proximity, etc.) and in estimated effect.

A5 Majority of studies had 200-300 participants, smallest study n=120. Estimated 95%CIs were relatively tight.

A6 Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. Although no points were deducted, it was assumed that some degree of publication bias is likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future (although an attempt was made to include unpublished studies, despite not being comprehensive); 2) there are anecdotal examples of unpublished negative studies on IGRAs; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

B1 6 studies were identified: 1 study evaluated both TSPOT.TB and QFT-G, while 1 study evaluated TSPOT.TB. The remaining 4 studies all evaluated QFT-GIT.

B2 Only the 4 published studies could be assessed for quality, 50% reported on timing of blood draw prior to TST, 50% reported blinding had been done for assessment of test results and 50% reported industry involvement.

B3 All studies, except one done in low-income setting were done in upper-middle income settings. Some indirectness in the choice of reference standard was noted.

B4 Serious heterogeneity in characterization of exposure gradient (some based on index case’s smear status, some based on sleeping proximity, etc.) and in estimated effect.

B5 All but one large study (n=2211) had between 82-301 participants. Studies estimated wide 95%CI, and majority were not significant.

B6 Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. Although no points were deducted, it was assumed some degree of publication bias was likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future (although an attempt was made to include unpublished studies, despite not being comprehensive); 2) there are anecdotal examples of unpublished negative studies on IGRAs; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

C1 2 studies were included; both studies evaluated the QFT, one study using the QFT-GIT and the other the QFT-G.
Use of IGRAs in contact screening and outbreak investigation

C1 A study was unpublished and hence not suitable for quality assessment, the other study was a longitudinal study that followed HCWs after a nosocomial infection. Population was representative, blood draw was done prior to TST, and there was no industry involvement, however, blinding was not reported.

C3 Both studies were done in Upper middle income settings, however one was a nosocomial outbreak involving health care workers and may not be generalizable to other contact settings including household contacts, especially in low income settings. While we did not downgrade for reference standard, we acknowledge there is some indirectness in the choice of reference standard.

C4 Serious heterogeneity between estimated ORs for exposure and conversions, one study shows an positive association between conversions and exposure, while the other shows a significant protective effect of exposure for conversions.

C5 45% CIs are tight and significant for the large unpublished (n=2211), however, CIs range from 0.18-21.12 and 0.69-122.38 for the smaller hospital outbreak study (n=39).

C6 Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias cannot be ruled out. Although we did not deduct points, we assumed some degree of publication bias is likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future (although we made an attempt to include unpublished studies, our attempt was not comprehensive); 2) there are anecdotal examples of unpublished negative studies on IGRAs; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

D1 2 studies included both IGRAs, 3 studies evaluated only TSPOT.TB, while the rest evaluated a version of the QFT.

D2 11/14 studies did not report on whether personnel assessing test results had been blinded to previous test results or reference standard, and 5/14 studies reported industry involvement.

D3 Studies were conducted in low and middle income settings. TB exposure gradient does not necessarily classify the target condition (LTBI) correctly.

D4 47% of studies showed moderate agreement, while 26.5% showed poor agreement and 26.5% fair agreement. In 68% of comparisons, TST estimated a higher prevalence while in the remaining 32% IGRA estimated a higher prevalence of LTBI.

D5 Due to heterogeneity in effect estimates concordance could not be pooled. However, effects estimated for individuals studies were frequently not significant.

D6 Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias cannot be ruled out. Although points were not deducted, a degree of publication bias is likely because: 1) literature on IGRAs is rapidly exploding and currently unpublished studies may come out in future (although we made an attempt to include unpublished studies, our attempt was not comprehensive); 2) there are anecdotal examples of unpublished negative studies on IGRAs; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.
3.6 The predictive value of IGRAs for incident active TB

3.6.1 Objectives, reference standards and outcomes

Primary objectives

- To assess whether IGRAs can prospectively predict the development of active TB (incident TB) amongst individuals without active TB disease at first assessment;

- To determine whether IGRA predictive (prognostic) ability is significantly higher than that for the TST.

Secondary objectives

- To assess the variability in IGRA-positive and -negative TB rates of progression by demographic and clinical risk factors for TB: provision of IPT, age strata (adult or children), HIV infection, BCG status;

- To determine the influence of immunological phenotypes of discordant and concordant TST/IGRA pairs at baseline on subsequent TB rates;

- To determine the existence of a gradient relationship between quantitative IFN-gamma response levels and rates of progression to TB disease;

- To assess the variability in the rates of TB in IGRA-positive individuals with TST (negative and positive) results who were treated with IPT.

Reference standards

As there was no data at the highest level of the hierarchy of evidence, the focus was on rates of incident active TB disease in individuals with positive and negative IGRA results at baseline and how they compare to those with positive and negative baseline TST results

Primary outcomes

- Crude incidence rates of disease progression, analysed by relevant strata (eg. IGRA-positive and -negative, IGRA/TST discordant and concordant pairs, IGRA quantitative levels);

- Incidence rate ratios (IRR) for disease progression in IGRA-positive versus IGRA-negative individuals (and likewise for the TST);

When exploring differences in the predictive ability of IGRA compared to the TST, rate ratios for TB in test positives vs. test negatives were computed and pooled, where appropriate, for studies that performed both TST and IGRA.
Definitions

**Sensitivity:** The probability that a baseline IGRA test is positive amongst those who ultimately developed active TB during follow-up;

**Specificity:** The probability that a baseline IGRA test is negative in those that did not subsequently develop active TB.

Sensitivity and specificity were regarded as surrogates of patient-relevant outcomes important for assessing the frequency and impact of either a false-negative or false-positive IGRA result at baseline: A false-positive test result may result in unnecessary preventive therapy in an individual who would not have progressed to TB disease, while a false-negative result would mean progression to active TB disease that could have been prevented.

### 3.6.2 Search results

Six studies conducted in low- and middle-income countries were identified. Of these, three were conducted in low-income countries (The Gambia, Senegal, Zambia) and three were done in middle-income countries (Turkey, China, South Africa). Three of the six studies (n=7,392) evaluated commercial IGRA s. Two studies (South Africa, Zambia) used the most recent QuantiFERON Gold In Tube technology while the third (China) evaluated the T-Spot.TB assay.

### 3.6.3 Data analysis

Data was extracted from included articles according to study design, participants, country, period of recruitment, proportion BCG vaccinated, IGRA method (assay used, test version, cut-off point used, etc.), TST method (PPD dose, cut-point used, etc.). Outcome data included baseline TST and IGRA positivity, IGRA/TST concordance/discordance and rates of progression to active TB.

### 3.6.4 Study characteristics

Studies included were those which enrolled adults or children without TB at baseline, regardless of HIV infection status. The studies were of longitudinal design (prospective or retrospective cohorts) and included any period of follow-up using either active or passive strategies.

The populations included in the studies were all different but all studies followed-up cohorts of groups known to be at high risk of TB progression. The China study recruited an older (mean age 60 years), all-male, high-risk TB group with confirmed silicosis; the South Africa study recruited school-going adolescents between 12 and 18 years. The Zambia study included TB case-contacts (15 - 65+years), with 37% of the 721 individuals in the cohort being HIV infected. None of the other two studies included HIV infected individuals. Over 80% of the China and South Africa study cohorts completed follow-up (information not available for the Zambia study).

Isoniazid preventive therapy (IPT) was only provided in the China cohort (33% of 203 TST-positives).

### 3.6.5 Study quality

IGRA predictive value studies are prospective studies that are not focused on conventional diagnostic test accuracy estimates of sensitivity and specificity, or single unadjusted measures of positive and negative predictive values. A modified version of the Newcastle-Ottawa Quality Assessment Scale for cohort studies was therefore used.
Specific modifications were made to the selection and outcome items, in particular to ascertainment or verification of the reference standard for active TB. Although microbiological determination of TB is regarded as the gold standard, not all studies used conventional microbiology to assess outcome. Possible verification bias was therefore considered to ascertain whether some exposure groups were more likely to have been assessed differently.

Studies from countries that have incorporated IGRA use into their guidelines for TB are likely to demonstrate a greater association, and therefore, better predictive ability of IGRA with incident active TB disease; thus, studies were assessed to determine whether respective countries had already included IGRA in their guidelines for identifying latent TB infection and/or active TB disease. The incorporation bias item was drawn from the QUADAS tool.

Overall, included studies varied in quality particularly with regard to the comparability (adjustments made to effect measures) and outcome (ascertainment, losses to follow-up, reporting) components of the modified NOS. Only the China study reported microbiological confirmation (around 80%) of at least 50% of the diagnosed incident TB cases. The South Africa study incorporated IGRA results in their reference standard for active TB.

### 3.6.6 Incidence rates of TB during follow-up

Overall incidence rates (IR) of TB were 22.5/1000 person years (PY) and 20.7/1000 PY for the China and Zambia studies, respectively. The South Africa study did not report overall TB rates but did provide results stratified by IGRA exposure status. Pooled IR of TB appeared to be higher in IGRA-positive individuals (IR=16.47; 95CI 11.2-21.7; I²=98%, p<0.0001) vs. IGRA-negatives (IR=2.85; 95CI 0.9-4.8; Figure29). However, even in those with positive IGRA results, the vast majority of individuals did not progress to TB disease during follow-up. Sub-groups analyses were not performed due to the small number of studies.
Figure 28. Crude TB incidence rates stratified by IGRA status

<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Publication Year</th>
<th>IR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IGRA positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China (All TB)</td>
<td>Leung (2010)</td>
<td></td>
<td>32.00 (17.80, 52.30)</td>
</tr>
<tr>
<td>Zambia</td>
<td>Shanaube (2010)</td>
<td></td>
<td>29.80 (16.92, 52.46)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Mahomed (2010)</td>
<td></td>
<td>6.40 (4.50, 8.70)</td>
</tr>
<tr>
<td>Subtotal (I-squared = 97.99%, p = 0.000)</td>
<td></td>
<td></td>
<td>16.47 (11.24, 21.70)</td>
</tr>
<tr>
<td></td>
<td>IGRA negative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China (All TB)</td>
<td>Leung (2010)</td>
<td></td>
<td>7.10 (0.90, 25.50)</td>
</tr>
<tr>
<td>Zambia</td>
<td>Shanaube (2010)</td>
<td></td>
<td>11.00 (3.00, 28.00)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Mahomed (2010)</td>
<td></td>
<td>2.20 (1.20, 3.80)</td>
</tr>
<tr>
<td>Subtotal (I-squared = 34.6%, p = 0.217)</td>
<td></td>
<td></td>
<td>2.85 (0.86, 4.84)</td>
</tr>
</tbody>
</table>
3.6.7 The association between IGRA and incident active TB

Incident rate ratios
Pooled estimates should be interpreted with caution as there were only 3 studies per stratum. The strength of the association between IGRA positivity and incident TB was moderate, indicated by a pooled incidence rate ratio (IRR) of around 3, failing to reach significance as indicated by wide confidence intervals (95CI 0.7-5.6; $I^2$ for pooled estimate=0%, p=0.912) (Figure 29).

Figure 29. Crude, unadjusted TB incidence rate ratios for IGRA-positives vs. IGRA-negatives

![Graph showing crude, unadjusted TB incidence rate ratios for IGRA-positives vs. IGRA-negatives]

3.6.8 The predictive value of IGRA compared with TST

All three studies provided incidence rate ratios of TB stratified by IGRA as well as TST status at baseline. The association with subsequent incident TB in test-positive individuals compared to test-negatives appears higher for IGRA than for TST; however, this was not statistically significant as indicated by overlapping confidence intervals (IGRA: IRR=3.24; 95CI 0.62-5.85; $I^2$=0%; p=0.90; TST: IRR=2.28; 95CI 0.83-3.73; (Figure 10).
Figure 10. Crude, unadjusted incidence rate ratios for IGRAs compared to the TST

<table>
<thead>
<tr>
<th>Country</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGRAs</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>4.10 (1.83, 9.08)</td>
</tr>
<tr>
<td>Zambia</td>
<td>2.78 (0.84, 8.78)</td>
</tr>
<tr>
<td>South Africa</td>
<td>2.90 (1.59, 5.31)</td>
</tr>
<tr>
<td>Subtotal (I-squared = 0.09; p = 0.593)</td>
<td>3.24 (0.62, 1.59)</td>
</tr>
<tr>
<td>TST+</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>1.64 (0.58, 3.52)</td>
</tr>
<tr>
<td>Zambia</td>
<td>1.48 (0.47, 5.38)</td>
</tr>
<tr>
<td>South Africa</td>
<td>3.03 (1.90, 7.30)</td>
</tr>
<tr>
<td>Subtotal (I-squared = 0.09; p = 0.383)</td>
<td>2.28 (0.85, 7.78)</td>
</tr>
</tbody>
</table>
3.6.9 The influence of discordant-concordant TST/IGRA pairs at baseline on subsequent TB rates

Evaluating discordant results may represent the best way to assess whether IGRAs perform better than the TST in predicting risk of incident TB disease. TST+/IGRA- results are mainly thought to indicate remote LTBI. By contrast, TST-/IGRA+ may represent more recently acquired infection that may either subsequently clear or progress to active disease.

The Zambia and South Africa studies (combined N=5,861) explored rates of TB in paired concordant and discordant TST/IGRA results (Table 21). Double-positive results (TST-positive/IGRA-positive) seemed to yield higher rates of TB during follow-up compared to double-negative results (TST-negative/IGRA-negative).

The Zambia study reported higher rates in the discordant pair where IGRA was the positive test compared to when TST was the positive test (incident TB rate 29.7/1000PY; 95CI 13.4 – 66.2). In contrast the South Africa study reported marginally higher rates in IGRA-negative/TST-positive discordant pairs (incident TB rate 3.3/1000PY; 95CI 0.4-12.0) than in IGRA-positive/TST-negative pairs (incident TB 1.8/1000PY; 95CI 0.4-5.4). However, these differences are not significant as the confidence intervals are wide and overlap.
### Table 21. Discordant-concordant TST/IGRA pairs and incidence rates of TB

<table>
<thead>
<tr>
<th>Country</th>
<th>Phenotype</th>
<th>n/N</th>
<th>PY</th>
<th>IR per 1000PY</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zambia</td>
<td>IGRA+/TST-</td>
<td>6/191</td>
<td>201.7</td>
<td>29.7</td>
<td>13.36-66.20</td>
</tr>
<tr>
<td>10mmTST</td>
<td>IGRA-/TST+</td>
<td>0/42</td>
<td>44.6</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>14pg/ml</td>
<td>IGRA+/TST+</td>
<td>6/173</td>
<td>201.1</td>
<td>29.8</td>
<td>13.4-66.41</td>
</tr>
<tr>
<td></td>
<td>IGRA-/TST-</td>
<td>0/211</td>
<td>224.9</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>S. Africa</td>
<td>IGRA+/TST-</td>
<td>3/714</td>
<td>1628</td>
<td>1.8</td>
<td>0.4-5.4</td>
</tr>
<tr>
<td>10mm TST</td>
<td>IGRA-/TST+</td>
<td>2/259</td>
<td>601</td>
<td>3.3</td>
<td>0.4-12.0</td>
</tr>
<tr>
<td>14pg/ml</td>
<td>IGRA+/TST+</td>
<td>35/1955</td>
<td>5250</td>
<td>8</td>
<td>4.6-9.3</td>
</tr>
<tr>
<td></td>
<td>IGRA-/TST-</td>
<td>10/2316</td>
<td>4508</td>
<td>2.1</td>
<td>1.1-4.1</td>
</tr>
</tbody>
</table>

### 3.6.10 Gradient between rates of TB and quantitative IGRA levels

It has been proposed that the risk of subsequently developing TB following testing may vary according to the levels of IFN-gamma produced in response to RD1-antigens. In turn, the levels of IFN-gamma produced may depend on the mycobacterial burden or IFN-gamma may reflect active replicating bacilli. In case-contact studies the length or intensity of exposure may therefore correlate with the level of IFN-gamma produced.

The Zambia study (N=721) explored this hypothesis, suggesting that there was no exposure-gradient relationship between quantitative baseline IGRA levels and rates of subsequent TB. Paradoxically, TB rates appeared highest in the lowest IGRA quartile (0.35-0.64 IU/ml) at 73.8/1000PY (95CI 23.8-228.94). High background TB prevalence, (ie. higher infection pressure and higher re-infection possibility), high background HIV prevalence and the high proportion of HIV-infected individuals in the follow-up cohort may explain these results; however, comparisons across the different IGRA levels were not statistically significant.

### 3.6.11 Patient relevant outcomes

The diagnostic accuracy estimates of sensitivity and specificity are surrogates of patient-relevant outcomes important for assessing the frequency and impact of either a false-negative or false-positive IGRA result at baseline in respective cohorts: For example, a false-positive result may result in IPT prescription for several months and although safe is not without adverse effects, notably clinical hepatitis; A false-negative result will result in no IPT being provided and the individual exposed to an increased risk of developing active TB in the future.

As only three studies were available, summary ROC statistics could not be derived; therefore, a description of test accuracy estimates per individual study is provided: IGRA sensitivity for incident TB was 88% (95CI 64% - 99%) in China (T-SPOT.TB), 75% (95CI 48% - 93%) in Zambia (QFT-GIT) and 75% (95CI 61% - 86%) in South Africa. Specificity was low across the three studies at 35% (95CI 30% - 41%), 50% (95CI 46% - 54%) and 49% (95CI 48% - 51%) respectively. This means that more than 50% of individuals would unnecessarily receive IPT based on a positive IGRA result alone.
TST sensitivity for incident TB was similar at 76% (95CI 50% - 93%) and 73% (95CI 59% - 84%) for the China and South Africa studies, respectively. Specificity was low at 35% (95CI 29% -41%) in China and 58% (95CI 57% - 58%) in South Africa. This means that between 40% and 60% of individuals would unnecessarily receive IPT based on a positive TST alone. In the Zambia study results were different - TST sensitivity for subsequent TB disease was low at 44% (95CI 20% - 70%) while specificity was higher than in the other two studies (67%; 95CI 64% - 71%). The Zambia study acknowledged logistical issues at the clinical sites that possibly affected TST results.

3.6.12 The predictive value of serial testing
This could not be assessed as all three studies performed single time-point IGRA testing.

3.6.13 Summary of findings and GRADE evidence profiles

- Included studies vary in quality, particularly with regard to comparability (adjustments made to effect measures) and outcome (ascertainment of incident TB, losses to follow-up, and reporting of incidence rates vs. cumulative incidence);
- A reference standard that is not independent of the index test introduces differential ascertainment bias and is likely to complicate studies, especially those from countries where IGRA have been adopted as part of national TB work-up algorithms. Prognostic studies of IGRA conducted in such countries should give full discussions of the extent of this and also mention how that was considered in final analyses, if at all.
- The vast majority of individuals (>95%) with a positive IGRA results did not progress to active TB disease during follow-up, although a modest and statistically insignificant increase in incidence rates of TB in IGRA-positives compared to IGRA-negatives has been observed;
- Both IGRA and the TST appear to have only modest predictive value and do not help identify those who are at highest risk of progression to disease. Patient relevant outcomes based on sensitivity and specificity appear comparable between the two tests.

3.6.14 Final Recommendations
The GRADE evidence profiles are provided in Tables 22 and 23. Based on these assessments, the Expert Group concluded that the quality of evidence for the predictive value of IGRA was very low and recommended that these assays should not be used to identify individuals at risk of active TB disease in low- and middle-income countries.

<table>
<thead>
<tr>
<th>OVERALL QUALITY OF EVIDENCE</th>
<th>VERY LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRENGTH OF RECOMMENDATION</td>
<td>STRONG</td>
</tr>
</tbody>
</table>
Table 22. GRADE Evidence Profile: Predictive value of commercial IGRA for incident active TB in Low and Middle-Income Countries

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Outcome: Efficacy of preventive therapy based on IGRA results</td>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
<td></td>
</tr>
<tr>
<td>B. Outcome: Prospective predictive value of IGRA for the development of active incident TB? (Do IGRA positive results have a stronger association with subsequent development of active TB compared to IGRA negative results?)</td>
<td>7,392 (3)</td>
<td>Cohort studies</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>No serious inconsistency</td>
<td>Very Serious (-2)</td>
<td>Likely</td>
<td>Very low</td>
</tr>
<tr>
<td>C. Outcome: Predictive value of IGRA for the development of active incident TB compared to the TST (Are IGRA's (positive vs. negative) have a stronger statistical association with subsequent active TB than the TST (positive vs. negative)?)</td>
<td>7,392 (3)</td>
<td>Cohort studies</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>No serious inconsistency</td>
<td>Very Serious (-2)</td>
<td>Likely</td>
<td>Very low</td>
</tr>
<tr>
<td>D. Outcome: Predictive value of IGRA for subsequent TB when IGRA are evaluated as part of a multivariable clinical algorithm for predicting TB (Additive value of IGRA)</td>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Important (4-6)</td>
<td></td>
</tr>
<tr>
<td>E. Outcome: Quantitative IGRA levels and subsequent rates of TB</td>
<td>721 (1)</td>
<td>Cohort of TB case-contacts</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>Very Serious (-2)</td>
<td>Likely</td>
<td>Very low</td>
</tr>
<tr>
<td>F. Outcome: Immunological phenotypes of discordant-concordant TST/IGRA pairs and subsequent rates of TB</td>
<td>5,861 (2)</td>
<td>Cohort studies</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>Very Serious (-2)</td>
<td>Likely</td>
<td>Very low</td>
</tr>
<tr>
<td>G. Outcome: Sensitivity, Specificity, False positive rates etc for active TB (as surrogates of patient relevant outcomes)</td>
<td>7,392 (3)</td>
<td>Cohort studies</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>Very Serious (-2)</td>
<td>Likely</td>
<td>Very low</td>
</tr>
<tr>
<td>H. Outcome: Utility of repeated or serial IGRA results for predicting subsequent incident active TB</td>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Important (4-6)</td>
<td></td>
</tr>
</tbody>
</table>

Footnotes
Quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high quality observational studies and at moderate when these types of studies were absent. We then subtracted one point when there was a serious issue identified or two points when there was a very serious issue identified in any of the criteria used to judge the quality of evidence.

3 studies were eligible and thus included in the analysis; 1 published (China) and 2 unpublished (Zambia and South Africa). (N refers to numbers that entered follow-up)

Based on the Newcastle-Ottawa scale, study samples were considered to be representative of specific groups of interest (i.e., silicosis patients (China), case-contacts (Zambia), adolescent school-goers) within the population and IGRA exposure groups were drawn from the same sample and therefore unlikely to introduce any bias. However, studies varied with regard to the comparability (adjustments made to effect measures) and outcome (ascertainment, losses to follow-up, reporting) components of the modified NOS. Lack of proper ascertainment of the TB outcome is considered to be the most serious of limitations. A point is deducted.

The results of the studies could be generalized for the specific country/region and for those specific groups of interest. However, the small number of studies warrants caution; a point is deducted for indirectness.

All 3 studies showed similar results and with very little heterogeneity in the pooled incidence rate ratio ($I^2=0\%$, $p=0.912$). No points were deducted.

The number of incident TB cases was small in all studies and the rates of TB fairly moderate; confidence intervals for relative risk estimates were wide (precision > +/- 20%). This is a very serious limitation. Two points are deducted.

Data included did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias cannot be ruled out. Although no points were deducted, a degree of publication bias is likely because: 1) literature on IGRA is rapidly exploding and currently unpublished studies may come out in future (although we made an attempt to include unpublished studies, our attempt was not comprehensive; we are aware of at least one unpublished study that was not assessed for this review); 2) there are anecdotal examples of unpublished negative studies on IGRA; and 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

All three studies provided incidence rates of TB stratified by IGRA as well as TST status at baseline. (N refers to numbers that entered follow-up)

Serious limitations include lack of proper ascertainment of the TB outcome by smear and/culture, IGRA incorporated in the methods to diagnose TB (South Africa) and lack of adjustment of all confounders. A point is deducted.

The results of the studies could be generalized for the specific country/region and for those specific groups of interest. However, the small number of studies warrants caution; a point is deducted for indirectness.

The two tests perform comparably and any differences are not statistically significant as the 95% confidence intervals for the pooled IRRs overlap and there is no heterogeneity in the pooled estimates for either test (IGRA+: IRR=3.2, $I^2=0\%$, $p=0.899$ and TST+: IRR=2.3, $I^2=0\%$, $p=0.383$). No points deducted.

The confidence intervals of the pooled IRRs are wide (precision > +/- 20%). This is a very serious limitation. Two points are deducted.

Publication bias was not formally assessed, but is deemed likely. See 36.

Only the Zambian study examined if there was an exposure-gradient relationship between baseline quantitative IGRA levels and subsequent rates of TB in those levels. (N refers to numbers included in this stratified analysis)

Lack of proper ascertainment of the TB outcome by smear/culture for both studies. The Zambian study is unpublished and only an interim report was available, so quality could not be fully assessed. A point is deducted.

There is only one study. There is serious indirectness. A point is deducted.

There is only one study; inconsistency cannot be assessed. A point is deducted.

The 95% confidence intervals per IGRA stratum were extremely wide (precision > +/- 20%). Two points are deducted.

Publication bias was not formally assessed, but is deemed likely. See 36.

The Zambia and South Africa studies further explored rates for TB in paired concordant and discordant TST/IGRA results. (N refers to number included in this stratified analysis)
All 3 studies were included in this evaluation of patient-relevant outcomes. The diagnostic accuracy estimates of sensitivity and specificity etc are surrogates of patient-relevant outcomes important for assessing the frequency and impact of either a false negative or false positive IGRA result at baseline. A falsely positive outcome may result in possible isoniazid preventive therapy (IPT) prescription for a period of 6-9 months, depending on country guidelines. IPT, although safe, is not without serious adverse effects, notably, clinical hepatitits and the increased possibility of drug resistance in the future. Whilst a falsely negative result may result in no IPT being provided and the individual exposed to at least a 2-fold risk of developing TB in the future. Serious limitations include lack of proper ascertainment of the TB outcome by smear and/culture, IGRA incorporated in the methods to diagnose TB and lack of adjustment of all confounders for confounding results. Observed 95% confidence intervals around the rates per strata are wide (precision > +/- 20%).

Publication bias was not formally assessed, but is deemed likely. See B6.

Table 23. GRADE Summary of findings: Predictive value of commercial IGRA for incident active TB in low and middle-income Countries

<table>
<thead>
<tr>
<th>Outcome</th>
<th>N (No. of studies)</th>
<th>Principal Findings</th>
<th>What do these findings mean?</th>
<th>Quality of Evidence</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficacy of preventive therapy based on IGRA results</td>
<td>No studies</td>
<td></td>
<td>Moderate increase in incidence rates of TB in IGRA positives compared to IGRA negatives. This</td>
<td>Very low @@@OOO</td>
<td>Critical</td>
</tr>
<tr>
<td>Prospective predictive value of IGRA for the development of active incident TB? (Do IGRA positive results have a stronger)</td>
<td>7,392 (3)</td>
<td>1) IGRA positives results appear to have a moderate but higher statistical association with incident TB compared to IGRA negatives, pooled IRR=3.2 (95% CI 0.74-5.64), I2=0%, p=0.91. This estimate is not statistically significant. The</td>
<td>Moderate increase in incidence rates of TB in IGRA positives compared to IGRA negatives. This</td>
<td>Critical (7-9)</td>
<td>Critical</td>
</tr>
</tbody>
</table>
association with subsequent development of active TB compared to IGRA negative results?)

| Predictive value of IGRA for the development of active incident TB compared to the TST (Do IGRAs (positive vs. negative) have a stronger statistical association with subsequent active TB than the TST (positive vs. negative))? |
|---|---|
| confidence interval includes the null. Furthermore, the small number of studies, the heterogeneity of populations studied all warrants caution when interpreting the pooled results. Despite the lack of evidence for statistical heterogeneity. |
| 2) IGRA positives results appear to have higher rates of incident TB than IGRA negatives. A pooled IR (IGRA+)=$16.5$ (95% CI 11.24-21.7), $I^2=98$%, $p<0.0001$ and IR (IGRA-)=2.85 (95% CI 0.86-4.84), $I^2=35$, $p=0.217$. The 95% CI do not overlap suggesting the difference may be significant. However, this is based on just three studies with different populations. The pooled results should be interpreted with caution; there is low-high statistical heterogeneity. |
| translates to moderate risk of progression. There are too few studies to conclude this with certainty. |

<table>
<thead>
<tr>
<th>Predictive value of IGRA for the development of active incident TB compared to the TST (Do IGRAs (positive vs. negative) have a stronger statistical association with subsequent active TB than the TST (positive vs. negative))?</th>
<th>7,392 (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) IGRA+: Pooled IRR=$3.24$ (0.62-4.69); $I^2=0$, $p=0.90$</td>
<td></td>
</tr>
<tr>
<td>2) TST+: Pooled IRR=$2.3$ (0.83-3.73); $I^2=0$, $p=0.38$</td>
<td></td>
</tr>
<tr>
<td>IGRA+ and TST+ may have a similar strength of association with subsequent TB compared to test negative individuals.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictive value of IGRA for the development of active incident TB compared to the TST (Do IGRAs (positive vs. negative) have a stronger statistical association with subsequent active TB than the TST (positive vs. negative))?</th>
<th>Very low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical</td>
<td>(7-9)</td>
</tr>
<tr>
<td>Predictive value of IGRA for subsequent TB when IGRA are evaluated as part of a multivariable clinical algorithm for predicting TB (Additive value of IGRA)</td>
<td>No studies</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Quantitative IGRA levels and subsequent rates of TB</td>
<td>721 (1)</td>
</tr>
<tr>
<td>Immunological phenotypes of discordant-concordant TST/IGRA pairs and subsequent rates of TB</td>
<td>5,861 (2)</td>
</tr>
</tbody>
</table>

It suggests no exposure-gradient relationship between quantitative IGRA levels and rates of subsequent TB. Rates appeared highest in the lowest IGRA quartile, 0.35-0.64 IU/ml at 73.8/1000PY (23.8-228.94), and not at subsequent higher strata, 0.65-3.94 IU/ml at 30.1 (12.5-72.4), 3.95-10 IU/ml at 0 rate per/1000PY and the highest IGRA quartile of >10 IU/ml at 50/1000PY (18.8-133.1). However, comparisons across the strata are not statistically significant, as confidence intervals overlap and results should be interpreted with caution.

Rates of TB during follow-up may be higher in those with double positive TST+/IGRA+ results than in those with double negative results.

The Zambia study reported higher rates in the discordant pair where IGRA was the positive tests compared to when TST was the positive tests, 29.7/1000PY (13.4 – 66.2) and 0 for IGRA+/TST- and IGRA-/TST+, respectively. By contrast the South African study reported marginally higher rates in IGRA-/TST+ of 3.3/1000PY (0.4-12.0) than in IGRA+/TST- of...
1.8/1000PY (0.4-5.4). However, these differences are not significant as the confidence intervals are wide and overlap.

<table>
<thead>
<tr>
<th>Sensitivity, Specificity, False positive rates etc for active TB (as surrogates of patient relevant outcomes)</th>
<th>7,392 (3)</th>
<th>No pooled results.</th>
<th>IGRA have moderate sensitivity for subsequent TB in keeping with observed moderate rates. This is not different from the TST. False positive rate is similar for both tests. The proportions scored positive by IGRA and TST are similar for the China and South Africa studies. By contrast, the proportion IGRA+ is higher than TST+ for the Zambia study. However, lower TST results may have resulted from logistical issues.</th>
<th>Very low (4-6)</th>
<th>Important (4-6)</th>
</tr>
</thead>
</table>

IGRA sensitivity for incident TB was 88% (64-99), 75% (48-93) and 75% (61-86) for the China (T-Spot.TB), Zambia (QFT-GIT) and South Africa (QFT-GIT) studies, respectively. Specificity was low across the studies at 35% (30-41), 50% (46-54) and 49% (48-51). That means, the false positive rate (100-specificity) for the studies will be 65% (59-70), 50% (46-54) and 51% (49-52). Based on a positive IGRA alone, all these individuals would unnecessarily receive IPT.

TST sensitivity for incident TB was similar at 76% (50-93) and 73% (59-84) for the China and South Africa studies, respectively. Specificity for those studies was 35% (29-41) and 58% (57-58). The proportions that would unnecessarily receive IPT based on IPT alone would be 65% (59-71) and 42% (41-42) for the China and South Africa studies, respectively. By contrast sensitivity for subsequent TB disease was poorest for the Zambia study at 44% (20-70) with a specificity of 67% (64-71). The Zambia study acknowledged logistical issues at the clinical sites that possibly affected TST results.

Utility of repeated or serial IGRA results for predicting subsequent incident active TB | No studies | | | | |

| Utility of repeated or serial IGRA results for predicting subsequent incident active TB | No studies | | | | |
3.7 **Operational aspects on the use of IGRAs in high TB burden countries**

Only a few studies addressed these aspects, mainly in the discussion and not systematically:

- **Cost**
  Cost of IGRAs was mentioned by four studies, mainly stating that the assays are too expensive and therefore a limitation to their use.

- **Reproducibility**
  Only one study addressed reproducibility of T-SPOT by assessing inter-observer agreement, showing excellent correlation. No other study mentioned the issue of test reproducibility.

- **Transport time**
  Twelve studies reported on accepted transport times of samples to the lab, which were mainly <6 hrs, within the limit accepted by the test manufacturers. One study accepted 16 hrs and another 24 hrs transport times. None reported on the impact of the transport times (ie. delay between drawing the blood and initiating the IGRA test) and IGRA test results/performance.

- **Time to result**
  No study reported on time to result for IGRAs.

- **Impact of the use of IGRAs on treatment**
  Four studies reported on the impact of IGRAs on TB therapy. In two studies, IGRA results were reported to clinicians; one study did not discuss the consequences and in the other QFT- positive children received preventive chemotherapy. The other two studies commented on the reduced number of patients that would require preventive therapy if IGRAs were part of the diagnostic algorithm.

- **Feasibility**
  The following aspects related to the feasibility of IGRAs were highlighted:
  - Phlebotomy can be difficult, particularly in very young children;
  - Blood amounts required may be an issue, however tests were performed with <2 ml of blood (T-SPOT) in some studies;
  - Indeterminate results as well as failures due to low cell counts (T-SPOT) may be more frequent in younger children (<4yrs) and immune-suppressed children;
  - Strong interferon response in negative control tubes (high background results) in QFT may reflect the influence of other coincident diseases;
  - Standardization and generation of automated, quantitative results should render IGRAs more objective than TST;
  - A well-equipped laboratory, expensive equipment and training are required for IGRA test performance, which may cause logistical problems.
Annex 1. Meeting participants

Expert Group

Dr Richard A Adegbola
Senior Program Officer
Infectious Diseases, Global Health
Bill & Melinda Gates Foundation
P O Box 23350
Seattle, WA 98102
USA
Richard.adegbola@gatesfoundation.org

Dr Lakhbir Singh Chauhan
Deputy Director General of Health Services
Ministry of Health and Family Welfare
522 "C" Wing, 5th Floor
Nirman Bhavan
110011 - New Delhi
India
ddgtp@rntcp.org

Dr Daniela Cirillo
Head, Emerging Bacterial Pathogens Unit
San Raffaele del Monte Tabor Foundation (HSR), Emerging bacterial pathogens
Via Olgettina 60
20132- Milan
Italy
cirillo.daniela@hsr.it

Dr Anneke C Hesseling
Professor and Director: Paediatric TB Research Program, Desmond Tutu TB Centre
Department of Paediatrics and Child Health,
Faculty of Health Sciences
Stellenbosch University
Private Bag X1
Matieland, 7602
South Africa
annekeh@sun.ac.za

Dr Phillip Hill
McAuley Professor of International Health
Director, Centre for International Health
Department of preventive and Social Medicine
University of Otago School of Medicine
PO BOX 913, Dunedin 9054
New Zealand
philip.hill@otago.ac.nz

Mr Oluwamayowa Joel
Communication for Development Centre
73, Ikosi Road, Ketu,
Lagos State
Nigeria
oluwamayowajoe@gmail.com

Dr Suman Laal
Associate Professor of Pathology & Microbiology
NYU Langone Medical Center
c/o VA Medical Center
423 East 23rd Street, Room 18123N
New York, NY 10010
USA
Suman.Laal@nyumc.org

Dr Philip LoBue
Associate Director for Science
Division of Tuberculosis Elimination
National Center for STD, HIV/AIDS, Viral Hepatitis, and TB Prevention
Centers for Disease Control and Prevention
1600 Clifton Road Mailstop E-04
Atlanta, GA 30333
USA
pgl5@cdc.gov

Dr Holger Schünemann
Department of Clinical Epidemiology & Biostatistics
McMaster University Health Sciences Centre
Room 2C10B, 1200 Main Street, West Hamilton
Canada
schuneh@mcmaster.ca
Systematic reviewers and observers

Dr Adithya Cattamanchi
San Francisco General Hospital
Pulmonary Division – Room 5K1
1001 Potrero Ave
San Francisco, CA 94110
USA
acattamanchi@medsfgh.ucsf.edu

Dr Anand Date
HIV/AIDS Care & Treatment Branch
(HIV/TB)
Global AIDS Program
Centers for Disease Control & Prevention
1600 Clifton Road Mailstop E-04
Atlanta, GA 30333
USA
adate@cdc.gov

Dr Anne Detjen
Technical Consultant
The Union North America Office
International Union Against Tuberculosis
and Lung Disease
61 Broadway, Suite 1720
New York, NY 10006
USA
adetjen@theunion.org

Dr Richard Menzies
Montreal Chest Institute
3650 St. Urbain St.
Montreal, PQ
Canada H2X 2P4
Dick.Menzies@Mcgill.ca

Dr John Metcalfe
Division of Pulmonary and Critical Care Medicine
University of California, San Francisco
Division of Epidemiology
University of California, Berkeley
230 Santa Paula Ave.
San Francisco, CA 94127
USA
John.Metcalfe@ucsf.edu

Dr Richard J O’Brien
Foundation for New Innovative New Diagnostics
16 Avenue de Budé
1202 Geneva
Switzerland
rick.obrien@finddiagnostics.org

Dr Madhukar Pai
McGill University
Dept of Epidemiology & Biostatistics
1020 Pine Ave West
Montreal, QC H3A 1A2
Canada
madhukar.pai@mcgill.ca

Dr Molebogeng Rangaka
Institute of Infectious Diseases and Molecular Medicine
School of Health Sciences University of Cape Town
Observatory, 7925
South Africa
molebogeng.rangaka@lshtm.ac.uk
mrxrangaka@yahoo.co.uk

Dr Karen R Steingart
Physician Consultant
Curry International Tuberculosis Center
University of California, San Francisco
3180 18th Street, Suite 101
San Francisco, CA 94110-2028
USA
karenst@uw.edu

Dr Alice Zwerling
Montreal Chest Institute, Rm K3.09
3650 St Urbain, Montreal, Quebec H2X 2P4
Canada
alice.zwerling@mail.mcgill.ca

Dr Peter Godfrey-Faussett
Department of Infection & Tropical Diseases
London School of Hygiene & Tropical Medicine
Keppel Street
WC1E 7HT - London
United Kingdom
pgf@lshtm.ac.uk
Annex 2. Declarations of Interest

None declared

R. Adegbola
A. Cattamanchi
L. Chauhan
D. Cirillo
A. Hesseling
P. Hill
P. LoBue
H. Schüneman
J. Oluwamayowa
S. Laal

Declared, insignificant

R. O’Brien (FIND support to academia to develop POC serodiagnostic test, FIND biomarker discovery project)

Declared, significant (observer status)

D. Dowdy (relevant research, participation in systematic review)
M. Pai (relevant research, participation in systematic reviews)
J. Metcalf (principal systematic reviewer)
K. Steingart (principal systematic reviewer)
P. Godfrey-Fausset (relevant research)
A. Zwerling (principal systematic reviewer)
Annex 3. Selection of studies evaluating the use of IGRA in the diagnosis of active TB

Titles/abstracts identified and screened for full-text retrieval: 789

Excluded based on title and abstract: 621

Full papers retrieved for more detailed evaluation: 168

Excluded: 134

*Reasons*
- Children: 1
- Duplicate data: 2
- Extrapulmonary TB: 2
- Less than 10 TB patients: 3
- LTBI: 93
- Noncommercial IGRA: 8
- Nonstandard IGRA method: 3
- Older generation IGRA: 22

Added from prior systematic review: 17

Pulmonary TB, all countries: 51

High income countries: 32

Unpublished investigations: 3

Pulmonary TB, Low/middle income countries
Papers 22 (studies 33)
Included studies


Excluded studies
(Reasons for exclusion in parenthesis after reference)

1. Targeted tuberculin testing and treatment of latent tuberculosis infection. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, July 1999. This is a Joint Statement of the American Thoracic Society (ATS) and the Centers for Disease Control and Prevention (CDC). This statement was endorsed by the Council of the Infectious Diseases Society of America. (IDSA), September 1999, and the sections of this statement. Am J Respir Crit Care Med. 2000 Apr;161(4 Pt 2):S221-47. (LTBI)


126


55. Huang SP. Enzyme-linked immunosopt. Chin_J_Infect_Chemo. 2009. (Nonstandard IGRA method)


59. Kabeer BS, Sikhamani R, Raja A. Comparison of interferon gamma and interferon gamma-inducible protein-10 secretion in HIV-tuberculosis patients. AIDS. 2009 Dec 10. (Duplicate)


79. Luetkemeyer AF, Charlebois ED, Flores LL, Bangsberg DR, Deeks SG, Martin JN, et al. Comparison of an interferon-gamma release assay with tuberculin skin testing in HIV-infected individuals. Am J Respir Crit Care Med. 2007 Apr;175(7):737-42. (LTBI)


98. Pai M. Guidelines on IGRAs: concordant or discordant? 2nd Global Symposium on IGRAs. 2009. (LTBI)


125. Torres H, Zapico M, Vivas S, Mostaza J, Blanco J, Ruiz de Morales J. Aplicación clínica de una prueba de producción de interferón gamma para el diagnóstico de


Annex 4. Selection of studies evaluating the use of IGRAs in children

Included studies


Studies excluded after full text review
(Reasons for exclusion in parenthesis after reference)


27. Ozekinci T, Ozbek E, Celik Y. Comparison of tuberculin skin test and a specific T-cell-based test, T-Spot (TM).TB, for the diagnosis of latent tuberculosis infection. Journal of International Medical Research 2007;35:696-703. (Data not available)


Annex 5. Selection of studies evaluating the use of IGRAs for the diagnosis of LTBI in HIV-positive individuals

Included studies


2. Clark SA, Martin SL, Pozniak A, et al. Tuberculosis antigen-specific immune responses can be detected using enzyme-linked immunospot technology in human


**Studies excluded after full text review**

(Reasons for exclusion in parenthesis after reference)


46. Kurup SK, Buggage RR, Clarke GL, Ursea R, Lim WK, Nussenblatt RB. Gamma interferon assay as an alternative to PPD skin testing in selected patients with


Annex 6. Selection of studies evaluating the use of IGRAs for tuberculosis screening of health care workers

IGRA studies identified from electronic databases sources

IGRA studies not done in health care workers n=490

IGRA studies with Health Care Workers n=56

Unpublished articles and conference presentations eligible for inclusion n=4

Did not meet eligibility criteria n=12 (point source outbreaks (7), treatment monitoring (2), short term)

Unique studies included in final review
Included studies


Annex 7. Selection of studies evaluating the use of IGRAs in LTBI screening in contact and outbreak investigations

Included studies


Studies excluded after full text review
(Reasons for exclusion in parenthesis after reference)


15. Diel R, Loddenkemper R, Meywald-Walter K, Gottschalk R, Nienhaus A. Comparative Performance of Tuberculin Skin Test, QuantiFERON-TB-Gold In Tube Assay, and T-
Spot.TB Test in Contact Investigations for Tuberculosis. Chest. 2008 Nov 18. (High-income country setting)


43. Janssens JP. Interferon-gamma release assay tests to rule out active tuberculosis. Eur Respir J. 2007 Jul;30(1):183-4; author reply 4-5. (High-income country setting)


71. Silverman MS, Reynolds D, Kavask PA, Garay J, Daly A, Davis I. Use of an interferon-gamma based assay to assess bladder cancer patients treated with intravesical BCG and exposed to tuberculosis. Clin Biochem. 2007 Apr 27. (High-income country setting)


Annex 8. Selection of studies evaluating the predictive value of IGRAs for incident active TB disease in low, middle and high income countries

- **Total identified published records that met search criteria up to 31 May 2010** (n = 724)
- **Additional records identified through other sources** (n = 3)

- **Abstracts and Reports Retrieved for Screening**

- **Records screened** (n = 727)
  - **Records excluded** (n = 708)

- **Full-text articles assessed for eligibility** (n = 19)
  - **Full-text articles excluded** (n = 7)

- **Studies included in qualitative synthesis** (n = 12)

- **Studies included in quantitative synthesis (meta-analysis)** (n = 6 and 7, for estimation of rates and incidence rate ratios, respectively)
  - (n = 12, for cumulative incidence risk ratios)
Included studies


Studies excluded after full text review
(Reasons for exclusion in parenthesis after reference)

1. Clark SA, Martin SL, Pozniak A, Steel A, Ward B, et al. (2007) Tuberculosis antigen-specific immune responses can be detected using enzyme-linked immunospot technology in human immunodeficiency virus (HIV)-1 patients with advanced disease. Clin Exp Immunol 150: 238-244. (Predictive value of IGRA was not a primary objective, furthermore, a mixture of patients that included those who were on TB treatment already, those who were TB suspects, and ‘healthy’ patients was followed-up; the risk of subsequent TB was not stratified by the various sub-groups (prevalent and incident cases were combined).
4. Haldar P (2009) Contact screening with single-step TIGRA testing and risk of active TB infection: The Leicester Cohort. Thorax. (Conference Abstract) (Study identified only in abstract form, but could not be included because of insufficient information)


7. Ordway DJ, Costa L, Martins M, Silveira H, Amaral L, et al. (2004) Increased Interleukin-4 production by CD8 and gammadelta T cells in health-care workers is associated with the subsequent development of active tuberculosis. J Infect Dis 190: 756-766. (RD1 antigens were not used)