Interferon-\(\gamma\) Release Assays for Active Pulmonary Tuberculosis Diagnosis in Adults in Low- and Middle-Income Countries: Systematic Review and Meta-analysis

John Z. Metcalfe,1,2 Charles K. Everett,1 Karen R. Steingart,2 Adithya Cattamanchi,1,2 Laurence Huang,1,3 Philip C. Hopewell,1,2 and Madhukar Pai4

Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital,2 Department of Health Services, University of Washington School of Public Health, Seattle,3 HIV/AIDS Division, San Francisco General Hospital, University of California; and4 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

Background. The diagnostic value of interferon-\(\gamma\) release assays (IGRAs) for active tuberculosis in low- and middle-income countries is unclear.

Methods. We searched multiple databases for studies published through May 2010 that evaluated the diagnostic performance of QuantiFERON-TB Gold In-Tube (QFT-GIT) and T-SPOT.TB (T-SPOT) among adults with suspected active pulmonary tuberculosis or patients with confirmed cases in low- and middle-income countries. We summarized test performance characteristics with use of forest plots, hierarchical summary receiver operating characteristic (HSROC) curves, and bivariate random effects models.

Results. Our search identified 789 citations, of which 27 observational studies (17 QFT-GIT and 10 T-SPOT) evaluating 591 human immunodeficiency virus (HIV)–uninfected and 844 HIV-infected individuals met inclusion criteria. Among HIV-infected patients, HSROC/bivariate pooled sensitivity estimates (highest quality data) were 76% (95% confidence interval [CI], 45%–92%) for T-SPOT and 60% (95% CI, 34%–82%) for QFT-GIT. HSROC/bivariate pooled specificity estimates were low for both IGRA platforms among all participants (T-SPOT, 61% [95% CI, 40%–79%]; QFT-GIT, 52% [95% CI, 41%–62%]) and among HIV-infected persons (T-SPOT, 52% [95% CI, 40%–63%]; QFT-GIT, 50% [95% CI, 35%–65%]). There was no consistent evidence that either IGRA was more sensitive than the tuberculin skin test for active tuberculosis diagnosis.

Conclusions. In low- and middle-income countries, neither the tuberculin skin test nor IGRAs have value for active tuberculosis diagnosis in adults, especially in the context of HIV coinfection.

Interferon-\(\gamma\) release assays (IGRAs) are the first new diagnostic test for latent tuberculosis (LTBI) in >100 years. Newest generation IGRAs measure interferon (IFN)–\(\gamma\) secretion after exposure of whole blood (QuantiFERON-TB Gold In-Tube [QFT-GIT], Cellestis) or peripheral blood mononuclear cells (T-SPOT.TB [T-SPOT], Oxford Immunotec) to antigens encoded in the region of difference–1 (RD1), a portion of the Mycobacterium tuberculosis genome absent among all bacille Calmette-Guérin (BCG) strains and most non-tuberculous mycobacteria [1]. We have shown in previous systematic reviews that compared with the tuberculin skin test (TST), IGRAs have higher specificity for LTBI in settings with low tuberculosis incidence, better correlation with surrogate measures of M. tuberculosis exposure, and less cross-reactivity with the BCG vaccine [2–4]. Thus, in recent years, IGRAs have become widely endorsed in high-income countries for diagnosis of LTBI [5–7]. However, IGRAs were explicitly designed to replace the TST in diagnosis of LTBI and were not intended for active tuberculosis, which is a microbiological diagnosis. Furthermore, diagnosis and treatment of LTBI remains...
limited in scope in most low- and middle-income countries, where detection and management of active tuberculosis is of highest priority for national tuberculosis programs. Because IGRAsthe TST, cannot distinguish LTBI from active tuberculosis [8–10], these tests can be expected to have poor specificity for active tuberculosis in all high-burden settings because of a high background prevalence of LTBI [11]. Additional differences in patient spectrum, such as anergy due to advanced disease, malnutrition, and human immunodeficiency virus (HIV)–associated immune suppression, or characteristics of the setting, such as laboratory procedures and infrastructure, may also contribute to a lower performance of IGRA observed in these settings [12]. However, private sector laboratories in high-burden countries increasingly use IGRAsthe TST, for active tuberculosis diagnosis [13], and many investigators continue to recommend the use of IGRAsthe TST, for active tuberculosis diagnosis [14–17].

Because of unclear benefits and potential costs to patients and national tuberculosis programs, we conducted a systemic review and meta-analysis to determine IGRA test performance in persons with suspected or confirmed active pulmonary tuberculosis living in low- and middle-income settings.

METHODS

Overview
Because of the absence of studies evaluating patient-important outcomes in persons with suspected tuberculosis who were randomized to treatment on the basis of IGRA results, we focused our review on the diagnostic accuracy of IGRAsthe TST, for active tuberculosis. We observed standard guidelines and methods for systematic reviews and meta-analyses of diagnostic tests [18–21].

Search Methods
We previously published systematic and narrative reviews on the accuracy and performance of IGRAsthe TST, in various subgroups [2–4, 10, 12]. We updated the previous literature searches to identify all studies evaluating IGRAsthe TST, published through May 2010. We searched PubMed, Embase, Biosis, and Web of Science for studies in all languages. The search terms used included “interferon-gamma release assay,” “T cell–based assay,” “antigen-specific T cell,” “T cell response,” “T-cell response,” “interferon,” “interferon-gamma,” “gamma-interferon,” “IFN,” “elispot,” “ESAT-6,” “CFP-10,” “culture filtrate protein,” “enzyme-linked immunosorbent spot,” “Quantiferon,” “Quantiferon-TB,” “tuberculosis,” and “Mycobacterium tuberculosis.” In addition to database searches, we reviewed bibliographies of reviews and guidelines, screened citations of all included studies, searched clinicaltrials.gov for ongoing studies, and contacted both experts in the field and IGRA manufacturers to identify additional published and unpublished studies. We requested pertinent information not reported in the original publication from the primary authors of all studies included in the review.

Study Selection and Data Collection
We included studies that evaluated the performance of the most recent generation of commercial, RD1 antigen-based IGRAsthe TST, (QFT-GIT and T-SPOT) among adults (age ≥15 years) with suspected active pulmonary tuberculosis or confirmed tuberculosis in low- and middle-income countries [22]; the World Bank Country Classification was considered as a surrogate for national tuberculosis incidence. HIV infection was established either by documented serological testing or self-report. We excluded (1) studies that evaluated noncommercial (in-house) IGRAsthe TST, purified protein derivative–based IGRAsthe TST, QuantiFERONTB Gold (2G), and IGRAsthe TST, performed using specimens other than blood; (2) longitudinal data focused on the effect of antituberculosis treatment on IGRA response; (3) studies including <10 eligible individuals; (4) studies focused on extrapulmonary tuberculosis or children (age <15 years); (5) studies reporting insufficient data to determine diagnostic accuracy measures; and (6) conference abstracts, letters without original data, and reviews.

At least 2 reviewers (J. Z. M., C. K. E., K. R. S., or A. C.) independently screened the accumulated citations for relevance, reviewed full-text articles using the prespecified eligibility criteria, and extracted data with use of a standardized form. The reviewers resolved disagreements about study selection and data extraction by consensus.

Assessment of Study Quality
Because primary outcomes for this systematic review focus on test accuracy, we evaluated study quality with use of a subset of relevant criteria from the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool, a validated tool for diagnostic accuracy studies [23]. Because of growing concerns about conflicts of interest in diagnostic studies and guidelines [24, 25], we also reported whether IGRA manufacturers had any involvement with the design or conduct of each study, including donation of materials, monetary support, work and/or financial relationships with study authors, and participation in data analysis.

Outcome Definitions
Well-designed diagnostic accuracy studies focus on a representative target population in whom genuine diagnostic uncertainty exists (ie, patients for whom clinicians would apply the test in the course of regular clinical practice) [26]. There is evidence that diagnostic studies that include only known patients with the condition of interest and healthy control subjects without this condition tend to overestimate test accuracy [27]. Therefore, we considered studies simultaneously evaluating IGRA sensitivity and specificity among persons with suspected active tuberculosis to represent the highest quality evidence, whereas studies evaluating IGRA performance among patients with known active tuberculosis (for sensitivity) were considered to be of lesser quality. Because of our focus on active tuberculosis diagnostic accuracy and the high prevalence of LTBI in settings with a high
tuberculosis burden, IGRA specificity was estimated exclusively among studies enrolling persons with suspected active tuberculosis for whom the diagnostic examination ultimately showed no evidence of active disease.

A hierarchy of reference standards for active tuberculosis was developed a priori to judge the quality of each individual assessment of IGRA diagnostic accuracy. From most to least favorable, these reference standards included (1) culture confirmation or sputum smear positivity in settings with high tuberculosis incidence (≥50 cases/100,000 population), where sputum smear microscopy has been shown to have high specificity [28]; (2) sputum smear positivity without culture confirmation in settings with low or intermediate tuberculosis incidence (<50 cases/100,000 population); and (3) clinical diagnosis based on presenting symptoms, radiologic findings, and/or response to tuberculosis treatment without microbiological confirmation. Because the TST remains in widespread use and indeterminate IGRA results may affect assay performance in low-income settings, we also evaluated (1) observed differences in sensitivity for active tuberculosis diagnosis between IGRA and TST, and (2) the proportion of IGRA results among patients with active disease that were indeterminate.

We used the following definitions for primary outcomes: (1) sensitivity was defined as the proportion of individuals with a positive IGRA result among those with culture-positive tuberculosis (we included indeterminate IGRA results in the denominator if they occurred in individuals with culture-positive tuberculosis), and (2) specificity was defined as the proportion of individuals with a negative IGRA result among those who had active tuberculosis disease ruled out (indeterminate IGRA results were excluded from analysis). With use of the Grading of Recommendations Assessment, Development and Evaluation framework [26], these measures can be interpreted as surrogates for patient-important outcomes.

**Data Synthesis and Meta-Analysis**

Multiple sources of heterogeneity frequently exist when summarizing estimates from studies of diagnostic tests [29]. We adopted the following approach to account for expected heterogeneity. First, when possible, we separately synthesized data for each commercial IGRA and by HIV status. The prespecified subgroups minimize heterogeneity related to differences in testing platform (enzyme-linked immunosorbent assay vs enzyme-linked immunospot assay), 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, we visually assessed heterogeneity with use of forest plots, characterized the variation in study results attributable to heterogeneity (I-squared value), and statistically tested for heterogeneity (χ² test) [29]. Third, we calculated pooled sensitivity and specificity estimates with use of random effects modeling, which provides more conservative estimates than does fixed effects modeling when heterogeneity is a concern [19, 30].

For each individual study, we assessed all outcomes for which data were available. First, we generated forest plots to display the individual study estimates and their 95% confidence intervals (CIs). Second, we used bivariate random effects regression models [31] when both sensitivity and specificity could be reported from the same population of tuberculosis suspects. Because pooling sensitivity and specificity separately can produce biased estimates of test accuracy [19], we preferred to generate pooled estimates when both sensitivity and specificity were reported in a study and ranked this as higher-quality evidence. Third, we generated hierarchical summary receiver operating characteristic (HSROC) curves to summarize the global test performance [30]. Because of the need to summarize 2 correlated measures (eg, sensitivity and specificity) and because substantial between-study heterogeneity is common, meta-analysis of diagnostic accuracy requires different and more complex methods than do traditional meta-analytic techniques. Graphically illustrating the trade-off between sensitivity and specificity, HSROC curves differ from traditional ROC curves in allowing accuracy to vary by each individual study (ie, allowing for random effects and, thus, 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 [32]. These 2 methods generally produce similar results and are both recommended by the Cochrane Diagnostic Test Accuracy Working Methods group [20]. We calculated pooled estimates when at least 4 studies were available in any subgroup and summarized individual study results when <4 studies were available. We performed all analyses with use of Stata, version 11 (StataCorp). For bivariate random effects regression and HSROC analyses, we used the user-written “metandi” program for Stata [31].

**RESULTS**

**Search Results**

The initial search yielded 789 citations (Figure 1). After full-text review of 168 articles, 19 [15, 17, 33–49] were determined to meet eligibility criteria for IGRA evaluation of active tuberculosis in low- and middle-income settings. Because some articles included >1 commercial IGRA, there were 27 unique evaluations (referred to as studies; 17 of QFT-GIT and 10 of T-SPOT) that included a total of 590 HIV-uninfected and 844 HIV-infected individuals.

**Study Characteristics**

Of the total studies, 7 (26%) were from low-income countries and 20 (74%) were from middle-income countries. Fourteen
studies (52%) included HIV-infected individuals, and 21 (78%) studies included ambulatory patients (Table 1). IGRAs were performed for persons suspected of having active tuberculosis in 14 studies (52%) [34, 36–38, 40, 41, 46, 47, 49] and in persons with known active tuberculosis in 13 studies (48%) [15, 17, 33, 35, 39, 42–45, 48]. A list of excluded studies and reasons for exclusion is available from the authors on request.

**Study Quality**

The majority of studies satisfied the QUADAS criteria assessed (Figure 2), with the exception of patient spectrum (biased sampling) and blinding. Sixteen studies (59%) did not enroll a representative spectrum of patients, and 9 (33%) did not clearly report whether assessment of the reference standard was performed with blinding to IGRA results. Industry involvement was unknown in 5 studies (19%) and acknowledged in 8 (30%), including donation of IGRA kits (6 studies) and work and/or financial relationships between authors and IGRA manufacturers (2 studies).

**Sensitivity and Specificity Estimation Among Persons With Suspected Tuberculosis**

We identified a total of 14 studies that simultaneously estimated sensitivity and specificity among persons with suspected tuberculosis, and test accuracy estimates were pooled using bivariate random effects and/or HSROC methods (these studies were ranked as high-quality evidence). Overall, studies enrolling persons with suspected active tuberculosis revealed a sensitivity of 83% (95% CI, 63%–94%) and specificity of 61% (95% CI, 40%–79%) for T-SPOT (6 studies) and a sensitivity of 69% (95% CI, 52%–83%) and specificity of 52% (95% CI, 41%–63%) for QFT-GIT (8 studies).

**Sensitivity**

With the exception of 2 studies [36, 47], the sensitivity of IGRAs was assessed on the basis of a positive culture result (21 studies [78%]) or a positive sputum acid-fast bacilli smear result in a setting with high tuberculosis incidence (4 studies [15%]). Among studies performed in patients with known active tuberculosis, 6 (46%) included patients who had been treated for >1 week.

**HIV-Infected Persons.**

Nine studies assessed IGRA sensitivity among HIV-infected persons with suspected active tuberculosis. HSROC and/or bivariate pooled sensitivity estimates were higher for T-SPOT (76%; 95% CI, 45%–92%; 4 studies [34, 37, 40, 41]) than for QFT-GIT (60%; 95% CI, 34%–82%; 5 studies [37, 38, 40, 41, 49]) (Figure 3). Pooled sensitivity estimates did not change appreciably for either T-SPOT (68%; 95% CI, 56%–80%; 5 studies [15, 34, 40–42]) or QFT-GIT (65%; 95% CI, 52%–77%; 7 studies [33, 38, 40, 41, 48, 49]) when studies evaluating patients with known active tuberculosis were included in the analysis (Figure 4). Pooled sensitivity estimates for both T-SPOT (I-squared, 72%; P < .01) and QFT-GIT (I-squared, 76%; P < .001) showed significant heterogeneity.

**HIV-Uninfected Persons.**

Five studies assessed IGRA sensitivity among HIV-uninfected persons with suspected active tuberculosis; data were insufficient to report HSROC and/or bivariate pooled sensitivity estimates for either QFT-GIT [36, 37, 47] or T-SPOT [37, 46]. Pooled sensitivity estimates were similar for T-SPOT (88%; 95% CI, 81%–95%; 4 studies [17, 37, 43, 46]) and QFT-GIT (84%; 95% CI, 78%–91%; 9 studies [10, 33, 35–37, 39, 45, 47, 48]) when studies evaluating patients with known active tuberculosis were included in the analysis (Figure 5). Pooled sensitivity estimates showed significant heterogeneity for QFT-GIT (I-squared, 60%; P = .01) but not for T-SPOT (I-squared, 28%; P = .25).

**Comparisons of QFT-GIT and T-SPOT Sensitivity.** Overall, 4 studies (3 involving HIV-infected patients [37, 40, 41] and 1 involving HIV-uninfected persons [37]) reported comparisons of T-SPOT and QFT-GIT sensitivity. T-SPOT sensitivity was higher but not significantly different from QFT-GIT sensitivity (sensitivity difference, 19%; 95% CI, −17% to 56%; P = .3) (Table 2). Results were similar when restricted to HIV-infected individuals.

**Comparisons of TST and IGRA Sensitivity.** Overall, 9 studies reported comparisons of TST and IGRA (3 T-SPOT and 6 QFT-GIT) sensitivity. TST sensitivity in the 5 studies [17, 39,
involving HIV-uninfected patients was higher (78%; 95% CI, 71%–86%) than that in the 4 studies (15, 38, 45, 48) involving HIV-infected patients (45%; 95% CI, 15%–75%). IGRA sensitivity was not statistically different from TST sensitivity for either T-SPOT (sensitivity difference, 23%; 95% CI, 0%–45%; P < .05) or QFT-GIT (sensitivity difference, 7%; 95% CI, −9% to 23%; P = .37) (Figure 6). There was significant heterogeneity for both estimates (I-squared, 75%; P < .001). Data were insufficient to form HIV-stratified pooled sensitivity difference estimates for either IGRA.

### Specificity

All specificity estimates were determined in persons with suspected tuberculosis with use of HSROC and/or bivariate techniques. Overall, pooled specificity was low for both T-SPOT (61%; 95% CI, 40%–79%; 6 studies) and QFT-GIT (52%; 95% CI, 41%–62%; 8 studies). When restricted to HIV-infected persons with suspected active tuberculosis, pooled specificity for T-SPOT (52%; 95% CI, 40%–63%; 4 studies [34, 37, 40, 41]) was similar to that for QFT-GIT (50%; 95% CI, 35%–65%; 5 studies [37, 38, 40, 41, 49]) (Figure 3). An insufficient number of studies were available to estimate pooled specificity for HIV-uninfected patients.

### Proportion of Indeterminate IGRA Results

The proportion of indeterminate IGRA results among patients with suspected or confirmed active tuberculosis varied considerably (range of 0%–26% among studies enrolling ≥50 participants). The proportion of indeterminate results was low (4%; 95% CI, 1%–7%) among HIV-uninfected patients, regardless of IGRA platform (Figure 1; online only). However,
the proportion of indeterminate results was considerably higher among HIV-infected patients for both QFT-GIT (15%; 95% CI, 9%–21%; 8 studies) and T-SPOT (9%; 95% CI, 0%–17%; 6 studies) (Figure 2; online only). Results were similar for HIV-infected patients when stratified by persons with suspected tuberculosis and persons with known active tuberculosis.

DISCUSSION

The vast majority of the estimated annual 9.4 million new cases of active tuberculosis and 1.7 million tuberculosis-related deaths occur in low- and middle-income countries [50]. Because of resource constraints, public health policies have appropriately placed limited emphasis on diagnosis and treatment of LTBI in these settings. Clinical use of IGRAs, however, has expanded dramatically in recent years, especially in the private sector [13]. Because of their high burden of disease and emerging economies, these countries (eg, India, South Africa, Brazil, and China) represent a potentially lucrative market for commercial IGRAs. Although IGRAs are intended for LTBI and not active tuberculosis disease, and although these tests cannot distinguish between latent infection and active disease, there is concern about increasing use of IGRAs for active tuberculosis in high-burden countries. In this systematic review focused on individuals living in low- and middle-income countries, the highest-quality evidence from persons with suspected tuberculosis demonstrated sensitivity of 69%–83% and specificity of 52%–61% for IGRAs.
in the diagnosis of active tuberculosis. Furthermore, there was no consistent evidence that either IGRA was more sensitive than the TST for active tuberculosis diagnosis.

The majority of evidence for the diagnostic accuracy of IGRAs to date has been summarized from high-income settings where active tuberculosis has been used as a surrogate reference standard for LTBI diagnosis [4, 14]. However, diagnostic test performance (eg, sensitivity and specificity) can be expected to vary according to disease prevalence and other population characteristics [51, 52]. Likewise, clinicians have been advised to base their decisions on studies that most closely match their own clinical circumstances [53].

IGRAs were designed as diagnostic tests of LTBI, though the lack of an accepted gold standard for LTBI has been a significant

<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>65 (52, 76)</td>
<td>16</td>
</tr>
<tr>
<td>Kaiser 2009</td>
<td>India</td>
<td>66 (59, 80)</td>
<td>15</td>
</tr>
<tr>
<td>Leidi 2009</td>
<td>Uganda</td>
<td>74 (49, 91)</td>
<td>12</td>
</tr>
<tr>
<td>Markova 2009</td>
<td>Bulgaria</td>
<td>92 (64, 100)</td>
<td>14</td>
</tr>
<tr>
<td>Ruby 2008</td>
<td>Zambia</td>
<td>63 (49, 75)</td>
<td>16</td>
</tr>
<tr>
<td>Tsiouna 2006</td>
<td>South Africa</td>
<td>85 (44, 83)</td>
<td>13</td>
</tr>
<tr>
<td>Veldsman 2009</td>
<td>South Africa</td>
<td>30 (15, 49)</td>
<td>14</td>
</tr>
<tr>
<td>Subtotal (I-squared=76%, p&lt;0.001)</td>
<td>65 (52, 77)</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>TSPOT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattamanchi 2010</td>
<td>Uganda</td>
<td>54 (45, 64)</td>
<td>25</td>
</tr>
<tr>
<td>Jiang 2009</td>
<td>China</td>
<td>66 (47, 81)</td>
<td>19</td>
</tr>
<tr>
<td>Leidi 2009</td>
<td>Uganda</td>
<td>89 (67, 99)</td>
<td>20</td>
</tr>
<tr>
<td>Markova 2009</td>
<td>Bulgaria</td>
<td>92 (32, 86)</td>
<td>12</td>
</tr>
<tr>
<td>Oni 2010</td>
<td>South Africa</td>
<td>68 (57, 78)</td>
<td>25</td>
</tr>
<tr>
<td>Subtotal (I-squared=72%, p=0.01)</td>
<td>58 (56, 80)</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB in human immunodeficiency virus (HIV)-infected persons with confirmed active tuberculosis in low- and middle-income countries. The forest plots display the sensitivity estimates obtained from individual studies and pooled estimates derived from random effects (DerSimonian–Laird) modeling.

<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>15</td>
</tr>
<tr>
<td>Chegou 2009</td>
<td>South Africa</td>
<td>96 (78, 100)</td>
<td>13</td>
</tr>
<tr>
<td>Chen 2009</td>
<td>China</td>
<td>85 (71, 84)</td>
<td>12</td>
</tr>
<tr>
<td>Dheda 2009</td>
<td>South Africa</td>
<td>73 (45, 92)</td>
<td>5</td>
</tr>
<tr>
<td>Katyar 2008</td>
<td>India</td>
<td>95 (87, 99)</td>
<td>18</td>
</tr>
<tr>
<td>Pai 2007</td>
<td>India</td>
<td>74 (60, 84)</td>
<td>12</td>
</tr>
<tr>
<td>Ruby 2008</td>
<td>Zambia</td>
<td>84 (68, 94)</td>
<td>11</td>
</tr>
<tr>
<td>Tahereh 2010</td>
<td>Iran</td>
<td>77 (59, 90)</td>
<td>9</td>
</tr>
<tr>
<td>Tsiouna 2006</td>
<td>South Africa</td>
<td>77 (46, 95)</td>
<td>5</td>
</tr>
<tr>
<td>Subtotal (I-squared = 60%, p=0.01)</td>
<td>84 (78, 91)</td>
<td>100</td>
<td></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>10</td>
</tr>
<tr>
<td>Ozerkici 2007</td>
<td>Turkey</td>
<td>93 (76, 99)</td>
<td>26</td>
</tr>
<tr>
<td>Shao-ling 2009</td>
<td>China</td>
<td>91 (71, 99)</td>
<td>19</td>
</tr>
<tr>
<td>Soyat 2008</td>
<td>Turkey</td>
<td>81 (72, 88)</td>
<td>40</td>
</tr>
<tr>
<td>Subtotal (I-squared = 28%, p=0.25)</td>
<td>88 (81, 95)</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB among human immunodeficiency virus (HIV)-uninfected persons with confirmed active tuberculosis in low- and middle-income countries. The forest plots show the sensitivity estimates obtained from individual studies and pooled estimates derived from random effects (DerSimonian–Laird) modeling.
Regardless of HIV status. These data suggest that 1 in 2 patients with suspected tuberculosis. As expected, because of the higher background LTBI prevalence and the known inability of IGRA specificity of both IGRAs for active tuberculosis was low, re-sults diagnosis is best estimated only in studies evaluating per-
sonal studies and pooled estimates derived from random effects (DerSimonian–Laird) modeling. *Studies involving human immunodeficiency virus (HIV)–infected patients.

limitation in establishing test performance. In contrast, adequate and commonly used reference standards exist for diagnosing active tuberculosis. Among studies that enrolled persons with suspected active tuberculosis (ie, patients with diagnostic un-certainty), both IGRA demonstrat-suboptimal rule-out value for active tuberculosis. In other words, approximately 1 in 4 patients with culture-confirmed active tuberculosis can be expected to have negative IGRA results in low- and middle-income countries; this has consequences for patients in terms of mor-bidity and mortality. Although high-quality data were limited, sensitivity of both IGRA was lower among HIV-infected pa-tients (60%–70%), suggesting that ~1 in 3 HIV-infected pa-tients with active tuberculosis will have negative IGRA results. The few available comparisons between QFT-GIT and T-SPOT revealed higher sensitivity for the T-SPOT platform, although this difference did not reach statistical significance. Lastly, comparisons with pooled estimates of TST sensitivity were diff-icult to interpret because of substantial heterogeneity. Our results, however, suggest that neither IGRA platform may be more sensitive than the TST for active tuberculosis diagnosis in low- and middle-income countries.

IGRA specificity in diagnosing LTBI, estimated among indi-viduals at low risk for tuberculosis exposure in settings with low tuberculosis incidence (high-income settings), is known to be high (≥98%) [4]. In contrast, specificity for active tuberculosis diagnosis is best estimated only in studies evaluating per-sons with suspected tuberculosis. As expected, because of the higher background LTBI prevalence and the known inability of IGRA to differentiate LTBI from active tuberculosis [10], the specificity of both IGRA for active tuberculosis was low, reg-ardless of HIV status. These data suggest that 1 in 2 patients without active tuberculosis will have positive IGRA results; this has consequences for patients because of unnecessary therapy for tuberculosis and its attendant risks. Studies demonstrating activated T-cell IFN-γ response throughout the entire spectrum of tuberculosis, from latency to active disease [54], lend biologic plausibility to our findings. Even in the spectrum of latent tu-berculosis infection [55], activated T-cell IFN-γ responses occur throughout each phase, with the possible exception of the innate immune response (which eliminates M. tuberculosis without priming a T-cell immune response).

The goal of our systematic review was to critically evaluate the diagnostic accuracy of IGRAs for active tuberculosis diagnosis in low- and middle-income settings. However, there are inherent limitations to sensitivity, specificity, and predictive values as measures of test performance. These measures are unable to determine the extent to which a test may improve on readily available clinical information [56] or the degree to which patient-important outcomes are improved by test results [26]. Although limited, available data suggest that IGRA may add little to the conventional diagnostic investigation for active tu-berculosis in settings with low [57] and high tuberculosis in-cidence [58]. Additional work is necessary to confirm this.

Our meta-analysis has several limitations. First, as with pre-vious systematic reviews [4, 14], heterogeneity was substantial for the primary outcomes of sensitivity and specificity. We used empirical random effects weighting, excluded all studies con-tributing <10 eligible individuals, and separately synthesized data for currently manufactured IGRA to minimize heteroge-neity. Second, World Bank income classification is an imperfect surrogate for national tuberculosis incidence. Although no standard criteria currently exist for defining countries with high tuberculosis incidence, our results were fundamentally un-changed when restricted to nations with a World Health Orga-nization (WHO)-defined annual tuberculosis incidence of ≥50 cases/100 000 population [50]. Third, it is likely that un-published data and ongoing studies were missed. It is also possible that studies that found poor IGRA performance were less likely to be published. Because of the lack of statistical methods to account for publication bias in diagnostic meta-analyses, it would be prudent to assume some degree of over-estimation of our estimates resulting from publication bias. Fourth, our review did not include evidence on use of IGRA in 2 patient subgroups in which conventional tests for active tu-berculosis perform poorly: children and patients with suspected extrapulmonary tuberculosis. Lastly, we did not identify any studies directly measuring the impact of IGRA on patient-im-portant outcomes.

In conclusion, as in the case of the TST, the data suggest no role for using IGRAs for active tuberculosis diagnosis for adults living in low- and middle-income countries. These data should help inform evidence-based policies on the role of IGRA in active tuberculosis diagnosis in low- and middle-income...
settings. Indeed, a WHO Expert Group considering this evidence recently recommended that IGRAs should not be used as a replacement for conventional microbiological diagnosis of pulmonary and extrapulmonary tuberculosis in low- and middle-income countries [59].

Notes

Acknowledgments. We thank the authors of all studies included in the review for kindly responding to our requests for additional information; George Yen, for his help with translation; and UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, WHO Stop TB Department, and New Diagnostics Working Group, Stop TB Partnership, for supporting this work.

Financial Support. This work was supported in part by the National Institutes of Health (UCSF-CTSI KL2 RR024130 to J. Z. M., K23HL094141 to A. C., and K24 HL087713 to L. H.) and a New Investigator Award from the Canadian Institutes of Health Research (to M. P.).

Potential conflicts of interest. K. R. S. serves as Coordinator of the Evidence Synthesis subgroup of Stop TB Partnership’s New Diagnostics Working Group; M. P. serves as cochair of the Stop TB Partnership’s New Diagnostics Working Group and as consultant to the Bill & Melinda Gates Foundation. All other authors: no conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References


29. Lijmer JG, Bossuyt PM. Exploring sources of heterogeneity in systematic reviews of diagnostic accuracy included in systematic reviews. JAMA 2002; 288:1126–32.


60. WHO. Interferon-gamma release assays for Active Tuberculosis • JID 2011:204 (Suppl 4) • S1129