2011

Use of tuberculosis interferon-gamma release assays (IGRAs) in low- and middle-income countries

Policy Statement

World Health Organization

Contents

Executive summary

Background

Research over the past decade has resulted in the development of two commercial interferon-gamma release assays (IGRAs), based on the principle that the T-cells of individuals who have acquired TB infection respond to re-stimulation with *Mycobacterium tuberculosis*-specific antigens by secreting interferon gamma (IFN-γ). 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 (ELISAs) measuring the amount of IFN-γ 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, UK) measures the number of peripheral mononuclear cells that produce INF-γ after stimulation with ESAT-6 and CFP-10.

Commercial IGRAs are FDA-approved as indirect and adjunct tests for TB infection, in conjunction with risk assessment, radiography and other medical and diagnostic evaluations. 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 IGRA use in low- and middle-income countries - typically with high TB- and/or HIV-burden - yet IGRAs are being marketed and promoted, especially in the private sector.

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. 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 (WHO-STB) therefore commissioned systematic reviews on the use of IGRAs in low- and middle-income countries, 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 of the GRADE evidence synthesis process are summarised below.

This Policy Statement applies to the use of commercial IGRAs in low- and middle-income countries only. Several international guidelines on IGRA use in high-income countries are available. This Policy Statement is not intended to apply to high-income countries or to supersede their national guidelines.

Overall conclusions

- **There is insufficient data and low quality evidence on the performance of IGRAs in low- and middle-income countries, typically those with a high TB and/or HIV burden;**
- **IGRAs and the TST cannot accurately predict the risk of infected individuals developing active TB disease;**
- **Neither IGRAs nor the TST should be used for the diagnosis of active TB disease;**
- **IGRAs are more costly and technically complex to do than the TST. Given comparable performance but increased cost, replacing the TST by IGRAs as a public health intervention in resource-constrained settings is not recommended.**

Summary of study results in low- and middle-income countries

Use of IGRAs in diagnosis of active TB: IGRAs were explicitly designed to replace the tuberculin skin test (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).

The quality of evidence for use of IGRAS (and the TST) in diagnosis of active TB was low. There was no consistent evidence that IGRAs were 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 to the diagnosis of active TB beyond readily available patient data and conventional microbiological tests.

Policy recommendation: **IGRAs (and the TST) should not be used in low- and middle-income countries for the diagnosis of pulmonary or extra-pulmonary TB, nor for the diagnostic work-up of adults (including HIV-positive individuals) suspected of active TB in these settings (strong recommendation). This recommendation places a high value on avoiding the consequences of unnecessary treatment (high false-positives) given the low specificity of IGRAs (and the TST) in these settings.**

Use of IGRAs in children: Two small studies prospectively estimated the incidence of active TB in children who had been tested with IGRAs. The quality of evidence for use of IGRAS in children was very low and conflicting results were reported. When exposure was used as the reference standard for LTBI, 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.

Policy recommendation: **IGRAs should not replace the TST in low- and middle-income countries for the diagnosis of latent TB infection in children, nor for the diagnostic work-up of children (irrespective of HIV status) suspected of active TB in these settings (strong recommendation). It should also be noted that there may be additional harms associated with blood collection in children and that issues such as acceptability and cost had not been adequately addressed in any studies.**

Use of IGRAs in HIV-infected individuals: 37 studies were identified that included 5,736 HIV-infected Individuals; however, despite the multitude of studies the quality of evidence for use of IGRAS in individuals living with HIV infection was very low. In persons with active TB (used as a surrogate reference standard for LTBI), pooled sensitivity estimates were higher for T-SPOT (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 the TST in head-to-head comparisons and the impact of advanced 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 T-SPOT 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 T-SPOT and increased by 15% (95% CI -11, 41) with TST. All tests therefore seemed to be affected by CD4+ cell count.

Policy recommendation: **IGRAs should not replace the TST in low- and middle-income countries for the diagnosis of latent TB infection in individuals living with HIV infection (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 screening of HCWs for LTBI in low- and middle-income countries and the quality of evidence was very low. Two cross-sectional studies compared IGRA and TST performance in HCWs. 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 burden TB and/or HIV settings.

Policy recommendation: **IGRAs should not be used in health care worker screening programmes in low- and middle-income countries (strong recommendation)**.

Use of IGRAs in contact screening and outbreak investigations: 16 studies (14 original manuscripts and 2 unpublished studies) evaluated IGRAs in contact screening and outbreak investigations in low- and middle-income countries. The quality of evidence for use of IGRAs for LTBI screening in contact and outbreak investigations was very low. 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 four studies reported a statistically significant difference between positivity rates estimated by TST, T-SPOT 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 (adjusted odds ratio) varied across studies. Results indicated that concordance between TST and IGRAs ranged widely.

Policy recommendation: **IGRAs should not replace the TST in low- and middle-income countries for the screening of latent TB infection in adult and paediatric contacts, or in outbreak investigations (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 quality of evidence for the predictive value of IGRAS was very low. 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; 95% CI 0.62-5.85; I2=0%; p=0.90; TST: IRR=2.28; 95% CI 0.83-3.73); 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.

Policy recommendation: **Neither IGRAs nor the TST should be used in low- and middle-income countries for the identification of individuals at risk of developing active TB (strong recommendation)**.

Acknowledgements

Declarations of Interest

Individuals were selected to be members of the Expert Group to represent and balance important perspectives for the process of formulating recommendations. The Expert Group therefore included technical experts, end-users, patient representatives and evidence synthesis methodologists.

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.

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. DOI statements were summarised by the co-chair (Karin Weyer, WHO-STB) of the Expert Group meeting at the start of the meeting.

P Hill and R O'Brien declared conflicts of interest that were deemed to be insignificant: P Hill declared receipt of kits from Cellestis and Oxford Immunotec for research projects, and R O'Brien declared FIND support to academia to develop a point of care serodiagnostic test, including the FIND biomarker discovery project.

Selected individuals with intellectual and/or research involvement in the use of TB interferon-γ release assays (IGRAs) in low- and middle-income settings were invited as observers to provide technical input and answer technical questions. P Godfrey-Fausett declared a research grant for the investigation of the use of the QuantiFERON-TB Gold In-Tube assay in Zambia and South Africa, and M Pai declared conduct of research studies on IGRAs. 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 report, nor in preparation of the STAG-TB documentation or preparation of the final WHO Policy Statement.

The systematic reviewers (A Cattamanchi, A Date, A Detjen, D Dowdy, R Menzies, J Metcalfe, M Pai, M Rangaka, K Steingart and A Zwerling) were deemed to have a conflict of interest and consequently were observers to the meeting, providing technical clarifications on the findings of the systematic reviews. They did not participate in the GRADE evaluation process, did not contribute to the meeting discussions where recommendations were developed, and did not provide comments on the final WHO Policy Statement.

USE OF TUBERCULOSIS INTERFERON-GAMMA RELEASE ASSAYS (IGRAs) IN LOW- AND MIDDLE-INCOME COUNTRIES

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 detection. 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 *M. 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). ESAT-6 and CFP-10 are strong targets of the cellular immune response in patients with *M. tuberculosis* infection. In such persons, sensitized memory/effector T cells produce interferon-gamma (IFN-γ) 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-γ produced in response to specific *M. tuberculosis* antigens (QFT-G: ESAT-6 and CFP-10, QFT-GIT: ESAT-6, 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-γ 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 with the BCG vaccine. Commercial IGRAs are FDA-approved as indirect and adjunct tests for TB infection, in conjunction with risk assessment, radiography and other medical and diagnostic evaluations. 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 low- and middle-income countries (typically characterised by high TB- and/or HIV-burden), where IGRAs 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 therefore commissioned systematic reviews on the use of IGRAs in low- and middle-income 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 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 detecting 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 (e.g. 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 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 investigation: 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 (ie. 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 IGRAs 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. Methods

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) evaluate the risks and benefits of using IGRAs in low- and middle-income countries; and c) identify gaps to be addressed in future research. Based on the Expert Group findings, the third and final step involved WHO policy guidance on the use of these tests, presented to the WHO Strategic and Technical Advisory Group for TB (STAG-TB) for consideration, with 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 and was co-chaired by WHO-STB and a clinical epidemiologist with expertise and extensive experience in evidence synthesis and guideline development.

To comply with current standards for evidence assessment in formulation of policy recommendations, the GRADE system (www.gradeworkinggroup.org), adopted by WHO for all policy and guidelines development, was used.

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. Recognising that test results may be surrogates for patient-important outcomes, the Expert Group evaluated IGRA accuracy while also drawing inferences on the likely impact of these tests on patient outcomes, as reflected by false-negatives (ie. cases of LTBI missed) or false-positives.

Systematic review and meta-analyses

Systematic reviews were done following detailed protocols with predefined questions relevant to the individual topics. Summaries of methodologies followed for each topic are given in the respective sections below. Detailed methodology is described in the Expert Group Meeting Report available at www.who.int/tb/laboratory/policy_statements/en/index.html.

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 to evaluate the role of IGRAs depending on the individual topic (ie. 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).

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 publications was requested from the primary authors of all studies included by the systematic reviewers.

Study selection: Studies that evaluated the performance of currently available commercial IGRAs, published in all languages and in all low- and middle-income countries, 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.

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- and 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.

2.2   Decision-making during the Expert Group meeting and external review

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

The Expert Group meeting was co-chaired by the WHO-STB secretariat and an evidence synthesis expert. Decisions were based on consensus. Concerns and opinions by Expert Group members were noted and included in the final meeting report. The detailed meeting report was prepared by the WHO-STB secretariat and underwent several iterations (managed by the secretariat) before being signed off by all Expert Group members.

Recommendations from the Expert Group meeting were presented to WHO STAG-TB. STAG-TB endorsed the recommendations and requested WHO to proceed with the development of final policy guidance. This was circulated to the Expert Group and STAG-TB members and comments incorporated as relevant.

The final policy guidance document was approved by the WHO Guidelines Review Committee (GRC), having satisfied the GRC requirements for guideline development.[i]

2.3. Scope of the policy guidance

This document provides a pragmatic summary of the evidence and recommendations related to the use of IGRAs in low- and middle-income countries and should be read in conjunction with the detailed findings from the Expert Group Meeting Report.

This policy guidance should be used to inform the use of IGRAs in low- and middle-income countries. It is intended for National TB Programme Managers and Laboratory Directors, external laboratory consultants, donor agencies, technical advisors, laboratory technicians, laboratory equipment procurement officers, and private sector service providers. Individuals responsible for programme planning, budgeting, resource mobilization, and training activities for TB diagnostic services may also benefit from using this document.

Date of review: 2016

---

[i]GRC statement: This guideline was developed in compliance with the process for evidence gathering, assessment and formulation of recommendations, as outlined in the WHO Handbook for Guideline Development (current version).

3. Evidence base for policy formulation

3.1 Use of IGRAs in diagnosis of active TB

3.1.1 Study characteristics

Studies included were those that evaluated the performance of the most recent generation of commercial, RD1 antigen based IGRAs (QuantiFERON-TB Gold In-Tube (QFT-GIT) [Celestis, Victoria, Australia] and T-SPOT [Oxford Immunotec, Oxford, United Kingdom]) among adult (>15 years) active pulmonary TB suspects or cases in low- and middle-income countries.

Studies excluded were those that evaluated non-commercial IGRAs, PPD-based IGRAs, QuantiFERON-TB Gold (2G), IGRAs performed in specimens other than blood; those reporting longitudinal data focused on the effect of anti-TB treatment on IGRA response; studies including <10 eligible individuals; studies focused on extrapulmonary tuberculosis in children; studies reporting insufficient data to determine diagnostic accuracy measures; and conference abstracts, letters without original data and reviews.

The initial search yielded 789 citations. After full-text review of 185 papers evaluating IGRAs for the diagnosis of active TB, 22 were determined to meet eligibility criteria, covering 33 unique evaluations of one or more IGRAs (hereafter referred to as studies) in 19 published and 3 unpublished reports. Of the 33 studies, 10 (30%) were from low-income countries, and 23 (70%) were from middle-income countries. Seventeen studies (52%) included HIV-infected individuals (n=1,057), and 27 (82%) studies involved ambulatory subjects (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. Because of the focus on diagnostic accuracy for active TB and the high prevalence of LTBI in high TB-burden settings, IGRA specificity was estimated exclusively among studies enrolling TB suspects where the diagnostic workup ultimately showed no evidence of active disease.

3.1.2 Summary of results

The results demonstrated that 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 could 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 for the diagnosis of active TB. Among studies that enrolled TB suspects (ie. patients with diagnostic uncertainty), both IGRAs demonstrated suboptimal 'rule-out' values for active TB;

- Even 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 would 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 would be IGRA-positive, with adverse 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 (eg. demographics, symptoms, or chest radiograph findings), 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, data for extra-pulmonary TB being non-existent; nevertheless, consensus by the Expert Group was that recommendations for pulmonary TB could reasonably be extrapolated to extra-pulmonary TB;

- Industry involvement was unknown in 18% studies and acknowledged in 27% studies, including donation of IGRA kits as well as work/financial relationships between authors and IGRA manufacturers.

3.1.3   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 population.

It is possible 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) and indirect assessment of patient impact (false-positive and false-negative results). None of the studies reviewed provided information on patient-important outcomes, ie. 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.

3.1.4 Grade evidence profiles and final policy recommendations

The GRADE evidence profiles are provided in Tables 1 and 2. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRAs and the TST in diagnosis of active TB was low and recommended that these tests should not be used in low- and middle-income countries as a replacement for conventional microbiological diagnosis of pulmonary and extra-pulmonary TB (strong recommendation).

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

The systematic review results have subsequently been published.[1]

Policy recommendation: **IGRAs (and the TST) should not be used in low- and middle-income countries for the diagnosis of pulmonary or extra-pulmonary TB, nor for the diagnostic work-up of adults (including HIV-positive individuals) suspected of active TB in these settings (strong recommendation). This recommendation places a high value on avoiding the consequences of unnecessary treatment (high false-positives) given the low specificity of IGRAs (and the TST) in these settings.**

3.2 Use of IGRAs in children

3.2.1 Study characteristics

The initial search yielded 234 citations. After full-text review of 68 papers evaluating IGRAs in children, 32 were determined to meet eligibility criteria, covering 33 unique evaluations of one or more IGRAs (hereafter referred to as studies) in 18 countries. Of the 33 studies, three were from low-income countries, and 11 were from middle-income countries. The incidence of smear-positive TB was <25/100,000 in 18 of these countries and >=25/100,000 in the remaining countries. Studies performed in high-income countries included between 11% and 100% immigrant children from countries with higher burdens of TB.

All studies included in the review assessed either or both commercial IGRAs, QuantiFERON (QFT, in its Gold and In-Tube version) and T-SPOT.TB (T-SPOT) as well as the TST in children. Very few studies clearly reported on the sampling methods (consecutive, random or convenience) and representativeness of the patient spectrum. Blinding of clinicians to IGRA results were absent for most studies. Wide variation was evident on the criteria used for the definition of the reference standard (active TB).

Among studies in low- and middle-income countries analysing the test performance for latent TB infection, 4 studies used "exposed" and "unexposed" as comparison groups and 5 studies allowed analysis of the correlation between different grades of exposure and test results. Six studies from low- and middle-income countries were included in the analysis of test performance in TB disease, with varying definitions for each group of TB suspects/patients and for the "no TB" categories.

3.2.2 Summary of results

The majority of IGRA studies in children had been performed in high-income countries and extrapolation to low- and middle-income settings with high background TB infection rates was not appropriate. However, based on available data, the results indicated that in low- and middle-income countries:

- IGRAs and the TST had very similar accuracy for diagnosis of LTBI and active TB in children;

- Major methodological inconsistencies between studies had a negative effect on the comparability of studies and results. A key constraint was 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 IGRAs over TST in detecting LTBI in exposed or unexposed individuals or in a gradient of exposure was not detected;

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

- Both IGRAs and TST showed lower sensitivity in HIV-infected children in one study assessed;

- Overall, the ability of TST and IGRAs 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 IGRAs. Cost was noted as an important and limiting factor. Aspects inherent to the use of IGRAs in children, such as the difficulty of phlebotomy and the amount of blood needed in young children, are relevant implementation considerations.

- A third of studies were supported by manufacturers of IGRAs, mainly through donation of test kits.

3.2.3 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 IGRAs 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 possible 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 LTBI. None of the studies reviewed provided information on patient-important outcomes, ie. showing that IGRAs or the TST 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.

3.2.4   Grade evidence profiles and final policy recommendations

The GRADE evidence profiles are provided in Tables 3 to 6. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRAS in children was very low and recommended that these tests should not be used in low- and middle-income countries as an alternative to TST in paediatric TB for the diagnosis of latent TB infection, nor 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 noted that there may be additional harms associated with blood collection in children and that issues such as acceptability and cost had not been adequately addressed in any studies.

The systematic review results have subsequently been published.[2]

Policy recommendation: **IGRAs should not replace the TST in low- and middle-income countries for the diagnosis of latent TB infection in children, nor for the diagnostic work-up of children (irrespective of HIV status) suspected of active TB in these settings (strong recommendation). It should also be noted that there may be additional harms associated with blood collection in children and that issues such as acceptability and cost had not been adequately addressed in any studies.**

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

3.3.1 Study characteristics

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, covering 37 unique evaluations (hereafter referred to as studies). Of these, 22 studies were conducted in low- and middle-income countries.

There was a high degree of variation in study design and study populations. 15/22 (68%) of studies included only ambulatory HIV-positive individuals. IGRAs were performed in persons with or suspected of having active TB in 12 studies, 6 studies evaluated asymptomatic HIV-positive persons for LTBI, and 4 studies considered both asymptomatic as well as symptomatic individuals with HIV co-infection.

3.3.2 Summary of results

Results indicated that in low- and middle-income countries:

- The optimal test for identifying HIV-infected persons who could benefit from IPT remains an unanswered question 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 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 and there was no data to show that individuals with TST-negative/IGRA-positive results had improved outcomes on IPT. Data on the impact of immunosuppression on IGRA validity remains unclear;

- Seven (32%) studies reported industry involvement, including donation of IGRA test kits and work/financial relationships between IGRA manufacturers and principal authors.

3.3.3   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.

3.3.4   Grade evidence profiles and final policy recommendations

The GRADE evidence profiles are provided in Tables 7 and 8. 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 in low- and middle-income countries as a replacement for TST for the assessment of LTBI (strong recommendation).

The systematic review results have subsequently been published.[3]

Policy recommendation: **IGRAs should not replace the TST in low- and middle-income countries for the diagnosis of latent TB infection in individuals living with HIV infection (strong recommendation). This recommendation also applies to HIV-positive children based on the generalisation of data from adults.**

3.4   Use of IGRAs for screening of health care workers

3.4.1   Study characteristics

The initial search yielded 546 citations. After full-text review of 56 papers evaluating commercial IGRAs in health care workers (HCWs), 48 were deemed to have met the eligibility criteria. Of these, only five (12%) were done in low- and middle-income settings.

Studies varied greatly in design, execution, and reported outcomes. IGRA performance varied greatly across populations; therefore, results were also stratified by TB incidence (>100 estimated incident TB cases/ 100,000 population; <= 100/100,000 as reported to WHO) in the countries where the studies were done. Due to the variety of study designs and HCW screening guidelines, study populations included HCWs with widely differing risks of TB exposure.

3.4.2   Summary of results

Results indicated that in low- and middle-income countries:

- Prevalence of LTBI in HCWs depended on the test used and the particular TB incidence setting. Two cross-sectional studies comparing IGRA and TST positivity rates in HCWs showed high TST positivity rates (40% to 66%) and slightly lower rate for IGRA positivity (statistically significant in only one study, which also showed the lowest rate of BCG vaccination among participants);

- Both the TST and IGRAs appeared to be associated with markers of TB exposure, but the magnitude of associations varied greatly; TST performance was adversely affected by BCG vaccination while IGRA performance seemed to be unaffected;

- Both IGRAs and the TST had suboptimal sensitivity and discordant results were common. IFN-γ g responses seemed to have natural variation and tended to fluctuate around the cut-off, causing apparent IGRA conversions and reversions. The exact cause of the conversions and reversions remained 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 was complicated by lack of data on optimum cut-offs for serial testing, and unclear interpretation and prognosis of conversions and reversions;

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

- There were no data to show that IGRAs performed better at identifying incidence of new TB infections among HCWs than the TST, irrespective of HIV status.

3.4.3   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 were seriously limited.

3.4.4 Grade evidence profiles and final policy recommendations

The GRADE evidence profiles are provided in Tables 9 and 10. 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 in these countries (strong recommendation). The Expert Group also noted the lack of WHO policy on using the TST in health care worker screening programmes.

The systematic review results have subsequently been published.[4]

Policy recommendation: **IGRAs should not be used in health care worker screening programmes in low- and middle-income countries (strong recommendation)**.

3.5 Use of IGRAs in contact screening and outbreak investigations

3.5.1 Study characteristics

The initial search yielded 608 citations. After full-text review of 99 papers evaluating commercial IGRAs in screening of contacts and outbreak investigations, 65 studies conducted in high-income countries were excluded, as were 18 studies using pre-commercial and in-house IGRAs. 16 studies were deemed to have met the eligibility criteria.

Most studies were small (39-301 participants); however, the inclusion of one unpublished study doubled the total sample size (2,211 study participants). All studies included BCG vaccinated participants. HIV status was frequently unreported, but when it was documented, rates were low (0-1.5%) with the exception of the large unpublished study where the reported HIV infection rate was around 38% in the adult study population, and one study reporting an HIV infection rate of 5% in the paediatric study population.

Only one study did not include household contacts but evaluated HCWs exposed to a smear-positive TB case. The remaining 15 studies all included household contacts, while three studies also included school or work contacts. Nine (56%) of the studies exclusively examined child contacts, three studies included both child and adult contacts, and four studies exclusively included adult contacts. Most studies involved only contacts of confirmed active TB cases; however, five studies recruited a comparison group with no known TB exposure.

Studies varied in quality, with several quality indicators frequently unreported. For example, only three of 14 studies reported that study personnel were blinded to other test results or TB exposure when performing and interpreting test results.

3.5.2 Summary of results

Results indicated that in low- and middle-income countries:

- The prevalence of positive tests varied greatly between studies and across assays. Prevalence of positive TST results ranged from 22% in children less than 5yrs to 84% in adult HCWs exposed to a smear-positive TB case. Prevalence of positive IGRA results ranged from 10% to 75% respectively. The majority of studies showed comparable LTBI prevalence by TST or IGRA in contacts;

- 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 (effect) varied across studies;

- IGRAs appeared to be dynamic assays with frequent conversions and reversions;

- Both IGRAs and TST seemed to have similar and modest predictive value.

- Five of 15 studies reported industry involvement, most frequently the donation of IGRA test kits. One study reported one of its authors having been a paid consultant of the manufacturer of the IGRA assay evaluated.

3.5.3   Strengths and limitations of the evidence base

Due to significant 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 IGRAs. Studies that assessed associations between exposure and test positivity used different categorisation of exposure variables, making it difficult to compare results across studies.

3.5.4   Grade evidence profiles and final policy recommendations

The GRADE evidence profiles are provided in Table 11. Based on these assessments, the Expert Group concluded that the quality of evidence for use of IGRAS for LTBI screening in contact and outbreak investigations was very low and recommended that these tests should not be used in low- and middle-income countries as a replacement for TST, neither in adults nor children investigated as close contacts of patients with confirmed active TB (strong recommendation).

Policy recommendation: **IGRAs should not replace the TST in low- and middle-income countries for the screening of latent TB infection in adult and paediatric contacts, or in outbreak investigations (strong recommendation)**.

3.6 The predictive value of IGRAs for incident active TB

3.6.1 Study characteristics

The initial search yielded 722 citations. After full-text review of 14 papers evaluating the predictive value of commercial IGRAs for active TB, 8 studies conducted in high-income countries were excluded, as were three studies using in-house IGRAs.

Three studies were deemed to have met the eligibility criteria. The at-risk populations included in the three studies were all different (older males with confirmed silicosis, school-going adolescents, and adult TB contacts including HIV-infected individuals). 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), leading to possible verification bias. One study incorporated IGRA results in their reference standard for TB, leading to incorporation bias.

3.6.2 Summary of results

Results indicated that in low- and middle-income countries:

* The vast majority of individuals (>95%) with a positive IGRA results did not progress to active TB disease during follow-up, although a modest but statistically insignificant increase in incidence rates of TB in IGRA- positives compared to IGRA-negatives had been observed;

* IGRA sensitivity for incident TB ranged from 75% to 88% (95% CI 46% - 99% depending on the country/study population), while IGRA specificity ranged from 35% to 51% (95% CI 30% - 54% depending on the country/study population). TST sensitivity for incident TB was similar, ranging from 73% to 76% (95% CI 50% to 93% depending on the country/study population). Specificity was equally low, ranging from 35% to 58% (95% CI 29% - 58% depending on the country/study population). One study reported lower TST sensitivity and higher specificity but acknowledged that logistical issues at the clinical sites could have affected the TST results;

* Both IGRAs and the TST appeared to have only modest predictive value and did not help to identify those who are at highest risk of progression to TB disease. Patient relevant outcomes based on sensitivity and specificity appeared comparable between the two tests.

3.6.3 Grade evidence profiles and final policy recommendations

The GRADE evidence profiles are provided in Table 12 and 13. Based on these assessments, the Expert Group concluded that the quality of evidence for the predictive value of IGRAs was very low and recommended that these assays should not be used in low- and middle-income countries to identify individuals at risk of active TB disease(strong recommendation).

The systematic review results have subsequently been published.[5]

Policy recommendation: **Neither IGRAs nor the TST should be used in low- and middle-income countries for the identification of individuals at risk of developing active TB (strong recommendation).**

4. Operational aspects on the use of IGRAs

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

- *Cost*
  Cost of IGRAs was mentioned by four studies, 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.

5. Overall conclusions

- There is insufficient data and low quality evidence on the performance of IGRAs in low- and middle-income countries, typically those with a high TB and/or HIV burden;

- IGRAs and the TST cannot accurately predict the risk of infected individuals developing active TB disease;

- Neither IGRAs nor the TST should be used for the diagnosis of active TB disease;

- IGRAs are more costly and technically complex to do than the TST. Given comparable performance but increased cost, replacing the TST by IGRAs as a public health intervention in resource-constrained settings is not recommended.

6. Implications for further research

Targeted further research to identify IGRAs with improved accuracy is strongly encouraged. Such research should be based on adequate study design and including quality principles such as representative suspect populations, prospective follow-up and adequate, explicit blinding. It is also strongly recommended that proof-of-principle studies be followed by evidence produced from prospectively implemented and well designed evaluation and demonstration studies, including assessment of patient impact.

7. GRADE tables

8.   Selected references

1. Metcalfe J, Everett C, Steingart K, Cattamanchi A, Huang L, Hopewell P, Pai M. Interferon – gamma release assays for active pulmonary TB diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. JID 2011; 204(suppl 4), November 15.

2. Mandalakas A, Detjen A, Hessling A, Bendetti A, Menzies D. Interferon-gamma release assays and childhood tuberculosis: systematic review and meta-analysis. Int J Tuberc Lung Dis 2011; 15(8): 1018-1032.

3. Cattamanchi A, Smith R, Steingart K, Metcalfe J, Date A, Coleman C. Marston B, Huang L, Hopewell P, Pai M. Interferon–gamma release assays for the diagnosis of latent tuberculosis in HIV-infected individuals: A systematic review and meta-analysis. J Acquir Immune Defic Syndr 2011; 56:230-238.

4. Zwerling A, van den Hof S, Scholten J, Cobelens F, Menzies D, Pai M. Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review. Thorax 2011; published online on 12 January, doi:10.1136/thx.2010.143180.

5. Rangaka M, Wilkinson K, Ling D, Menzies D, Mwansa-Kambafwile J, Fielding K, Wilkinson R, Pai M. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. Lancet 2011; published online on 16 August 2011, doi: 10.1016/S1473-3099(11)70210-9.

Table 1. 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

No of Participants (studies)	Study design	Limitations	Indirectness	Inconsistency	Imprecision	Publication Bias	Quality of evidence (GRADE)	Importance
A. Outcome: Diagnostic accuracy								
True Positives								
2067 (19) [A1]	Cross-sectional	No Serious Limitation [A2]	No Serious Indirectness [A3]	Serious [A4] (-1)	Serious [A5] (-1)	Likely [A6]	Low ⊕⊕◯◯	Critical (7-9)
True Negatives								
2067 (19) [A1]	Cross-sectional	No Serious Limitation [A2]	No Serious Indirectness [A3]	Serious [A4] (-1)	Serious [A5] (-1)	Likely [A6]	Low ⊕⊕◯◯	Critical (7-9)
False Positives								
2067 (19) [A1]	Cross-sectional	No Serious Limitation [A2]	No Serious Indirectness [A3]	Serious [A4] (-1)	Serious [A5] (-1)	Likely [A6]	Low ⊕⊕◯◯	Critical (7-9)
False Negatives								
2067 (19) [A1]	Cross-sectional	No Serious Limitation [A2]	No Serious Indirectness [A3]	Serious [A4] (-1)	Serious [A5] (-1)	Likely [A6]	Low ⊕⊕◯◯	Critical (7-9)
B. Outcome: Proportion indeterminate tests								
2872 (33) [B1]	Cross-sectional	Serious [B2] (-1)	No Serious Indirectness [B3]	Serious [B4] (-1)	No Serious Imprecision [B5]	Likely [B6]	Low ⊕⊕◯◯	Critical (7-9)
C. Outcome: Incremental value								
943 (2) [C1]	Cohort	Serious [C2] (-1)	No Serious Indirectness [C3]	No Serious Inconsistency [C4]	Serious [C5] (-1)	Unlikely [C6]	Low ⊕⊕◯◯	Critical (7-9)

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: 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.

[A1] 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 enroll 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 enroll 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%, I^2 78%, p<0.001), and HIV-infected subjects evaluated with both QFT-GIT (range 3-40%, I^2 72%, p<0.001) and T-SPOT (range 0-25%, I^2 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 enroll a representative spectrum of patients. Model specification was undertaken for both studies using traditional parametric statistical methods.

[C3] See footnote [A3]. 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 2.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

Outcomes: TP, TN, FP, FN Subgroups	Effect % (95% CI)	No. of participants (studies)	What do these results mean given 10% prevalence among suspects being screened for TB?	What do these results mean given 30% prevalence among suspects being screened for TB?	Quality of Evidence
T-SPOT.TB; HIV-infected	Sensitivity 78% (56, 91) Specificity 55% (45, 64)	549 (5)	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.	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.	Low ⊕⊕◯◯
T-SPOT.TB; HIV-uninfected	Insufficient data for pooled estimates	364 (3)	--	--	--
QuantiFERON-TB Gold In-Tube; HIV-infected	Sensitivity 62% (41,79) Specificity 51% (39, 64)	469 (6)	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 the 900 patients without TB, 459 (TN) will not be treated; 441 (FP) will be unnecessarily treated.	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 700 patients without TB, 357 (TN) will not be treated; 343 (FP) will be unnecessarily treated.	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.	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.	Low ⊕⊕◯◯
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 countires.	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 proportion of indeterminate TSPOT test results in low- and middle-income countries will be low and similar to high-income countries.	Low ⊕⊕◯◯
	QuantiFERON-TB Gold	16% (10-21%)	728 (10)	In low- and middle-income	Low

	In-Tube, HIV-infected Subjects			countries, the proportion of indeterminate QFT-GIT results among HIV-infected subjects can be expected to be high - in about 16% of the patients tested, clinicians will not be able to use the QFT results for decision making.	⊕⊕◯◯
	T-SPOT.TB, HIV-infected Subjects	8% (1-15%)	666 (7)	In low- and middle-income countries, the proportion of indeterminate TSPOT results among HIV-infected subjects can be expected to be high - in about 8% of patients tested, clinicians will not be able to use the TSPOT results for decision making.	Low ⊕⊕◯◯
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).		943 (2)	This evidence suggests that after consideration of readily available patient data, neither commercial IGRA can be expected to be useful in diagnosing active pulmonary TB in patients living in low-and middle-income countries.	Low ⊕⊕◯◯

* Value is IGRA minus TST.

Table 3. GRADE Evidence Profile: The performance of IGRAs for the diagnosis of latent tuberculosis infection in children in low- and middle-income countries

No of participants (studies)	Study design	Limitations	Indirectness	Inconsistency	Imprecision	Publication bias	Quality of evidence (GRADE)[1]	Importance
A. Risk of progression to active TB								
No studies in LMIC								Critical (7-9)
B. Outcome: Performance of IGRAs in studies using a dichotomous measure of exposure as reference standard for LTBI (exposed/unexposed)								
229 (4)[B1]	Mainly cross-sectional	Not serious[B2]	Not serious[B3]	Serious[B4] (-1)	Very serious[B5] (-2)	Likely[2]	Very Low ⊕○○○	Critical (7-9)
C. Outcome: Performance of IGRAs in studies assessing different gradients of TB exposure as reference standard for LTBI								
1057 (5)[C1]	Cross sectional	Not serious[C2]	Not serious[C3]	Serious[C4]	Very serious[C5]	Likely[2]	Very Low ⊕○○○	Critical (7-9)

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 T-SPOT, 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 4. GRADE Summary of Findings – IGRAs for the diagnosis of latent tuberculosis infection in children in low- and middle-income countries

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)

Outcome	No. Participants	Principal Findings	What do these findings mean?	Quality of Evidence	Importance
Predictive value for active TB		No studies in LMIC			Critical (7-9)
Performance of IGRAs against dichotomous measure of exposure	QFT-G/QFT-GIT: 59 (3) T-SPOT: 170 (4) TST (10mm): 159 (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 $\oplus\bigcirc\bigcirc\bigcirc$	Critical (7-9)
Performance of IGRAs against exposure gradient	QFT-G/QFT-GIT: 773 (5) T-SPOT: 225 (1) TST (10 mm) 871 (5)	1. Pooled correlation between test and exposure gradient: • QFT-G/QFT-GIT: 0.28 (95%CI 0.06-0.86, I^2 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, I^2 0.65) 2. Regression slopes • QFT-G/QFT-GIT: 1.84 (95%CI 1.38-2.44, I^2 0.66) • T-SPOT: 1.63 (95%CI 1.12-2.39) • TST (10 mm): 1.73 (95% CI 1.36-2.20, I^2 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 $\oplus\bigcirc\bigcirc\bigcirc$	Critical (7-9)

Table 5. GRADE Evidence Profile: The diagnostic accuracy of IGRAs for the diagnosis of active tuberculosis in children in low- and middle-income countries

No of Participants (Studies)	Study design	Limitations	Indirectness	Inconsistency	Imprecision	Publication Bias	Quality of Evidence (GRADE)[1]	Importance
A: What is the sensitivity of IGRAs in children with active TB?								
207 (6)[A1]	Mainly cross-sectional	Serious[A2] (-1)	Not serious[A3]	Not serious[A4]	Very serious[A5] (-2)	Likely[2]	Very Low ⊕◯◯◯	Critical (7-9)
B: What is the specificity of IGRAs in children without TB?								
519 (4)[B1]	Mainly cross-sectional	Serious[B2] (-1)	Not serious[B3]	Serious[B4]	Serious[B5] (-1)	Likely[2]	Very Low ⊕◯◯◯	Critical (7-9)
C What is the proportion of indeterminate IGRA results among children assessed for active TB?								
656 (5)[C1]	Mainly cross sectional	Serious[C2] (-1)	Not serious[C3]	Not serious[C4]	Serious[C5]	Likely[2]	Very Low ⊕◯◯◯	Important (4-6)
D: What is the diagnostic accuracy of IGRAs in HIV-infected children?								
36 (1)[D1]	Cross-sectional	Serious[D2] (1)	Not serious[D3]	Not applicable[D4]	Very serious[D5]	Likely[2]	Very Low ⊕◯◯◯	Important (4-6)
E: What is the diagnostic accuracy of IGRAs in children < 5 years?								
471 (2)[E1]	Cross-sectional	Serious[E2] (-1)	Serious[E3] Not serious	Not applicable[E4]	Very serious[E5]	Likely[2]	Very Low ⊕◯◯◯	Important (4-6)

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.

[A1] 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.

[A2] 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)

[A3] 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).

[A4] 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-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.

[A5] 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)

[B1] 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.

[B2] 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)

[B3] 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)

[B4] 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).

[B5] 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)

[C1] 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.

[C2] 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)

C3 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 generalisable 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)

C4 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)

C5 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)

D1 One study assessed QuantiFERON-TB Gold in 36 HIV-infected children with active TB in Romania (an upper middle-income country.

D2 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)

D3 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 generalisable to low-income countries. Sensitivity of IGRAs is only a surrogate for patient-important outcomes. False negative results, particularly in HIV-infected children, may results 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)

D4 Only one study – inconsistency therefore cannot be assessed.

D5 The 95% CI for sensitivity of QFT-G in 36 HIV-infected children was very wide (63%, 95%CI 16-100). (2 points subtracted)

E1 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).

E2 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)

E3 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)

E4 Heterogeneity could not be assessed since each test was only assessed in one study.

E5 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 6. GRADE Summary of Findings – IGRAs for the diagnosis of active tuberculosis in children in low- and middle-income countries

Review question: What is the diagnostic accuracy of IGRAs for the diagnosis of active TB in children in LMIC?
Patients/population: TB suspects or active TB patients and control group with 'no TB' in low and middle income countries
Setting: Mainly mixed, in- and outpatients
Index test: QuantiFERON-TB Gold [QFT-G], QuantiFERON-TB Gold In-Tube [QFT-GIT], and T-SPOT.*TB* [T-SPOT].
Importance: 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.
Reference standard: Culture confirmed TB and probable TB versus 'no TB'
Studies: Cross-sectional or case-control studies

Outcome	Index test	No. of Participants (Studies)	Effect % (95% CI) Main findings	What do these results mean given a 10% prevalence among suspects being screened for TB?	What do these results mean given a 30% prevalence among suspects being screened for TB?	Quality of Evidence
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 QFT-GIT 90% (83-95)	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 ⊕○○○

	TST	Sensitivity 168 (5)	Pooled sensitivity 65% (31-99)	IGRAs do not perform significantly different from TST	
		Specificity 490 (3)	Pooled specificity 90% (82-98)		
What is the proportion of indeterminate IGRAs among children assessed for active TB?	T-SPOT	132 (2)	Indeterminates/total number of tests 9/132 = 6.82 % Range of % indeterminates across studies 0-8%	**What do these results mean?** On average, indeterminate IGRA results are below 10% but can be high in certain populations, such as in one study performed in 100% HIV-infected children with active TB, showing 25% indeterminates.	Very Low ⊕○○○
	QFT-G/ QFT-GIT	QFT-G 36 (1) QFT-GIT 488 (2)	Indeterminates/total number of tests QFT-G: 9/36 = 25% QFT-GIT: 20/488=4.1% (Range of % indeterminates across studies 0-5%)		

Performance of IGRAs in HIV-infected children	T-SPOT	*No studies*	*No studies*			Very Low
	QFT-G	Sensitivity 36 (1) Specificity *No studies*	<u>Sensitivity</u> QFT-G: 63% (16-100) • 47% (0-100) if indeterminates counted as FN <u>Pooled specificity</u> *No studies*	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.	With a prevalence of 30%, 300/1000 will have TB. Of these, 189 will be correctly identified with QFT-G, 111 will be missed.	⊕○○○
	TST	Sensitivity 36 (1) Specificity *No studies*	<u>Sensitivity</u> 39% (0-100) <u>Specificity</u> *No studies*	Sensitivity of TST is lower than of QFT-G, but confidence intervals are wide and overlap.		
Performance in children <5yrs	T-SPOT	Sensitivity 134 (3) Specificity 97 (2)	<u>Pooled sensitivity</u> 77% (23-100) <u>Pooled specificity</u> 93% (83-100)	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.	With a prevalence of 30%, 300/1000 will have TB. Of these, 231 will be correctly identified with T-SPOT, 69 will be missed. Of 700 children with out TB, 651 will not be treated, and 49 will be unnecessarily treated.	Very Low ⊕○○○

		QFT-GIT	Sensitivity 36 (1)	Pooled sensitivity 35% (30-40)	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.	With a prevalence of 30%, 300/1000 will have TB. Of these, 105 will be correctly identified by QFT, 195 will be missed. Of 700 children without TB, 595 will not be treated, 105 will be unnecessarily treated.	
			Specificity 327 (1)	Pooled specificity 85% (81-89)			
		TST	Sensitivity 99 (2)	Pooled sensitivity 41% (0-85)	Sensitivity and specificity of T-SPOT are higher than of QFT-GIT or TST, but the difference is not significant (overlapping confidence intervals)		
			Specificity 395 (2)	Pooled specificity 83% (81-86)			

Table 7. GRADE Evidence Profile: The role of IGRAs in the diagnosis of latent tuberculosis infection in HIV-infected individuals in low- and middle-income countries

No of Participants (Studies)	Study design	Limitations	Indirectness	Inconsistency	Imprecision	Publication bias	Quality of evidence (GRADE)[1]	Importance
A. Outcome: Predictive value of IGRAs for active TB								
1100 (3)[B1] LMIC: 306 (1)	Prospective cohort	Serious[B2] (-1)	Serious[B3] (-1)	None[B4]	Serious[B5] (-1)	Likely[B6]	Very Low ⊕○○○	Critical (7-9)
B. Outcome: Sensitivity for active TB (as a surrogate reference standard for LTBI)								
1523 (18)[D1] LMIC: 1056 (16)	Mainly cross-sectional	No serious limitations[D2]	Serious[D3] (-1)	Very Serious[D4] (-2)	Serious[D5] (-1)	Likely[D6]	Very Low ⊕○○○	Important (4-6)
C. Outcome: Concordance with TST								
2158 (15)[E1] LMIC: 401 (5)	Cross-sectional	No serious limitations[E2]	Very Serious[E3] (-2)	Serious[E4] (-1)	None[E5] (-1)	Likely[E6]	Very Low ⊕○○○	Important (4-6)

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: 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.

[B1] 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).

[B2] 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.

[B3] Two studies were carried out in high income countries; hence the findings may not be generalizable to low/middle income countries.

[B4] 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.

[B5] 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.

[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. Some degree of publication bias was assumed 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; 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.

[D1] 18 studies were identified: 9 evaluated TSPOT and 9 evaluated QFT-GIT.
[D2] 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.
[D3] 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.
[D4] 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.
[D5] The 95% confidence interval for pooled sensitivity was wide for both TSPOT (72%, 95% CI 62-81%) and QFT-GIT (61%, 47-75%) in low/middle income countries.
[D6] 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)..

[E1] 15 studies were identified: 9 evaluated TSPOT and 6 evaluated QFT-GIT.
[E2] Study limitations were evaluated using the QUADAS scale. A majority of studies satisfied all QUADAS criteria assessed.
[E3] 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 IGRAs and TST is a poor surrogate for patient-important outcomes.
[E4] 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%).
[E5] The 95% confidence interval for pooled concordance was within +/10% in most sub-groups.
[E6] 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 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; 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 8.GRADE Summary of Findings – Role of IGRAs in the diagnosis of latent tuberculosis infection in HIV-infected individuals in low- and middle-income countries

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)

Outcome	N	Principal Findings	What do these findings mean?	Quality of Evidence	Importance
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)

Table 9.   GRADE Evidence Profile: IGRAs for tuberculosis screening of healthcare workers in low- and middle-income countries

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

216 (1) [C1]	Longitudinal	No serious limitations [C2]	Serious [C3] (-1)	No serious inconsistency [C4]	Very Serious [C5] (-2)	Likely [C6]	Very Low ⊕○○○	Important (4-6)

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] 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 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.

[B1] 4 cross-sectional studies were identified: 3 evaluated a previous version of the QFT, 1 study evaluated only the T-SPOT.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 IGRAs 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 IGRAs (Range: 50-81%).

[B5] 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

[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, 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.

C1 1 longitudinal study was included which assessed concordance between TST and IGRA conversions, using the QFT test.

C2 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.

C3 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.

C4 This study estimated fair concordance between QFT and TST conversions (96%).

C5 Only 1 study was identified with a small number of participants (n=216).

C6 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.

Table 10. GRADE Summary of Findings – IGRAs for tuberculosis screening of healthcare workers in low- and middle-income countries

Review question: What is the role of IGRAs in the diagnosis of latent tuberculosis infection (LTBI) in health care workers (HCWs)?
Study Population: Healthcare workers being screened for LTBI, all ages, from middle and low income countries.
Setting: Occupational screening of HCWs for LTBI
Index test: QuantiFERON-Gold or Gold In-tube (QFT) and T-SPOT.TB
Importance: The performance of IGRAs in diagnosing LTBI in HCWs is uncertain, it is unclear if IGRAs should be used in HCWs to identify those who could benefit from preventive therapy. In particular, it is unclear whether IGRA conversions identify those who could benefit from preventive therapy.
Reference standard: See hierarchy of reference standards (Figure 1)
Studies: Observational studies (longitudinal cohort, cross-sectional, case-control)

Outcome	No. Partici-pants	Principal Findings	What do these findings mean?	Quality of Evidence	Importance
Efficacy of preventive therapy based on IGRA test results	No Studies in HCWs				Critical (7-9)
Predictive value of IGRA for active TB	No Studies in HCWs				Critical (7-9)
Correlation between IGRA positivity and occupational TB exposure	991 (2 studies)	1) T-SPOT.TB: No studies evaluated T-SPOT.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.	Data were limited on T-SPOT.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.	Low ⊕⊕○○	Critical (7-9)

Correlation between IGRA conversions and occupational TB exposure	No Studies in HCWs				Critical (7-9)
Sensitivity for active TB (as a surrogate reference standard for LTBI)	No Studies in HCWs				Important (4-6)
Concordance between TST and IGRAs	1,357 (4 studies)	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.	Concordance was fair to poor in low and middle income settings. Both tests provide similar estimates of prevalence in low and middle income countries.	Very Low ⊕○○○	Important (4-6)
Concordance between IGRA and TST conversions	216 (1 study)	This study found 96% agreement between test conversions (QFT & TST).	IGRA and TST conversions show moderate concordance. Data are limited in all settings.	Very Low ⊕○○○	Important (4-6)

Table 11. GRADE Evidence Profile: Performance of IGRAs for the diagnosis of LTBI in contacts of active TB in low-and middle-income countries

No of Participants (Studies)	Study design	Limitations	Indirectness	Inconsistency	Imprecision	Publication bias	Quality of evidence (GRADE)[1]	Importance
A. Efficacy of preventive therapy based on IGRA test results								
No Studies								Critical (7-9)
B. Predictive value of IGRA for active TB								
9 studies:	Covered in	Predictive SR:	Rangaka et al					Critical (7-9)
C. Outcome: Correlation between IGRAs and different gradients of TB exposure (ordinal, continuous, etc.)								
3,868 (9)[A1]	Cross-sectional	Serious[A2] (-1)	No Serious indirectness[A3]	Serious[A4] (-1)	No serious imprecision[A5]	Likely[A6]	Low ⊕⊕○○	Critical (7-9)
D. Outcome: Correlation between IGRAs and TB exposure as a dichotomous variable								
3,145 (6)[B1]	Mainly cross-sectional	Serious[B2] (-1)	No Serious indirectness[B3]	Serious[B4] (-1)	Serious[B5] (-1)	Likely[B6]	Very Low ⊕○○○	Critical (7-9)
E. Outcome: Correlation between IGRA conversions and TB exposure								
309 (2)[C1]	Longitudinal	Serious[C2] (-1)	No Serious indirectness[C3]	Very Serious[C4] (-2)	Serious[C5] (-1)	Likely[C6]	Very Low ⊕○○○	Critical (7-9)
F. Outcome: Sensitivity for active TB (as a surrogate reference standard for LTBI)								
No Studies								Important (4-6)
G. Outcome: Concordance with tuberculin skin test (TST)								
5,080 (16)[D1]	Mainly cross-sectional	Serious[D2] (-1)	Very Serious[D3] (-2)	Very Serious[D4] (-2)	Serious[D5] (-1)	Likely[D6]	Very Low ⊕○○○	Important (4-6)

[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 T-SPOT.TB and QFT-GIT, 2 studies evaluated T-SPOT.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 T-SPOT.TB and QFT-G, while 1 study evaluated T-SPOT.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.

[C2]1 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 generalizeable 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 a positive association between conversions and exposure, while the other shows a significant protective effect of exposure for conversions.
[C5]95% 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 T-SPOT.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% IGRAs estimated a higher prevalence of LTBI.
[D5]Due to heterogeneity in effect estimates concordance could not be pooled. However, effects estimated for individual 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.

Table 12. GRADE Evidence Profile: Predictive value of commercial IGRA for incident active TB in low- and middle-Income countries

No of Participants (Studies)	Study design	Limitations	Indirectness	Inconsistency	Imprecision	Publication Bias	Quality of Evidence (GRADE)[1]	Importance
A. Outcome: Efficacy of preventive therapy based on IGRA results								
No studies								Critical (7-9)
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?)								
7,392 (3) [B1]	Cohort studies	Serious (-1) [B2]	Serious (-1) [B3]	No serious inconsistency [B4]	Very Serious (-2) [B5]	Likely [6]	Very low ⊕○○○	Critical (7-9)
C. Outcome: Predictive value of IGRA for the development of active incident TB compared to the TST (Are IGRAs (positive vs. negative) have a stronger statistical association with subsequent active TB than the TST (positive vs. negative)?								
7,392 (3) [C1]	Cohort studies	Serious (-1) [C2]	Serious (-1) [C3]	No serious inconsistency [C4]	Very Serious (-2) [C5]	Likely [C6]	Very low ⊕○○○	Critical (7-9)
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)								
No studies								Important (4-6)
E. Outcome: Quantitative IGRA levels and subsequent rates of TB								
721 (1) [E1]	Cohort of TB case-contacts	Serious (-1) [E2]	Serious (-1) [E3]	Serious (-1) [E4]	Very Serious (-2) [E5]	Likely [E6]	Very low ⊕○○○	Important (4-6)
F. Outcome: Immunological phenotypes of discordant-concordant TST/IGRA pairs and subsequent rates of TB								
5,861 (2) [F1]	Cohort studies	Serious (-1) [F2]	Serious (-1) [F3]	Serious (-1) [F4]	Very Serious (-2) [F5]	Likely [F6]	Very low ⊕○○○	Important (4-6)
G. Outcome: Sensitivity, Specificity, False positive rates etc for active TB (as surrogates of patient relevant outcomes)								
7,392 (3) [G1]	Cohort studies	Serious (-1) [G2]	Serious (-1) [G3]	Serious (-1) [G4]	Very Serious (-2) [G5]	Likely [G6]	Very low ⊕○○○	Important (4-6)
H. Outcome: Utility of repeated or serial IGRA results for predicting subsequent incident active TB								
No studies								Important (4-6)

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: 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.

[B1] 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)

[B2] 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.

[B3] 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.

[B4] 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.

[B5] 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.

[B6] 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 (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 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.

[C1] 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)

[C2] 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.

[C3] 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.

[C4] 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.

[C5] The confidence intervals of the pooled IRRs are wide (precision > +/- 20%). This is a very serious limitation. Two points are deducted.

[C6] Publication bias was not formally assessed, but is deemed likely. See [B6].

[E1] 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)

[E2] 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.

[E3] There is only one study. There is serious indirectness. A point is deducted.

[E4] There is only one study; inconsistency cannot be assessed. A point is deducted.

[E5] The 95% confidence intervals per IGRA stratum were extremely wide (precision > +/- 20%). Two points are deducted.

[E6] Publication bias was not formally assessed, but is deemed likely. See [B6].

[F1] 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)

[F2] 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

[F3] Although results may be generalizable to similar L/MIC, there are only two studies. A point is deducted.

[F4] Rates of TB during follow-up may be higher in those with double positive TST+/IGRA+ results than in those with double negative results. Both studies seem to suggest this. However, contrasting results are seen with regard to discordant pairs. Pooled estimates were not derived. The inconsistency in results is deemed serious; a point is deducted.

[F5] Observed 95% confidence intervals around the rates per strata are wide (precision > +/- 20%).

[F6] Publication bias was not formally assessed, but is deemed likely. See B6.

[G1] 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 6-9months, depending on country guidelines. IPT, although safe, is not without serious adverse effects, notably, clinical hepatitis 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.

[G2] 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 most studies. A point is deducted

[G3] Although results may be generalizable to similar L/MIC, there are only three studies. A point is deducted.

[G4] There is heterogeneity in individual studies' test accuracy estimates (e.g. specificity/false positive rates). A point is deducted.

[G5] The summary estimates of sensitivity and specificity are moderate and the confidence intervals are wide (precision > +/- 20%). Two points are deducted.

[G6] Publication bias was not formally assessed, but is deemed likely. See [B6].

Table 13. GRADE Summary of Findings: Predictive value of commercial IGRA for incident active TB in low and middle-income countries

Review question: What is the predictive value of interferon-gamma release assays for incident active tuberculosis disease in low and middle-income countries?
Patients/population: Studies of adults or children without TB at baseline and regardless of HIV infection status.
Setting: Community-based cohort in a high-burden country, high-risk for TB individuals attending outpatients clinics and school-going adolescents residing in a high-burden country
Index test: Latest Commercial IGRA (QuantiFERON Gold In Tube and T-SPOT.TB)
Importance: The predictive value of IGRAs for subsequent incident TB is uncertain. Longitudinal studies on the predictive (prognostic) value of a positive IGRA are emerging. Data from these studies provide the initial evidence to refute or support the use of IGRAs in targeting chemoprophylaxis for IGRA-positive individuals.
Reference standard: Development of TB. See hierarchy of reference standards.
Studies: Any longitudinal study design (e.g. prospective or retrospective cohort), low and middle-income countries. Follow-up (of any length) should be described. This can either be active or passive follow-up.

Outcome	N (No. of studies)	Principal Findings	What do these findings mean?	Quality of Evidence	Importance
Efficacy of preventive therapy based on IGRA results	No studies				Critical (7-9)
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?)	7,392 (3)	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 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), I2=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	Moderate increase in incidence rates of TB in IGRA positives compared to IGRA negatives. This translates to moderate risk of progression. There are too few studies to conclude this with certainty. However, even in those with positive IGRA results, the vast majority of individuals did not progress to TB disease during follow-up.	Very low ⊕◯◯◯	Critical (7-9)

		interpreted with caution; there is low-high statistical heterogeneity.			
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)?	7,392 (3)	1) IGRA+: Pooled IRR=3.24 (0.62-4.69); I^2=0%, p=0.90 2) TST+: Pooled IRR=2.3 (0.83-3.73); I^2=0%, p=0.38 The derived estimates are not statistically significant; the confidence intervals include the null. The pooled estimates should also be interpreted cautiously: there are only three studies; heterogeneous populations and study methods	**IGRA+ and TST+ may have a similar strength of association with subsequent TB compared to test negative individuals.**	Very low ⊕○○○	Critical (7-9)
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)	No studies				Important (4-6)
Quantitative IGRA levels and subsequent rates of TB	721 (1)	No pooled estimates: there is only one study 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.	Inconclusive results. Number of studies assessed is too small.	Very low ⊕○○○	Important (4-6)

Immunological phenotypes of discordant-concordant TST/IGRA pairs and subsequent rates of TB	5,861 (2)	No pooled estimates. 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.	Inconclusive results. Numbers of studies is too small and/or the rate of TB observed per strata too low.	Very low $\oplus\bigcirc\bigcirc\bigcirc$	Important (4-6)
Sensitivity, Specificity, False positive rates etc for active TB (as surrogates of patient relevant outcomes)	7,392 (3)	No pooled results. 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	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+	Very low $\oplus\bigcirc\bigcirc\bigcirc$	Important (4-6)

		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.	for the Zambia study. However, lower TST results may have resulted from logistical issues.		
Utility of repeated or serial IGRA results for predicting subsequent incident active TB	No studies				Important (4-6)

Annex 1: List of Participants- Expert Group meeting

Expert Group meeting on use of interferon-γ release assays (IGRAs) in tuberculosis control in low- and middle-income settings

Geneva, Switzerland, 19-20 July 2010

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 Catharina Boehme
Foundation for New Innovative New Diagnostics
(FIND)
16 Avenue de Budé
1202 Geneva
Switzerland
catharina.boehme@finddiagnostics.org

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

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 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 10006USA
adetjen@theunion.org

Dr David Dowdy
1236 3rd Ave., Apt. #3
San Francisco, CA 94122
USA
david.dowdy@ucsf.edu

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

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
oluwamayjoel@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 Richard Menzies
Montreal Chest Institute
Room K1.24
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 Rick 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 Holger Schünemann
Department of Clinical Epidemiology &
Biostatistics
McMaster University Health Sciences Centre,
Room 2C10B 1200 Main Street West Hamilton
Canada
schuneh@mcmaster.ca

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 Lakhbir Singh Chauhan
Deputy Director General of Health Services
Ministry of Health and Family Welfare
Nirman Bhavan
110011 - New Delhi
India
ddgtb@rntcp.org

World Health Organization

WHO HQ Staff
Dr Chris Gilpin, STB/TBL
gilpinc@who.int

Mr Jean Iragena, STB/TBL
iragenaj@who.int

Dr Regina Kulier, Secretariat, GRC
kulierr@who.int

Dr Christian Lienhardt, TBP
lienhardtc@who.int

Dr Fuad Mirzayev, STB/TBL
mirzayevf@who.int

Dr Mario Raviglione, STB
raviglionem@who.int

Dr Karin Weyer STB/TBL
weyerk@who.int

Dr Matteo Zignol, STB/TBS
zignolm@who.int

WHO/TDR
Dr Luis Cuevas
cuevasl@who.int

Dr Jane Cunningham
cunninghamj@who.int

Dr Andy Ramsay
ramsaya@who.int

Dr Soumya Swaminathan
swaminathans@who.int

Annex 2: List of Participants- STAG-TB

10th Meeting Strategic and Technical Advisory Group for Tuberculosis (STAG-TB)

27-29 September 2010, WHO Headquarters, Geneva, Switzerland

Dr Salah Al Awaidy
Director
Department of Communicable
Disease Surveillance & Control
Oman

Dr Kenneth Castro
Director, Division of TB Elimination
Centers for Disease Control and
Prevention
USA

Dr Jeremiah Muhwa Chakaya
(STAG-TB Chair)
Technical Expert
National Leprosy and TB
Programme
Ministry of Health
Kenya

Ms Lucy Chesire
TB Advocacy Adviser
Kenya AIDS NGOs Consortium
KANCO
Kenya

Dr Elizabeth Corbett
Reader in Infectious and Tropical
Diseases
London School of Tropical Medicine
& Hygiene
and MLW Research Programme
Malawi

Dr Charles L. Daley
Head, Division of Mycobacterial and
Respiratory Infections
National Jewish Health
USA

Dr Pamela Das
Executive Editor
The Lancet
United Kingdom

Prof. Francis Drobniewski
Director, Health Protection Agency
National Mycobacterium Reference
Unit
Institute for Cell and Molecular
Sciences,
United Kingdom

Dr Wafaa El-Sadr
CIDER
Mailman School of Public Health
Columbia University
USA

Dr Paula I. Fujiwara
(STAG-TB Vice Chair)
Director, Department of HIV and
Senior Advisor
The Union
France

Dr Yuthichai Kasetjaroen
Director
Bureau of Tuberculosis
Ministry of Health
Thailand

Prof Vladimir Malakhov
National Center for External Quality
Assessment in Laboratory Testing
of Russian Federation
Russian Federation

Dr Mao Tan Eang
Advisor to the Minister of Health
Director, National Center for
Tuberculosis and Leprosy Control
Ministry of Health
Cambodia

Dr Giovanni Battista Migliori
Director
WHO Collaborating Centre for
Tuberculosis and Lung Diseases
Fondazione Salvatore Maugeri
Italy

Dr Megan Murray
Associate Professor of
Epidemiology
Harvard University School of Public
Health
Department of Epidemiology
USA

Dr Yogan Pillay
Deputy Director General
Strategic Health Programmes
Department of Health
South Africa

Dr Ren Minghui
Director-General
Department of International
Cooperation
Ministry of Health
People's Republic of China

Dr Rajendra Shukla
(unable to attend)
Joint Secretary
Ministry of Health & Family Welfare
India

Dr Pedro Guillermo Suarez
TB & TB-HIV/AIDS Division
Center for Health Services
Management Sciences for Health
USA

Dr Marieke van der Werf
Head, Unit Research, Senior
Epidemiologist
KNCV Tuberculosis Foundation
The Netherlands

Dr Rosalind G. Vianzon
National TB Programme Manager
National Center for Disease Control
and Prevention
Department of Health
Philippines

Dr Tido Von Schön-Angerer
Campaign for Access to Essential
Medicines
Medicins Sans Frontieres
Switzerland

World Health
Organization