Tuberculosis (TB)

TB diagnostics and laboratory strengthening - WHO policy

The use of liquid medium for culture and DST, 2007

WHO recommends, as a step-wise approach:

1. The use of liquid medium for culture and DST in middle- and low-income countries.

2. The rapid species identification to address the needs for culture and drug susceptibility testing (DST).

Taking into consideration that liquid systems will be implemented in a phased manner, integrated into a country specific comprehensive plan for laboratory capacity strengthening and addressing the following key issues:

1. Appropriate biosafety level;
2. detailed customer plan describing guarantees and commitments of the manufacturer;
3. appropriate training of staff;
4. maintenance of infrastructure and equipment in laboratories;
5. quick transportation of samples from the peripheral to the culture laboratory;
6. rapid communication of results.

Use of Liquid TB culture and drug susceptibility testing (DST) in low- and medium- income settings

International TB laboratory experts and representatives of partner organizations recommend the use of TB liquid culture and DST in low income settings. Liquid culture systems are the standard of care for TB diagnosis and patient management in industrialized countries.

Liquid culture and DST systems are more complex and sensitive than solid culture and DST media. Increased bacterial contamination and an increased frequency of nontuberculous mycobacterial (NTM) isolation must

be addressed. A rapid method to differentiate M. tuberculosis complex from other mycobacterial species is essential.

**Background**

Laboratory diagnosis of TB largely relies on the direct microscopic examination of sputum specimens. However, the technique, although specific, has low and variable sensitivity and cannot identify drug-resistant strains. Mycobacterial culture is more sensitive but growth of TB bacilli on traditional solid medium requires 4-8 weeks and consequently delays appropriate treatment in the absence of a confirmed diagnosis.

Expanding culture capacity is urgently needed to address challenges due to the epidemics of HIV-associated TB and drug resistant TB, especially in resource-limited settings.

Liquid culture systems reduce the delays in obtaining results to days rather than weeks. For DST, the delay may be reduced to as little as 10 days, compared to 28-42 days with conventional solid media. Liquid systems are more sensitive for detection of mycobacteria and may increase the case yield by 10% over solid media. With increased sensitivity and reduced delays, liquid systems may contribute significantly to improved patient management.

Liquid systems are, however, more prone to contamination by other microorganisms. In experienced laboratories, approximately 5-10% of specimens cannot yield results because of contamination. Procedures to prevent cross-contamination (due to carryover of bacilli from positive to negative specimens) should also be strictly followed, especially where more positive specimens are processed in high-incidence countries.

**Phased implementation**

The decision to implement a liquid culture and DST system should be based on need and be consistent with a country’s plan for TB laboratory capacity strengthening and expansion. Such plans should be considered only in countries with a strong network of quality-assured microscopy.

In most circumstances, the first priority would be to implement the system in the NRL, assuming that the NRL is currently supervising quality-assured (QA) microscopy of the laboratory network and performing QA TB culture and DST. This would provide valuable experience that could be applied if a decision were made to scale up the system. In those countries where a liquid culture and DST system is currently in use at the NRL, the decision to scale up should be informed by the accumulated experience at the NRL with the liquid culture system.

Subsequent expansion of liquid culture and DST capacity would logically be to regional TB culture and DST laboratories. The extent of scale up should be determined by need and availability of funding, and again consistent with a national laboratory plan.
Species identification

It is imperative that all mycobacterial isolates be speciated at least to the level of M. tuberculosis complex vs. NTM. When using liquid culture, with the expectation that time-to-detection will be significantly reduced, it is also imperative that a rapid and affordable method of species identification be used. Where identification of NTM is needed, standard biochemical tests or other methods can be considered.

Key issues for WHO

Ministries of Health and their partners urgently require guidance from WHO on the use of liquid culture and other means to improve diagnostic capacity. Based on an expert consultation organized by WHO (26 March 2007) with key experts and agencies in this field, and building on recent comprehensive review of available evidence regarding the efficacy, effectiveness and feasibility of implementing liquid culture technology in high TB burden settings, recommendations were made.

Recommendations

1. Adoption of liquid culture systems should be decided by Ministries of Health in the context of a comprehensive and detailed country plan for TB laboratory capacity strengthening.

2. Country plans for expansion of TB culture and drug-susceptibility testing (DST) should be based on a strong network of quality-assured microscopy, the cornerstone for TB diagnosis.

3. Laboratories should have demonstrated experience in culture and DST using conventional methods.

4. Phased implementation is recommended with the National Reference Laboratory (NRL) as a first priority and further scale-up to regional laboratories based on the NRL experience and consistent with the national country plan.

5. Adequate infrastructure and equipment should be provided, especially regarding laboratory biosafety. Specimen processing for culture purposes has to be performed in appropriate Biological Safety Cabinets (BSCs), at least in Biosafety Level 2 (BSL2) facilities. Processing of cultures for conventional species identification, subculturing and phenotypic DST must be performed in BSL3 facilities [Ref. 9], since culture suspensions required for these activities generate highly infectious aerosols with a high concentrations of TB bacilli. The successful establishment, staffing and maintenance of BSL3 laboratories is demanding and costly.

6. Commercial liquid culture systems must include a detailed commercial sales contract which guarantees ample and continuous supply, optimal shipment conditions and logistics for custom clearance.

7. A customer support plan should detail measures that guarantee - by the supplier - equipment installation, maintenance, reparation and provision of training, training materials and technical support.

WHO will include the use of liquid culture in all relevant technical documents (e.g. Standard Operational Procedures, training material for culture and DST) and will support countries in assessing their needs and building capacity to use liquid cultures.

References


TB diagnostics and laboratory strengthening - WHO policy

Definition of a new sputum smear-positive TB case, 2007

The revised definition of a new sputum smear-positive pulmonary TB case is based on the presence of at least one acid fast bacilli (AFB+) in at least one sputum sample in countries with a well functioning external quality assurance (EQA) system.

Background

As highlighted in the Stop TB Strategy, quality-assured bacteriological examination is an essential element for diagnosis and management of TB patients harbouring susceptible or resistant bacilli. During the last two years, an increasing number of countries are scaling up external quality assurance programmes for smear microscopy by means of blinded re-checking of slides. As a result, the quality of smear microscopy examination reached a satisfactory level in some countries. Evidence suggests that countries with a functional EQA system have very low frequency of false positive cases.

Key issues for WHO action

A number of key meetings and workshops were held where the TB case definition was discussed. These meetings included the Stop TB Partnership Laboratory Strengthening Subgroup (SLCS), an expert group meeting organized by the UNION held in Belgium and a technical expert workshop held in the Netherlands. Recent scientific evidence [Ref. 1,2] was reviewed and it was concluded that where a functional EQA for smear microscopy is in place, the finding of a single AFB in at least one single sputum smear examination in a TB suspect would satisfy the criterion to report a patient as having "sputum smear-positive tuberculosis" and to subsequently start treatment.
It should be noted that the definition of bacteriological failures has not been reviewed; hence, no change in definition of failure cases is proposed at this stage.

Given this policy revision, WHO will:

• guide and support countries in making country-specific plans of action for modifying all normative, training, and recording and reporting tools;
• provide technical assistance to countries to upgrade and fully expand functional external quality assurance (EQA) systems for TB laboratory services;
• provide guidance on study design, and sampling methodologies, in order to evaluate new diagnostic technologies (in collaboration with the Special Programme for Research and Training in Tropical Diseases (TDR));
• monitor and evaluate the impact of the change of policy on case detection at country level.

References


2. Bonnet M, Ramsay A, Gagnidze L, Githui W, Guerin PJ, Varaine F. Reducing the number of sputa examined, and thresholds for positivity: An opportunity to optimize smear microscopy. Accepted for publication, *Int J Tuberc Lung Dis*
Tuberculosis (TB)

TB diagnostics and laboratory strengthening - WHO policy

Reduction of number of smears for the diagnosis of pulmonary TB, 2007

WHO recommends the number of specimens to be examined for screening of TB cases can be reduced from three to two, in places where a well-functioning external quality assurance (EQA) system exists, where the workload is very high and human resources are limited.

Related links

HOME - TB diagnostics and laboratory strengthening
WHO policy statements on strengthening TB laboratories

Detailed background information
pdf, 151kb

Background

The WHO Stop TB Strategy and the Global Plan to Stop TB, 2006-2015 recognizes the weakness of the health system as one of the greatest challenges to TB control and indeed to the achievement of the Millenium Development Goals (MDGs) in general. The Global Plan also recognizes that patients, particularly poor patients, face economic barriers in accessing TB control services and that patients with TB in many resource-limited settings face long and sometimes costly pathways to diagnosis. In most of these countries, the laboratory services are often neglected and may be considered to be among the weakest components of the health system.

The challenge is particularly great in sub-Saharan Africa, where the direct and indirect effects of the HIV epidemic exacerbate the human resource crisis and compound the essential but neglected component of adequate TB diagnostic capacity within the health system.

Key issues for WHO action

The current international policy on TB case detection recommends the examination of three sputum smears for the diagnosis of pulmonary tuberculosis (PTB). The present definition of a smear-positive case states "Tuberculosis in a patient with at least two initial sputum smear examinations (direct smear microscopy) positive for acid fast bacilli (AFB+)". [Ref. 1, 2]

A systematic review of 37 eligible studies that quantified the incremental diagnostic yield of serial sputum specimens was performed by Mase et al
and published recently. [Ref. 3] The results clearly demonstrated that the vast majority of TB cases (on average 85.8%) was detected with the first sputum specimen. With the second sputum specimen, the average incremental yield was 11.9%, while the incremental yield of the third specimen, when the first two specimens were negative, was 3.1%. [Ref. 3]

In a recent study conducted in Kenya, Bonnet et al. demonstrated that decreasing the number of smears examined for the detection of new pulmonary TB cases lead to a reduction of patient's visits to a clinic and the laboratory workload. Examining only two smears could therefore alleviate the workload of laboratories - particularly in countries with a high microscopy workload - by one third. [Ref. 4]

It is expected that microscopic analysis of two sputum smear samples will improve case findings through enhanced quality of service, decreased time for diagnosis and initiation of treatment and decreased number of patients dropping out of the diagnostic pathway.

**However, the reduction of the number of specimens examined for screening TB patients from three to two specimens should only be recommended in settings with a well-established laboratory network, a fully functional EQA programme for smear microscopy including on-site evaluation with the feed-back mechanism and where the workload is very high and human resources are limited.** [Ref. 5]

Given the new policy, WHO will:

- guide and support countries in making country-specific plans of action for modifying relevant normative, training, and recording and reporting tools;
- provide technical assistance to countries to upgrade and fully expand functional external quality assurance (EQA) systems for TB microscopy services;
- provide guidance on study design, and sampling methodologies, in order to evaluate new diagnostic technologies (in collaboration with the Special Programme for Research and Training in Tropical Diseases (TDR));
- monitor and evaluate the impact of the policy change on case detection at country level.

**References**


5. External quality assessment for AFB smear microscopy (CDC web site) [pdf 631kb]
Policy guidance on drug-susceptibility testing (DST) of second-line antituberculosis drugs

World Health Organization
Geneva
2008

WHO/HTM/TB/2008.392
Contents

Preamble 4
Introduction 6
Current knowledge 8
Rational use of DST in programmes for control of drug-resistant TB 12
Conclusions 18
References 19
Preamble

Results from drug-resistance surveys and ongoing surveillance show that drug-resistant tuberculosis (TB) is widespread geographically (1). Drug-resistant TB is a man-made problem of global concern – the result of mismanagement of antituberculosis drugs through poor TB control, drug-prescription errors and non-adherence of patients to treatment. However, the extent of the problem remains underestimated or unknown in many settings owing to insufficient laboratory capacity and inadequate policies to detect drug-resistant TB patients accurately and in a timely manner. Multidrug-resistant TB (MDR-TB) has become a serious threat to global TB control as a result of the difficulties in diagnosis and treatment and the associated high cost to TB control programmes. Documented transmission of MDR-TB to vulnerable populations and in high-burden HIV settings compounds this threat (2). The emergence of extensively drug-resistant TB (XDR-TB), with poor treatment outcomes, very high mortality in XDR-TB patients with concomitant HIV infection (3), and the risk of XDR-TB spread across country borders, has heightened global concern over a potentially untreatable epidemic that may jeopardize recent advances in global TB control.

Guided by the Stop-TB Partnership Working Group on MDR-TB and the Green Light Committee (GLC), concurrent efforts by various private, nongovernmental and public organizations focus on confronting the challenges of drug-resistant TB, and sharing information and strategies in an unprecedented, collaborative way. However, estimates by the World Health Organization (WHO) highlight the need for diagnostic capacity as one of the most crucial aspects in mobilizing an effective response to the challenges of drug-resistant TB, with fewer than 5% of existing MDR-TB cases estimated to be diagnosed (4). The weakest link in TB control remains the need for appropriate, affordable and sustainable laboratory services, and this has been brought into stark relief by the pressing need for an accelerated and extensive scale-up of MDR-TB programmes.

GLC-assisted projects in different epidemiological and resource-constrained settings have shown that the management of MDR-TB is feasible and effective, even in resource-constrained settings (5). However, major challenges remain in the area of laboratory capacity to meet the demand for scale-up of MDR-TB programmes within the context of routine TB control. Laboratory constraints are centred on: programmatic requirements such as infrastructure development, acquisition and maintenance of equipment, quality assurance and biosafety; an urgent need for reliable and reproducible methodologies for second-line drug-susceptibility testing; and the need for rational use of second-line DST in programmes about to engage in MDR-TB treatment. In order to address these issues, WHO has taken the lead in developing interim laboratory policy guidance for countries establishing or expanding MDR-TB treatment programmes.

1 MDR-TB: Mycobacterium tuberculosis complex isolates with in vitro resistance against isoniazid and rifampicin, with or without resistance to additional first-line anti-TB drugs.
2 XDR-TB: Mycobacterium tuberculosis complex isolates defined as multidrug-resistant, with additional in vitro resistance to a fluoroquinolone and one or more of the following injectable drugs: kanamycin, amikacin, capreomycin.
Aim

This document is intended to provide an interim policy framework for the laboratory component relevant to programmatic implementation of MDR-TB strategies. A detailed technical manual on laboratory methodology, laboratory biosafety and standard operating procedures related to second-line drug-susceptibility testing (DST) is also under preparation.

Process

When preparing this document, priority consideration was given to data from published studies; however, scientific literature is limited and extrapolation from expert opinion and experience within laboratories involved in second-line DST was very useful in developing current consensus on second-line DST procedures. To this end, a core group of international TB laboratory experts reviewed the available literature, shared experiences and provided consensus expert opinion on controversial technical issues.

Date of review

Given the paucity of scientific data on several aspects of second-line DST, the need for additional research and rapid translation of research findings into policy and practice is evident. This document therefore constitutes work in progress and will be complemented by ongoing and future research, guided by increased collaboration of partners involved in laboratory services, and subject to review by early 2010.

Conflict of interest

F. Drobniewski: grants from Becton Dickinson and Co. and association with Life Sciences for Diagnostics.
Introduction

One of the main aims of effective TB control is the prevention of drug resistance resulting from a variety of programmatic, health provider- and patient-related factors. Irregular drug supply, poor drug quality, clinical errors in drug prescription and a lack of patient adherence to treatment are known determinants of anti-TB drug resistance (5). Subsequent transmission of resistant bacilli is facilitated by inadequate infection control, especially in congregate settings. MDR-TB and XDR-TB outbreaks have almost invariably been linked with HIV infection (2, 3), resulting in exceptionally high patient mortality and highlighting the urgent need for rapid diagnosis and intervention in vulnerable populations.

Definitive diagnosis of MDR-TB and XDR-TB requires that *Mycobacterium tuberculosis* be isolated and identified, and drug-susceptibility testing (DST) completed. Using conventional methodologies, growth detection, identification of *M. tuberculosis* and DST may take weeks or even months. In addition, the interpretation of DST results for TB bacilli is complicated by the fact that organisms may be intra- or extracellular, may have a long generation time, may be dormant or active, and may be present in different tissues with variable drug-penetration ability. DST results may therefore not accurately reflect the bacterial population by the time the results become available, and cannot be exclusively relied upon to guide the design of treatment modalities.

Newer rapid phenotypic DST methods (e.g. direct tests, colorimetric methods, phage-based methods) and genotypic DST techniques (e.g. nucleic acid amplification assays, resistance mutation detection and sequence-based assays) are very promising but are either still in development, at early validation stage or in early field demonstration phase, and only aimed at first-line anti-TB drugs. While presenting an opportunity for rapid detection of MDR-TB, no tests for rapid identification of second-line drug resistance are yet available.

Conventional DST for first-line anti-TB drugs has been thoroughly studied and consensus has been reached on appropriate methodologies, critical drug concentrations, and reliability and reproducibility of testing. On the other hand, surveys of current practices for second-line DST in the global Supranational Reference Laboratory (SRL) Network as well as a few multicentre laboratory studies have revealed important differences with regard to methods, the critical concentrations of drugs, and the critical proportions of resistance (6–8). The reliability of drug-susceptibility testing for second-line drugs (SLDs) has therefore been questioned (9–10) and the urgent need to standardize methodologies, establish criteria for defining resistance and carry out proficiency testing is obvious. Recent studies have compared newer methodologies with conventional DST for selected SLDs and have suggested tentative critical concentrations for these drugs (6–8).

It should be noted however that no studies have systematically evaluated all available DST methods for all available SLDs, established critical concentrations for all available SLDs, or evaluated a large number of clinical isolates for microbiological and clinical end-points.
Countries embarking on diagnostic and treatment programmes for drug-resistant TB need policy guidelines on the rational use of DST, particularly for second-line drugs. Policy formulation has, however, been hampered by the following:

- Second-line DST has not been standardized internationally, owing to technical difficulties related to in vitro drug instability, drug loss caused by protein binding, heat inactivation, filter sterilization, incomplete dissolution and/or varying drug potency. Laboratory technique, medium pH, incubation temperature and incubation time also influence DST results. In addition, the drug critical concentration defining resistance is often very close to the minimal inhibitory concentration (MIC) required to achieve antimycobacterial activity, increasing the probability of misclassification of susceptibility or resistance, and leading to poor reproducibility of DST results.

- Only a few laboratories internationally have the required capacity and expertise to reliably test for all classes of available anti-TB drugs. These laboratories are largely limited to resource-rich settings. Many of the newer techniques are difficult to implement in the countries where they are most needed owing to high cost, technical complexity and lack of appropriately trained laboratory staff. As a result, conventional culture and DST methods using egg-based or agar-based media are still the most widely used in resource-limited settings, leading to long diagnostic delays. Even in sophisticated and well-resourced environments, wide variations in second-line DST systems and methods have been reported, reflecting the difficulties in securing reproducibility and optimizing the clinical relevance of DST results. In addition, the majority of newer techniques still need proper evaluation to verify their efficiency in different epidemiological settings.

- Many high-burden TB countries do not have access to the full range of second-line drugs because of financial, regulatory or other constraints. Fluoroquinolones, aminoglycosides and (to a much lesser extent) polypeptides are readily available in many countries, although specific drugs in these classes may not be. Cross-resistance between drugs in the same group further limits the selection of available drugs. In settings with limited access to SLDs, development of resistance to the most potent groups of SLDs (aminoglycosides, fluoroquinolones, polypeptides) therefore creates a situation where TB is virtually untreatable.
Current knowledge

Drug efficacy

The WHO Guidelines for the programmatic management of drug-resistant tuberculosis (5) categorize available anti-TB drugs in five groups, based on known efficacy (Table 1). The backbone of regimens for the treatment of MDR-TB consists of an injectable drug (aminoglycoside or polypeptide) and a fluoroquinolone, supported by at least two additional SLDs in order to ensure that the regimen includes at least four drugs confirmed or expected to be effective (5).

Table 1  Alternative method of grouping antituberculosis drugs

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 First-line oral agents</td>
<td>Isoniazid (H); rifampicin (R); ethambutol (E); pyrazinamide (Z); rifabutin (Rfb)(^a)</td>
</tr>
<tr>
<td>Group 2 Injectable agents</td>
<td>Kanamycin (Km); amikacin (Am); capreomycin (Cm); viomycin (Vm); streptomycin (S)</td>
</tr>
<tr>
<td>Group 3 Fluoroquinolones</td>
<td>Moxifloxacin (Mfx); levofloxacin (Lfx); ofloxacin (Ofx)</td>
</tr>
<tr>
<td>Group 4 Oral bacteriostatic second-line agents</td>
<td>Ethionamide (Eto); protionamide (Pto); cycloserine (Cs); terizidone (Trd); (p)-aminosalicylic acid (PAS)</td>
</tr>
<tr>
<td>Group 5 Agents with unclear role in DR-TB treatment (not recommended by WHO for routine use in DR-TB patients)</td>
<td>Clofazimine (Cfz); linezolid (Lzd); amoxicillin/clavulanate (Amx/Clv); thioacetazone (Thz); imipenem/cilastatin (Ipm/Cln); high-dose isoniazid (high-dose H);(^b) clarithromycin (Clr)</td>
</tr>
</tbody>
</table>

\(^a\) Rifabutin is not on the WHO List of Essential Medicines. It has been added here as it is used routinely in patients on protease inhibitors in many settings.

\(^b\) High-dose H is defined as 16–20 mg/kg/day.

Aminoglycosides, polypeptides and fluoroquinolones are bactericidal (able to achieve killing of mycobacteria), while thioamides, cycloserine/terizidone and \(p\)-aminosalicylic acid (PAS) are bacteriostatic (able to prevent growth of mycobacteria).

Once the injectable drugs and the fluoroquinolones are compromised by resistance, available treatment regimens become much weaker and the possibility for patient cure decreases significantly. Significantly more clinical data are needed to answer key questions relating to treatment outcomes in the presence of different combinations and permutations of drug resistance. However, evidence from Latvia shows that the rate of successful treatment outcome precipitously falls as resistance to the key SLDs increases – in a cohort of 820 MDR-TB patients who had completed treatment with
SLDs, two thirds overall had a successful outcome. However, in those patients with strains resistant to an injectable drug (kanamycin or capreomycin) and a fluoroquinolone (ofloxacin), less than 30% were successfully treated. Among those with strains resistant to kanamycin and ofloxacin specifically, only 24% had a successful treatment outcome (personal communication, Vaira Leimane, Latvia; Timothy Holtz, CDC).

Where bacterial resistance to aminoglycosides, polypeptides and fluoroquinolones is associated with additional resistance to the bacteriostatic drugs, treatment regimens are weakened even further and become virtually ineffective. As has been shown in South Africa (3), such infections are lethal for HIV-coinfected patients.

Reliability and reproducibility of second-line DST

Aminoglycosides, polypeptides and fluoroquinolones have been tested in different laboratory environments and shown to have relatively good reliability and reproducibility (6–8, 11–13). Data on the reproducibility and reliability of DST for the other SLDs are either much more limited or have not been established, or the methodology for testing does not exist.

Most importantly, correlation of second-line DST results with clinical response to treatment has not yet been adequately established. As a result, the prognostic relevance of in vitro resistance remains unclear for the majority of SLDs.

Cross-resistance

Cross-resistance between the older-generation fluoroquinolones is almost complete; limited evidence has suggested that the third-generation quinolones (notably moxifloxacin and gatifloxacin) may have enhanced clinical benefit – even in the presence of in vitro resistance to ofloxacin or ciprofloxacin – owing to their low MICs, enhanced antimycobacterial activity, and improved biochemical structure providing metabolic stability and long half-life, theoretically reducing the selection of resistant mutants (12, 13). This observation remains to be confirmed in controlled clinical trials.

Study findings on cross-resistance between the aminoglycosides and/or the polypeptides are contradictory. In summary, the genotypes associated with resistance to the aminoglycosides and the cyclic polypeptides are overlapping; therefore, a combination of an aminoglycoside and a polypeptide would be equivalent to using a single drug. However, a recent study (albeit small) showed that cross-resistance patterns and MICs vary among the different molecular mutations described (14).

Although emerging evidence shows a clear association between drug resistance and specific molecular mutations, this association does not actually prove a role for the individual mutations in drug resistance. Generalizing resistance to a class of SLDS based solely on resistance to a single drug in the class may therefore be misleading, as summarized below (14).

- Isolates that acquire resistance to capreomycin are usually susceptible to kanamycin and amikacin. A small proportion may be resistant to kanamycin and...
an even smaller proportion may be resistant to amikacin. The molecular basis for this observation has been described.

- Isolates that acquire resistance to amikacin essentially always have associated resistance to kanamycin and capreomycin. The molecular basis for this observation has been described.

- Isolates that acquire resistance to kanamycin show different levels of cross-resistance with amikacin and capreomycin. The molecular basis for some of these observations has been described;

- Isolates that acquire resistance to streptomycin are usually susceptible to kanamycin, amikacin and capreomycin. However, rare strains with apparently single-step mutations that confer resistance to both streptomycin and kanamycin have been observed, although the molecular mechanism is not known.

Table 2 summarizes current consensus on the reliability and reproducibility of DST for anti-TB drugs, based on a robust assessment of published studies combined with laboratory experience and expert opinion. The following broad criteria were used to assess the strength of available evidence, based on two or more criteria having been met for assigning a drug to a specific category.

I. Extensive published studies, extensive multicentre laboratory review, broad intermethod agreement, high stability of drug powder in vitro, consistent DST reliability and reproducibility, extensive clinical outcome data.

II. Extensive published studies, extensive multicentre laboratory review, limited intermethod agreement, variable DST reproducibility (and therefore reliability), variable stability of drug powder in vitro, less extensive clinical outcome data.

III. Less extensive published studies, limited multicentre laboratory review, limited intermethod agreement, limited data on DST reproducibility and reliability, limited data on drug powder stability in vitro, limited clinical outcome data.

IV. Limited or no published studies, limited multicentre laboratory review, limited data or questionable DST reproducibility (and therefore reliability), instability of drug powder in vitro, no clinical outcome data.

V. No published studies, no multicentre laboratory review, reproducibility and reliability impossible to assess, unknown stability of drug powder in vitro, no clinical outcome data.
<table>
<thead>
<tr>
<th>Drug group&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Drug</th>
<th>DST category</th>
<th>DST method available</th>
<th>DST critical concentrations (μg/ml)</th>
<th>BACTEC460</th>
<th>MGIT960</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-line oral anti-TB agents</td>
<td></td>
<td>Solid, liquid</td>
<td></td>
<td>Löwenstein-Jensen&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H11&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Isoniazid</td>
<td>I</td>
<td>Solid, liquid</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Rifampicin</td>
<td>I</td>
<td>Solid, liquid</td>
<td>40.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Ethambutol</td>
<td>II</td>
<td>Solid, liquid</td>
<td>2.0</td>
<td>5.0</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>Pyrazinamide</td>
<td>II</td>
<td>Liquid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injectable anti-TB agents</td>
<td></td>
<td>Solid, liquid</td>
<td></td>
<td>Löwenstein-Jensen&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H11&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Streptomycin</td>
<td>II</td>
<td>Solid, liquid</td>
<td>4.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Kanamycin</td>
<td>II</td>
<td>Solid, liquid</td>
<td>30.0</td>
<td>5.0</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Amikacin</td>
<td>II</td>
<td>Liquid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Capreomycin</td>
<td>II</td>
<td>Solid, liquid</td>
<td>40.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>Viomycin</td>
<td>V</td>
<td>None</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluoroquinolones</td>
<td>Ciprofloxacin&lt;sup&gt;d&lt;/sup&gt;</td>
<td>III</td>
<td>Solid, liquid</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Ofloxacin</td>
<td>III</td>
<td>Solid, liquid</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Levofloxacin</td>
<td>IV</td>
<td>Solid, liquid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Moxifloxacin</td>
<td>IV</td>
<td>Liquid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Gatifloxacin</td>
<td>IV</td>
<td>Solid</td>
<td>-</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Oral bacteriostatic second-line anti-TB agents</td>
<td>Solid, liquid</td>
<td></td>
<td>Löwenstein-Jensen&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H11&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Ethionamide</td>
<td>IV</td>
<td>Solid, liquid</td>
<td>40.0</td>
<td>5.0</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>Prothionamide</td>
<td>IV</td>
<td>Solid, liquid</td>
<td>40.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cycloserine</td>
<td>IV</td>
<td>Solid</td>
<td>40.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Terizidone</td>
<td>IV</td>
<td>None</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>P-aminoosalicylic acid</td>
<td>IV</td>
<td>Solid, liquid</td>
<td>1.0</td>
<td>2.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>Thioacetazone</td>
<td>V</td>
<td>None</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 5&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Antituberculosis agents with unclear efficacy (not recommended by WHO for routine use in MDR-TB patients)</td>
<td>Solid, liquid</td>
<td></td>
<td>Löwenstein-Jensen&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Middlebrook 7H11&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Ciprofloxacin</td>
<td>V</td>
<td>Liquid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Clindamycin/Clavulanate</td>
<td>V</td>
<td>None</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Clarithromycin</td>
<td>V</td>
<td>None</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Linezolid</td>
<td>V</td>
<td>Liquid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup> WHO Guidelines for the programmatic management of drug-resistant tuberculosis (5).

<sup>b</sup> Indirect proportion method recommended. Other solid media methods (resistance ratio, absolute concentration) have not been adequately validated for second-line drugs.

<sup>c</sup> Routine DST for group 4 and 5 drugs is not recommended.

<sup>d</sup> Ciprofloxacin is no longer recommended to treat drug-susceptible or drug-resistant TB (5).

<sup>e</sup> Gatifloxacin only to be used in exceptional circumstances (5).
Rational use of DST in programmes for control of drug-resistant TB

The WHO *Guidelines for the programmatic management of drug-resistant tuberculosis* (5) call for rational strategies for case-finding and treatment of drug-resistant TB patients. Access to quality-assured DST is a prerequisite, with representative data from drug-resistance surveillance or surveys as a minimum to guide the design of treatment modalities.

The following considerations are relevant to the laboratory component of MDR-TB programmes.

**Rapid rifampicin-resistance testing**

In most settings, particularly where fixed-dose combination (FDC) first-line anti-TB drugs are used, resistance to rifampicin is almost invariably associated with resistance to isoniazid. Detection of rifampicin resistance therefore serves as a reliable (although not complete) proxy of MDR-TB. The advantages of rapid rifampicin testing include earlier identification of patients on inappropriate first-line regimens, prompt screening of patients at risk of MDR-TB, and early interruption of MDR-TB transmission.

Several tests for rapid detection of rifampicin resistance have been validated in laboratory-based studies and are currently being evaluated for feasibility, cost-effectiveness and cost-benefit in large-scale field demonstration studies. The use of line-probe assays for rapid detection of rifampicin resistance has recently been endorsed by a WHO expert group convened to evaluate the evidence base (WHO Expert Group on molecular line-probe assays for rapid screening of patients at risk of MDR-TB, communication, 31 March 2008).

The use of rapid rifampicin-resistance testing is recommended in high-risk MDR-TB settings (including high-burden HIV settings); however, confirmation of MDR-TB by conventional DST is still regarded as the gold standard, and adequate laboratory capacity to ensure a quality-assured diagnosis of MDR-TB remains a fundamental requirement. As no rapid tests for the diagnosis of XDR-TB are yet available, adequate laboratory capacity for second-line DST also remains a fundamental need.

**Organization and funding of the laboratory network**

TB laboratory networks conventionally have a pyramidal structure based on an appropriate number of peripheral (level I) laboratories capable of sputum-smear microscopy, an appropriate number of intermediate (level II) laboratories capable of mycobacteriological culture, and a single (or more than one in large countries) national (level III) laboratory capable of DST (15).

Although diagnostic and treatment programmes for drug-resistant TB are based on a range of strategies, in vitro DST plays a key role in all of these. As a minimum, laboratory capacity to reliably detect isoniazid and rifampicin resistance is a prerequisite. Routine DST for second-line drugs is *not* recommended unless the
required laboratory infrastructure and capacity have been established, rigorous quality assurance is in place, and sustainable proficiency has been demonstrated.

Proficiency in DST is a combination of laboratory technique and workload. In order to retain proficiency and expertise, it is recommended that second-line DST only be performed if at least 200 specimens from high-risk patients per year are expected. This implies centralization of laboratory services for second-line DST or outsourcing such services (e.g. to one of the laboratories in the SRL Network) where programmes involve small numbers of MDR-TB patients.

Adequate resource allocation (human and financial) to laboratory services is essential. For mycobacteriological culture and DST, one of the most neglected areas is currently the availability of sufficient, adequately qualified and trained laboratory staff. Funding should also be adequate to ensure a safe and functioning laboratory infrastructure with appropriate and well-maintained equipment and sufficient laboratory consumables.

Transport of infectious substances

Specimens from patients suspected of having MDR-TB as well as cultures of \textit{M. tuberculosis} pose a significant public health risk if not properly transported. Cultures in particular constitute enriched infectious material containing large numbers of viable organisms, and the risk is compounded when cultures of resistant strains are transported.

International organizations such as the Universal Postal Union (UPU), the International Civil Aviation Organization (ICAO) and the International Air Transport Association (IATA) have developed strict guidelines and procedures to facilitate safe and expeditious shipment of infectious substances while at the same time ensuring public safety. These are guided by the United Nations \textit{Model regulations on the transport of dangerous goods} (16). IATA issues \textit{Infectious substances shipping guidelines} (17) regularly and imposes additional restrictions as necessary. These guidelines must be followed if a shipment is carried by members of IATA.

Exchange of \textit{M. tuberculosis} cultures between countries (e.g. for diagnostic DST, retesting or proficiency testing) is always subject to international regulations, which include national import and export regulations specific to individual countries.

Infectious substances and diagnostic specimens likely to contain infectious substances require triple packaging in accordance with the United Nations \textit{Model regulations} (16). Cultures of \textit{M. tuberculosis} should be shipped in screw-capped tubes as primary watertight containers, enclosed in a second watertight container and finally enclosed in an appropriate outer shipping container. Petri-dish cultures and large volumes of liquid cultures must not be shipped.

Compliance with shipment requirements is the responsibility of the shipper, who must be familiar with national and international regulations.

Hand-carriage of infectious substances is strictly prohibited by international air carriers, as is the use of diplomatic pouches (17).
Surveillance and surveys using DST

Surveillance of anti-TB drug resistance is essential for providing information on the magnitude and trends in drug resistance, for developing appropriate treatment guidelines and for evaluating the impact of control programme interventions. In addition, survey or surveillance data on second-line drug resistance are recommended in designing appropriate treatment modalities.

Screening all MDR-TB strains for second-line drug resistance is recommended as a minimum (18) using the hierarchy of DST recommended below. Because of the reliability and reproducibility of DST for aminoglycosides, polypeptides and fluoroquinolones, and since resistance to these drugs defines XDR, DST of these second-line drugs constitutes a priority for surveillance and surveys, based on the history of specific drugs within the different categories used in a country. Second-line DST for all drugs in categories I and II (Table 2) is, however, very useful during surveillance and surveys to establish a baseline and inform treatment decisions, even if specific drugs in the different categories may not have been used in the past.

Hierarchy of DST under programmatic conditions

Programmes for management of drug-resistant TB should decide which drugs to test for resistance, based on the most appropriate strategy for designing treatment regimens and taking the above-mentioned constraints of SLD-DST into account. It is also recommended that DST be limited to those drugs used in individual country treatment strategies, i.e. there is little point in establishing DST capacity for second-line drugs that are unavailable or not recommended by country-specific programmes.

Box 1 provides an outline for systematic DST of first- and second-line anti-TB drugs under routine programmatic conditions.
Box 1 Systematic approach to implementation of DST under routine programmatic conditions

<table>
<thead>
<tr>
<th>Step 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoniazid</td>
</tr>
<tr>
<td>Rifampicin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethambutol</td>
</tr>
<tr>
<td>Streptomycin</td>
</tr>
<tr>
<td>Pyrazinamide</td>
</tr>
</tbody>
</table>

Steps 1 and 2 may be merged if indicated by epidemiological considerations and/or treatment modalities (e.g. standardized or individualized MDR-TB regimens still involving first-line drugs) and if resources allow extended DST capacity.

<table>
<thead>
<tr>
<th>Step 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amikacin, kanamycin, capreomycin</td>
</tr>
<tr>
<td>Ofloxacin (or fluoroquinolone of choice in treatment strategy)</td>
</tr>
</tbody>
</table>

Steps 1 and 3 may be merged in settings where XDR is a concern in order to enable the rapid identification of XDR-TB patients.

Given the variability in cross-resistance reported for the aminoglycosides, it is recommended that all aminoglycosides (including streptomycin) as well as capreomycin be tested for resistance where possible.

Selection of the most appropriate fluoroquinolone for use in treatment modalities should be based on a representative survey or surveillance data. Only one fluoroquinolone needs to be tested owing to extensive cross-resistance.

Routine DST for group 4 drugs (ethionamide, prothionamide, cycloserine, terizidone, P-aminosalicylic acid and for group 5 drugs (clofazimine, amoxicillin-clavulanate, clarithromycin, linezolid) is not recommended.

Biosafety and laboratory infection control

Transmission of TB – including drug-resistant TB – is a recognized occupational risk for laboratory workers. Adequate biosafety measures and prevention of laboratory-acquired infection are therefore paramount.

The relative hazards of infective microorganisms handled in the laboratory are classified by WHO according to the risk of causing human disease, the potential for laboratory spread and whether effective treatment and prevention measures are available. Related biosafety levels for laboratories have been defined, taking into account the pathogenic agent, the facilities available, and the equipment, practices and procedures required to ensure a safe laboratory working environment (19).
Irrespective of the risk, laboratory standards require that:

- appropriate and specific administrative controls (including good laboratory practice, standard operating procedures and accident management plans) are in place and enforced;
- appropriate engineering controls are being used and are functioning adequately as designed;
- personal protective equipment is appropriate for the tasks being performed;
- proper waste management procedures are in place;
- proper procedures for general laboratory safety (including physical, electrical and chemical safety) are in place and enforced.

*M. tuberculosis* is classified by WHO as a risk group 3 laboratory pathogen, causing serious human disease but with limited possibilities of (laboratory) spread and with effective treatment and preventive measures available (19). Mycobacteriological culture and DST procedures generate high concentrations of organisms that pose an increased risk of aerosol spread. Given the limited treatment options for MDR-TB (and even more so for XDR-TB), laboratory procedures for culture and DST therefore necessitate special containment through biosafety level 3 precautions.

Biosafety level 3 containment requires the strengthening of laboratory operations and safety programmes, specifically those related to laboratory design, use of specialized equipment to prevent or contain aerosols, and health surveillance of laboratory staff. Published guidelines on biosafety level 3 precautions should be rigorously followed (19) and expert engineering consultation sought when establishing laboratory infrastructure for second-line DST, taking the following essential aspects into consideration:

- Access to the containment laboratory must be restricted, preferably through an anteroom (i.e. a specific area designed to maintain the pressure differential between the laboratory and the adjacent space).
- A controlled ventilation system must be installed which maintains a directional airflow into the laboratory, supported by a visual monitoring device showing that proper directional airflow is maintained at all times.
- The building ventilation system must be constructed so that air from the containment laboratory is not recirculated to other areas within the building. This is usually achieved through high-efficiency particulate air (HEPA) filtration, which needs specialized engineering input, validation and maintenance to ensure compliance with safety standards.
- Essential equipment for mycobacteriological culture and DST includes appropriate and well-maintained biological safety cabinets, centrifuges and other safety equipment to meet biosafety level 3 precaution standards.
• Health and medical surveillance of laboratory personnel involved in mycobacteriological culture and DST is strongly recommended. Surveillance should include a detailed medical history, targeted baseline health assessment, monitoring of respiratory signs and symptoms, and a proactive plan for appropriate medical investigations when indicated.

Quality assurance

A diagnosis of MDR-TB or XDR-TB has profound implications for the individual patient. Accuracy of the laboratory diagnosis is therefore crucial and a comprehensive laboratory quality assurance programme must be in place to ensure the accuracy, reliability and reproducibility of DST results.

Central reference laboratories involved in programmes for drug-resistant TB should establish formal links with one of the laboratories in the Supranational Reference Laboratory Network (SRLN) to help ensure the quality of laboratory services and validation of DST results. The SRLN currently consists of 26 laboratories, including a global coordinating centre in Belgium.

The SRLN ensures DST standards by a system of external quality assurance that should preferably be established prior to the implementation of MDR-TB programmes. As a minimum, external quality assurance with an SRL should consist of the following:

• an initial assessment visit;
• proficiency testing with an adequate number of coded isolates;
• periodic rechecking of isolates obtained within the MDR-TB programme.

Proficiency testing by the SRLN involves regular distribution to national reference laboratories of panels of coded *M. tuberculosis* strains with predefined drug-resistance profiles. The test results of the reference laboratory are compared with the coded SRL results in blinded fashion and specific performance indicators (sensitivity, specificity, reproducibility) calculated for each drug and for the reference laboratory as a whole.

As a minimum performance indicator, proficiency testing should identify correctly resistance to isoniazid and rifampicin in more than 90% in two out of three recent rounds of panels.

The SRLN is in agreement that panels for second-line proficiency testing should not include XDR strains of *M. tuberculosis*. Rather, panels with different permutations of mono resistance to second-line drugs are currently being developed that will be compiled to allow reliable assessment of the overall capability of national reference laboratories to identify XDR-TB. Panels including isolates with second-line drug resistance will be made available through the SRLN in 2008.
Conclusions

Quality-assured laboratory services constitute the backbone of programmes for drug-resistant TB. Implementation of such programmes necessitate that governments and donors adequately fund appropriate and safe laboratory infrastructures in which well-trained staff working to clear standard operating procedures are able to deliver accurate and timely drug-resistance results. The need remains to improve DST for second-line drugs and to configure screening and diagnostic algorithms into rational management programmes for drug-resistant TB. In addition, accelerated expansion and integration of laboratory services as a core component of TB control programmes is required to achieve the aims of the global MDR-TB response and maximize the potential of new technological developments.
References


Use of interferon-gamma release assays in support of TB diagnosis
Use of interferon-gamma release assays in support of TB diagnosis

Ad hoc scientific panel opinion
Contents

Executive summary ........................................................................................................................................ 2
1 Background ................................................................................................................................................ 4
  1.1 Current situation .................................................................................................................................... 4
  1.2 Objectives ............................................................................................................................................. 4
  1.3 Document background ........................................................................................................................... 5
  1.4 Document format ................................................................................................................................... 6
2 Background information on IGRAs ................................................................................................................ 7
  2.1 What are interferon-gamma release assays (IGRAs)? ................................................................................ 7
  2.2 Which IGRAs are available? .................................................................................................................... 7
  2.3 How are IGRAs performed? .................................................................................................................... 7
  2.4 What are the advantages and disadvantages of IGRAs? ............................................................................ 8
3 Panel opinions and summary of evidence .................................................................................................... 10
  3.1 Is there a role of IGRAs in the diagnosis of active TB? ............................................................................. 10
    3.1.1 Can IGRAs be used as a stand-alone tool to diagnose active TB disease? ........................................... 10
    3.1.2 Can IGRAs be used to support the diagnosis of active TB disease? ...................................................... 10
  3.2 Is there a role for IGRAs in the diagnosis of latent TB infection? .............................................................. 14
    3.2.1 What is the value of IGRA tests in identifying individuals for preventive treatment? .............................. 15
    3.2.2 Can IGRAs differentiate LTBI from active TB? ................................................................................... 17
    3.2.3 How should IGRAs be used in different population groups and settings? .............................................. 18
4 Future research needs and considerations ................................................................................................... 27
  4.1 Can IGRAs be used with extrasanguinous fluids to support the diagnosis of active TB? .............................. 27
  4.2 Large-scale population screening for LTBI .............................................................................................. 28
  4.3 Future research needs .......................................................................................................................... 29
References .................................................................................................................................................. 30
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCG</td>
<td>Bacillus Calmette-Guérin</td>
</tr>
<tr>
<td>CFP-10</td>
<td>Culture filtrate protein 10</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>ECDC</td>
<td>European Centre for Disease Prevention and Control</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>ELISPOT</td>
<td>Enzyme-linked immunosorbent spot</td>
</tr>
<tr>
<td>ESAT-6</td>
<td>Early secretory antigenic target-6</td>
</tr>
<tr>
<td>HAART</td>
<td>Highly active antiretroviral therapy</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>Interferon-gamma</td>
</tr>
<tr>
<td>IGRA</td>
<td>Interferon-gamma release assay</td>
</tr>
<tr>
<td>LTBI</td>
<td>Latent TB infection</td>
</tr>
<tr>
<td>MS</td>
<td>Member State</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative predictive value</td>
</tr>
<tr>
<td>NTM</td>
<td>Nontuberculous mycobacteria</td>
</tr>
<tr>
<td>NTP</td>
<td>National TB programme</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>PHA</td>
<td>Phytohemagglutinin A</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive predictive value</td>
</tr>
<tr>
<td>QFT-GIT</td>
<td>QuantiFERON-TB Gold In-Tube</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>TBNET</td>
<td>TB Network European Trials group</td>
</tr>
<tr>
<td>TST</td>
<td>Tuberculin skin test</td>
</tr>
</tbody>
</table>
Summary and background

This guidance document presents the evidence-based expert opinion of an ad hoc scientific panel on the use of the interferon-gamma release assays (IGRAs) for the diagnosis of latent tuberculosis (TB) infection and active TB. The panel expressed that IGRAs should not replace the existing standard diagnostic methods for the diagnosis of active TB and that a negative IGRA result does not exclude active TB disease. As to the diagnosis of LTBI, the panel expressed that IGRAs may be used in conjunction with an overall risk assessment in order to identify individuals for whom preventive treatment should be considered. Further opinions on the use of IGRAs in specific risk groups and populations are presented in the document.

IGRAs are relatively new diagnostic tools for TB. They were developed to support the diagnosis of latent TB infection but research is also ongoing for expanding its use to the diagnosis of active tuberculosis.

To assure optimal tuberculosis prevention and control it is essential that the introduction of new tools into national TB programmes and/or tuberculosis control strategies is based on solid scientific evidence. Uncertainty still remains as to the effectiveness of IGRAs. ECDC therefore identified the need to provide an EU-adapted guidance on the applicability and appropriateness of IGRAs, both for the identification of latent infection and active TB disease.
Executive summary

New tools to diagnose latent TB infection (LTBI) and active tuberculosis (TB) are needed. LTBI is commonly diagnosed by using the tuberculin skin test (TST). More recently, interferon-gamma release assays (IGRAs) were introduced for the diagnosis of LTBI.

IGRAs are blood-based tests that essentially measure the presence of specific *M. tuberculosis* reactive T-cells sensitised by a previous infection with *M. tuberculosis*. Two commercial IGRAs are available, the QuantiFERON-TB Gold In-Tube assay (QFT-GIT) (Cellestis Ltd., Australia) and the T-SPOT-TB (Oxford Immunotec, UK). Compared to the TST, IGRAs are not confounded by prior bacille Calmette-Guérin (BCG) vaccination and are less likely to be influenced by previous exposure to most nontuberculous mycobacteria (NTM) due to the target antigens selected to stimulate cellular immune responses.

European Union Member States are heterogeneous in terms of TB burden and characteristics of TB epidemiology, with intermediate-to-high (>20 per 100 000) and low (< 20 per 100 000) TB-incidence countries. This guidance document on the use of IGRAs in EU Member States, based on the most up-to-date scientific evidence available on the diagnosis of LTBI and active TB, was developed on the initiative of the European Centre for Disease Prevention and Control (ECDC).

The bulk of the evidence presented in this guidance document is based on two systematic reviews and meta-analyses assessing the role of IGRAs in the diagnosis of active TB and LTBI, conducted by the TB Network European trials group (TBNET) for and under the supervision of ECDC. Additional reviews and key studies covering other areas are also presented. ECDC originally developed the document as an FAQ list (frequently asked questions) and then presented to an ad hoc scientific panel of experts. The panel of experts was selected by ECDC's Chief Scientist and endorsed by the ECDC Director. Experts were selected based on their expertise in different areas of TB control. It was also ensured that panel members did not have a conflict of interest. During its meeting, the panel assessed the scientific evidence on IGRAs and expressed a unanimous opinion on the use of IGRAs in defined areas of TB control.

The opinion of the panel regarding the use of IGRAs as a stand-alone test for diagnosing active TB was as follows:

**Expert opinion**
Based on the evidence, IGRAs should not replace the standard diagnostic methods (including microbiology, molecular tests, and clinical and radiological assessment) for diagnosing active TB.

The opinion of the panel on the use of IGRAs to support the diagnosis of active TB disease was as follows:

**Expert opinion**
Based on the evidence, IGRAs do not have an added value in most clinical situations when combined with standard methods for diagnosing active TB.

However, based on limited evidence, in certain clinical situations (e.g. patients with extrapulmonary TB, patients who test negative for acid-fast bacilli in sputum and/or negative for *M. tuberculosis* on culture, TB diagnosis in children, or in the differential diagnosis of infection with NTM) IGRAs could contribute supplementary information as part of the diagnostic work-up.

Please note that a negative IGRA does not rule out active TB.

On the question of whether IGRAs have a role in diagnosing LTBI, specifically with the aim of identifying individuals for preventive therapy, the following opinion was stated:
The use of tools to diagnose LTBI must consider the accuracy of the test in specific risk groups (immunocompromised persons, children) and populations (high- or low-incidence settings, vaccination status). This document also presents the opinion of the panel on the use of IGRAs to diagnose LTBI in various populations and risk groups.

The aim of this EU-adapted guidance on the use of IGRAS for the diagnosis of LTBI and active TB is to present the most up-to-date evidence and expert opinion regarding IGRAs, providing Member States with support when considering the introduction of IGRAs to national TB programmes and/or tuberculosis control strategies.

Expert opinion

Based on the available results on positive predictive value (PPV) for progression, and taking into consideration the low statistical power and low number of studies, IGRAs may be used as part of the overall risk assessment to identify individuals for preventive treatment (e.g. immunocompromised persons, children, close contacts, and recently-exposed individuals).

Similarly, despite the limitations of available studies, the high NPV for progression of IGRAs indicates that at the time of testing and in the context of an overall risk assessment, progression to active TB in healthy immunocompetent individuals with negative IGRAs is very unlikely. Therefore, IGRAs may be used in this context.

Please note that, especially in risk groups and specific situations, a negative IGRA does not rule out LTBI. See Section 3.2.3 ‘How should IGRAs be used in different population groups and settings?’.
1 Background

1.1 Current situation

There is currently a need to identify and improve new tools for the diagnosis of latent TB infection (LTBI) and active tuberculosis (TB). LTBI is recognised as a complex clinical condition, in which the exact biological status of the TB bacilli is not fully understood. When a person is infected with Mycobacterium tuberculosis (M. tuberculosis), the bacilli are thought to persist in a subclinical status with minimal replication, a status in which the bacteria are unable to cause manifest clinical disease. Upon a shift in an individual’s immunologic status, M. tuberculosis is able to begin replicating and multiply to a number that causes disease, manifesting as active TB. LTBI is therefore defined as a clinical condition in an individual suspected of being infected, but having no manifestation of disease, and from whom M. tuberculosis bacilli cannot be identified through culture.

Active TB is diagnosed by evaluating an individual’s medical history, clinical symptoms, (chest) radiography, as well as the microbiologic and molecular identification of M. tuberculosis (through the detection of acid-fast bacilli in sputum, M. tuberculosis culture, and nucleic acid amplification). The diagnosis of active TB can often be challenging, with results remaining inconclusive (e.g. acid-fast bacilli sputum smear-negative), especially in specific risk groups, such as children and immunocompromised individuals. New useful, sensitive, and rapid tools to detect active TB are clearly needed.

Identifying LTBI aims at identifying individuals who would benefit from treatment, preventing future development of active TB disease. This in itself is an important reason for controlling the transmission of disease within a population, as it decreases the number of active TB cases that have the potential of transmitting the infection. The challenge of identifying LTBI-infected individuals lies in the lack of a diagnostic gold standard for LTBI.

There are currently two diagnostic methods that support the diagnosis of LTBI: the tuberculin skin test (TST) and interferon-gamma release assays (IGRAs). Both tests are immunological methods that detect an immune response to antigens and consequently do not allow a direct measure of persistent infection. The *in vivo* TST is based on the intracutaneous injection of M. tuberculosis antigens and subsequent identification of an immune reaction at the site of injection. A limitation of the TST is that the complex mixture of different antigens used are not specific for M. tuberculosis, and therefore local immunologic activity at the site of the antigen deposition does not differentiate between an existing immune response elicited by either, previous bacille Calmette-Guérin (BCG) vaccination, exposure to NTM, or M. tuberculosis infection. IGRAs are more recent *in vitro* assays that detect the presence of cellular immune responses towards M. tuberculosis-specific antigens. These include the early secretory antigenic target-6 (ESAT-6), culture filtrate protein 10 (CFP-10), and the TB7.7 antigens. In contrast to the TST, the antigens in IGRAs are absent in most of NTM (with the exception of M. flavesens, M. marinum, M. kansasi and M. szulgai), as well as from BCG strains. Although IGRAs cannot distinguish between active TB and LTBI, IGRA-results are not confounded by BCG vaccination and less likely to be confounded by exposure to NTM.

European Union Member States are heterogeneous in terms of TB burden and characteristics of TB epidemiology, with intermediate-to-high (>20 per 100 000), and low TB incidence countries (< 20 per 100 000). As for the application of TB diagnostic tools, a country’s TB epidemiology will influence how TB control programmes consider the use of IGRAs. A number of countries have already introduced this diagnostic tool in their national TB programmes and it has been the European Centre for Disease Prevention and Control’s experience that an increasing number of countries are currently considering the implementation of IGRAs and, as a result, requesting guidance and support. In 2007, European experts developed a consensus statement on the use of IGRAs based on the scientific evidence available at the time. Since then, substantially more studies have evaluated IGRAs and more countries have been considering the use of assays.

ECDC therefore identified the need to provide an EU-adapted guidance on the applicability and appropriateness of IGRAs, both for the identification of latent infection and active TB disease in order to support the Member States as they consider the introduction of IGRA’s in national TB programmes and/or tuberculosis control strategies.

It should also be noted that the introduction of new diagnostic tools, including IGRAs, requires the adjustment of policies and programmes to assure that the tools are properly adopted, introduced and implemented. This is beyond the scope of this document; however, resources describing all aspects of adoption and implementation are readily available. One example is the detailed framework on adoption, introduction and implementation of new tools for TB control developed by the Stop TB Partnership Retooling Taskforce.

1.2 Objectives

The aim of this guidance document is to present the most recent scientific evidence and expert opinion on the use of IGRAs for the diagnosis of LTBI as well as their applicability to the diagnosis of active TB. It presents several
aspects to consider when implementing IGRAs in national TB programmes, including the accuracy of the assays, their application within different patient groups and/or TB incidence settings, and future research needs.

The ultimate goal of this document is to present the most up-to-date evidence and subsequent expert opinion regarding IGRAs, in order to provide the Member States with support when considering the introduction of IGRAs in national TB programmes and/or tuberculosis control strategies.

1.3 Document background

Following ECDC’s procedures for developing evidence-based guidance, the ECDC convened an ad hoc scientific panel of experts in order to assess the most up-to-date scientific evidence on IGRAs, and subsequently express a unanimous opinion on the use of IGRAs in defined areas of TB control. Panel members were identified by ECDC’s Chief Scientist and endorsed by the Director of ECDC, based on their expertise in different areas of TB control, as well as their documented lack of conflicts of interest (see Table 1). All members signed a Declaration of Interest, which was reviewed by the Chief Scientist who confirmed that no member of the panel had a conflict of interest in regard to the topic of discussion. During the work of the panel, one member developed a conflict of interest: the expert’s status was changed to ‘observer’, and the expert was excluded from contributing to the conclusions of the panel.

The panel was independent from ECDC, which organised, hosted and observed the panel meeting.

Table 1: Members of the ad hoc scientific panel and observers

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markus Maeurer, Chair</td>
<td>Stockholm, Sweden</td>
<td>Karolinska Institutet and Karolinska Hospital</td>
</tr>
<tr>
<td>Gernot Rohde</td>
<td>Maastricht, Netherlands</td>
<td>Maastricht University Medical Centre</td>
</tr>
<tr>
<td>Andrew Ramsay</td>
<td>Geneva, Switzerland</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>Ibrahim Abubakar</td>
<td>London, United Kingdom</td>
<td>Health Protection Agency, Centre for Infections</td>
</tr>
<tr>
<td>Connie Erkens</td>
<td>The Hague, Netherlands</td>
<td>KNCV TB Foundation</td>
</tr>
<tr>
<td>Vera Katalinić-Janković</td>
<td>Zagreb, Croatia</td>
<td>European Reference Laboratory Network for TB (ERLN-TB)/Croatian National Institute of Public Health</td>
</tr>
<tr>
<td>Walter Haas</td>
<td>Berlin, Germany</td>
<td>Robert Koch Institute</td>
</tr>
<tr>
<td>Hans Gaines</td>
<td>Stockholm, Sweden</td>
<td>Swedish Institute for Infectious Disease Control</td>
</tr>
<tr>
<td>Anne Detjen</td>
<td>New York City, USA</td>
<td>International Union Against TB and Lung Disease</td>
</tr>
<tr>
<td>Francesco Blasi</td>
<td>Milan, Italy</td>
<td>Ospedale Maggiore (General Hospital), Milan</td>
</tr>
<tr>
<td>Christoph Lange*</td>
<td>Borstel, Germany</td>
<td>Research Centre Borstel/TBNET</td>
</tr>
<tr>
<td>Giovanni Sotgiu*</td>
<td>Sassari, Italy</td>
<td>University of Sassari/TBNET</td>
</tr>
<tr>
<td>Shreemanta Parida*</td>
<td>Berlin, Germany</td>
<td>(Former affiliation: Max Planck Institute for Infection Biology)</td>
</tr>
<tr>
<td>Elisabeth Whittaker*</td>
<td>London, United Kingdom</td>
<td>Imperial College London</td>
</tr>
</tbody>
</table>

* Observer

The panel’s task was to express an expert opinion on key questions referring to the accuracy of IGRAs and their applicability to national TB programmes. The evidence presented was based on systematic reviews and meta-analyses of the literature collated by ECDC. Where such studies were not available, key studies were presented. Summaries of the results and conclusions of these studies were developed by ECDC, using a ‘frequently asked questions’ format, and passed on to the panel before the meeting.

The scientific evidence on the accuracy of IGRAs formed the basis for the development of this guidance document. Two systematic reviews and a meta-analyses were conducted to assess the accuracy of IGRAs in diagnosing LTBI and active TB disease. The overall aims of these reviews were to:

- evaluate the evidence on the accuracy of using IGRAs in order to diagnose active TB1; and
- evaluate the evidence on the accuracy/usability of IGRAs for identifying LTBI and its comparative disadvantage/advantage to the TST2.

The objective of the first systematic review and the meta-analysis assessing IGRAs for the diagnosis of active TB was to compare the accuracy of QFT-GIT and T-SPOT.TB to TST in the diagnosis of active TB (in adults and in children) with blood samples, and to assess the accuracy of QFT-GIT and T-SPOT.TB with samples of extrasanguinous fluids in the diagnosis of active TB (in high-TB incidence and low-TB incidence settings)1.

The objective of the systematic review and meta-analysis assessing IGRAs in the diagnosis of LTBI was to compare the accuracy (specificity, positive predictive value for progression, negative predictive value, negative predictive
value for progression, and association with *M. tuberculosis* exposure and BCG vaccination) of QFT-GIT and T-SPOT.TB to TST in the diagnosis of LTBI.

These systematic reviews were conducted by the TB Network European Trials group (TBNET) for, and under the supervision of, ECDC through an open contract finalised in March 2010. The meta-analyses on the evidence from the two systematic reviews was developed by TBNET and ECDC and published in a peer-reviewed journal, the European Respiratory Journal. Representatives of the TBNET contractors were present as observers during the meeting to support the panel regarding the studies and results.

Also present at the meeting were two observers to support the panel in questions of TB in children and TB immunology.

Following the meeting, which was chaired by Professor Markus Maeurer, ECDC updated the document to include the panel's opinions as well as the considerations identified by the panel. The document was then sent to the panel members as well as the ECDC Advisory Forum for consultation and commenting.

### 1.4 Document format

The opinions of the *ad hoc* scientific panel are presented in Section 3 of this document. Preceding each opinion is a section summarising the panel's considerations, followed by an overview of the available evidence. As described above, the bulk of the evidence is based on the two TBNET/ECDC systematic reviews and the meta-analyses, which assess the accuracy of IGRAs in the diagnosis of active TB and LTBI (see Annex 1 and Annex 2). Where needed, the evidence has been complemented with other published meta-analyses or systematic reviews. When such studies were not available, key studies are presented.

The final section highlights future research needs and considerations regarding the use of IGRAs in the diagnosis of LTBI and active TB disease.

This document is based on the evidence and knowledge of IGRAs at the time of publication (early 2011). The current document will be updated when more evidence on IGRAs becomes available.
2 Background information on IGRAs

Upon infection with *M. tuberculosis*, the different subsets of immune cells (e.g. macrophages, T-cells) involved in the immune response directed against the bacilli do not fully eradicate the bacilli, but rather contain the infection. Macrophages play an important role in the first line of defence against pathogen-infection through their ability to ingest and subsequently kill pathogens. However, having developed immune escape mechanisms, *M. tuberculosis* bacilli have the ability to persist within macrophages, averting the attack by these host cells.

The cytokine interferon-gamma (IFN-γ) is produced by different cells of the immune system: CD4 T-cells, CD8 T-cells and Natural Killer cells. This cytokine is considered to play an important role in the elimination of *M. tuberculosis* by activating the production of reactive oxygen and nitrogen intermediates in macrophages, which in turn are involved in the destruction of bacterial pathogens. T-cells specifically recognizing *M. tuberculosis* antigens, particularly CD4 T-cells, produce IFN-γ essential for the activation of *M. tuberculosis*-infected macrophages which, upon activation, can target *M. tuberculosis* bacilli and control their growth.

In summary, infection with *M. tuberculosis* triggers a complex immune response that, in most individuals, leads to the containment of the infection and the establishment of a pool of long-lasting memory T-cells specifically directed against *M. tuberculosis* antigens.

2.1 What are interferon-gamma release assays (IGRAs)?

IGRAs are blood-based tests assessing the presence of effector and memory immune responses directed against the *M. tuberculosis* antigens ESAT-6, CFP-10 and, in one of the available tests, the TB7.7 antigen. The IGRAs have been shown to predominantly measure the presence of *M. tuberculosis*-specific effector memory T-cells, the presence of which are considered indicative of previous in vivo exposure to the bacilli. Blood samples might also contain central-memory T-cells specific to the *M. tuberculosis* antigens and thus be measured in the assays. The latter is however seen as less likely, as this subset of cells react more slowly to antigen exposure and is considered to first release other cytokines during the time-span of the assays (e.g. interleukin-2).

The IGRAs measure the presence of an adaptive immune response to *M. tuberculosis* antigens, and are thus only an indirect measure of *M. tuberculosis* exposure (evidence is still lacking as to whether an immune response corresponds to actual infection). IGRAs have been developed for the identification of an immune response to *M. tuberculosis*-specific antigens, considered to be a correlate of *M. tuberculosis* infection, and are licensed for the use on blood specimens.

2.2 Which IGRAs are available?

There are two commercially available IGRAs:

**QuantiFERON-TB Gold In-Tube assay (Cellestis Ltd, Australia)**

The QuantiFERON-TB Gold In-Tube assay (QFT-GIT), which has replaced the QuantiFERON-TB Gold assay, detects the level of IFN-γ produced in response to the *M. tuberculosis* antigens ESAT-6, CFP-10, and TB7.7, and uses the enzyme-linked immunosorbent assay (ELISA) detection method. This is an indirect measure of the presence of *M. tuberculosis* specific T-cells.

**T-SPOT.TB assay (Oxford Immunotec, UK)**

The T-SPOT.TB measures the number of IFN-γ producing T-cells in response to the *M. tuberculosis* antigens ESAT-6 and CFP-10, and is based on the enzyme-linked immunosorbent spot (ELISPOT) assay.

Only the two commercially available IGRAs (QFT-GIT and T-SPOT.TB, both with standardised and licensed protocols that allow the comparison of study results) are considered in this guidance document.

2.3 How are IGRAs performed?

IGRAs are performed on fresh blood specimens. The QFT-GIT is performed by drawing 1 ml of blood into one of each of the three manufacturer-precoated, heparinised tubes. Within 16 hours of blood collection, the tubes must be incubated for another 16 to 24 hours at 37 °C. The plasma is harvested after centrifugation (QFT-GIT collection tubes contain a gel plug that separates the plasma from the cells when centrifuged) and used (immediately, or later, provided there is adequate storage) to assess the concentration of IFN-γ by ELISA test. Results are interpreted according to the manufacturer’s recommendations (Table 1).
Table 1. Interpretation criteria for the QuantiFERON-TB Gold In-Tube assay (QFT-GIT)²⁰

<table>
<thead>
<tr>
<th>Result</th>
<th>IFN-γ concentration (International Units per ml, IU/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M. tuberculosis antigens</td>
</tr>
<tr>
<td>Positive</td>
<td>≥ 0.35 IU/ml and ≥ 25% over nil</td>
</tr>
<tr>
<td>Negative</td>
<td>&lt; 0.35 IU/ml or &lt; 25% over nil</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>&lt; 0.35 IU/ml or &lt; 25% over nil</td>
</tr>
<tr>
<td>Any</td>
<td>&gt; 8.0 IU/ml</td>
</tr>
</tbody>
</table>

M. tuberculosis antigens: mixture of peptides representing the entire amino acid sequences of ESAT-6 and CFP-10, and partially TB7.7; negative control (i.e. nil), positive control (phytohemagglutinin A, PHA).

For the T-SPOT.TB assay, 8 ml of blood are required and the assay must be performed within eight hours of blood collection (using, for example, heparinised tubes). Alternatively, the manufacturer also provides a reagent (T-Cell Xtend) which extends processing time to 32 hours after blood collection²¹. The T-cell-containing peripheral blood mononuclear cell (PBMC) fraction is separated from whole blood and distributed to the microtitre plate wells (250,000 cells/well) provided in the T-SPOT.TB assay kit. Following 16 to 20 hours (at 37°C with 5% CO² incubation), the number of IFN-γ-secreting T-cells (represented as spot-forming units) can be detected by ELISPOT assay. Results are interpreted according to the manufacturer’s recommendations (Table 2)²¹.

Table 2. Interpretation criteria for the T-SPOT.TB assay²¹

<table>
<thead>
<tr>
<th>Result</th>
<th>Spot count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M. tuberculosis antigens</td>
</tr>
<tr>
<td>ESAT-6</td>
<td>≥ 6 over nil and/or ≥ 6 over nil</td>
</tr>
<tr>
<td>CFP-10</td>
<td>≤ 5 over nil and/or ≤ 5 over nil</td>
</tr>
<tr>
<td>Borderline*</td>
<td>If for any antigen highest is 5 - 7 over nil</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>≤ 6 over nil and ≤ 6 over nil</td>
</tr>
<tr>
<td>Any</td>
<td>Any</td>
</tr>
</tbody>
</table>

* Retesting of patient is recommended

The presence of negative and positive controls (both in the QFT-GIT and the T-SPOT.TB tests) ensures that IGRAs are correctly performed. The negative control (no stimulation with the M. tuberculosis antigens, inducing no IFN-γ production) in IGRAs assesses the baseline level of IFN-γ present in the sample. The positive control (PHA, a T-cell-activating mitogen) in IGRAs assesses the performance of the test by measuring the ability of T-cells to produce IFN-γ, which may be impaired in immunocompromised patients.

Indeterminate results such as a high background detected in the negative control tube, or low responses in the positive-control tube (PHA), may be explained by technical factors (e.g. inappropriate storage of blood). Indeterminate results may also be explained by the immune status of the individual being tested. For instance, individuals with an impaired immune system (e.g. low T-cells numbers, decreased capacity to respond with IFN-γ production) might show such indeterminate results.

It is recommended that new blood samples are retested when a patient’s sample showed indeterminate results (or borderline results with T-SPOT.TB assay)²². If after retesting the results remain indeterminate, technical error may be ruled-out and T-cell anergy in the patient sample may be a possible explanation²³.

Please note that in the QFT-GIT assay, a standardised volume of blood, which will have a variable amount of cells depending on the sample, is tested, whereas in the T-SPOT.TB assay, a standardised number of cells are tested.

IGRAs were developed and licensed for use on blood. However, it is known that M. tuberculosis-specific T-cells are recruited at the site of infection, where their frequency is increased compared to peripheral blood. As a result, there is increased research activity on the applicability of IGRAs with extrasanguinous samples (e.g. pleural fluids, materials from bronchoalveolar lavage, ascitis, or liquor cerebrospinalis) (see Section 6.1).

2.4 What are the advantages and disadvantages of IGRAs?

As described above, IGRAs detect the presence of persistent cellular immune responses towards the M. tuberculosis-specific antigens ESAT-6, CFP-10 and TB7.7 (QFT-GIT), which are known to be absent in most of the NTM (except M. flavescens, M. marinum, M. kansasii and M. szulgai), as well as in BCG strains⁶⁻⁸. IGRAs cannot distinguish between active TB and LTBI, but results will be less likely to be confounded by an individuals’ previous exposure to NTM and will not be confounded by BCG vaccination. This feature in itself shows IGRAs advantage.
over the TST in identifying an individual’s true immune response to *M. tuberculosis*, especially in settings/populations with high NTM exposure and general BCG vaccination.

The advantages of IGRAs over the TST are several. Most importantly, individuals being tested are only required to present once to the healthcare facility (for drawing blood), increasing the likelihood that a final diagnosis is achieved. Furthermore, these *in vitro* assays have a rapid turn-around time and, being laboratory-based, follow standardised operational procedures. This decreases the effect of inter-personal variability when conducting the assays and aids in interpreting the results.

There are a number of disadvantages with IGRAs that should be considered when introducing them to national TB programmes and/or tuberculosis control strategies: IGRA testing requires drawing blood from individuals, and drawing a sufficient amount of blood from children is difficult. Also, IGRAs have to be conducted within a limited time frame. More specifically, the blood has to be tested in the laboratory within a given time-frame: 16 hours for the QFT-GIT, and eight hours for the T-SPOT.TB. The test incubation time also allows antigen-specific T-cells to react with the test antigens contained in the assay. This may trigger several immune functions, including T-cell proliferation and cytokine production, influencing the read-out of test results.

IGRAs have higher resource demands when compared to the TST, as they require laboratory access, trained personal, implementation of quality-assured procedures, and guaranteed continuous access to reagents.

As they are technically more demanding, IGRAs are a more costly diagnostic tool. However, the reading and analysis of test results can be done by batch (QFT-GIT: plasma can be frozen for later ELISA analysis), thus reducing the cost but also increasing the time until results are known. Furthermore, the higher specificity of IGRAs may decrease the number of false-positive test results when investigating LTBI, and therefore prevent further medical evaluations and treatment.

When used as diagnostic tools, two types of costs for IGRAs should be considered:

- direct costs of IGRAs; and
- overall costs, including direct and indirect costs of IGRAs.

**Direct costs of IGRAs**

In a cost-effectiveness analysis conducted by Pooran et al. (2010) on LTBI contact screening in the United Kingdom, IGRAs were substantially more expensive than TSTs:

- QFT-GIT testing (including test kit, consumables, processing, and phlebotomy): EUR 54 (GBP 45).
- T-SPOT.TB testing (including test kit, consumables, processing, and phlebotomy): EUR 66 (GBP 55).
- TST testing (including disposables, administration, and reading): EUR 19.30 (GBP 16.14).

**Direct and indirect costs of IGRAs**

A cost analysis by Linertova et al. assessing the costs of LTBI screening in Spanish healthcare workers in 2009 factored in the time spent on testing procedures as well as the hourly wages of healthcare workers. This study showed that the larger part of the cost of TST testing came from the time spent to perform the test and the reading of TST results, whereas the larger part of the cost of QuantiFERON-TB Gold (the predecessor of QFT-GIT) came from material and laboratory costs. The authors concluded that in the Spanish healthcare system the costs incurred by QuantiFERON-TB Gold and TST were similar when screening healthcare workers for LTBI.

Recent studies have assessed the cost effectiveness of IGRAs in the diagnosis of LTBI in contact screening in Germany and France, as well as healthcare worker screening in Israel. Although QFT-GIT was shown to be more expensive than TST, QFT-GIT led to fewer false-positive results, and thus consequently to fewer chest X-ray controls, fewer prescriptions of preventive treatment (in settings of high BCG-vaccination coverage), as well as fewer clinical visits. Taking into account adherence to preventive treatment, which in the study by Diel et al. was estimated to be as low as 24%, the studies concluded that QFT-GIT alone was more effective and cost-effective than TST alone.
3 Panel opinions and summary of evidence

For each question on the applicability of IGRAs, the ad hoc scientific panel assessed the presented evidence, identified a set of considerations, and expressed its opinion. The bulk of the evidence was based on two systematic reviews and meta-analyses conducted by TBNET for ECDC that assessed the accuracy of IGRAs in the diagnosis of active TB and LTBI (see Annex 1 and Annex 2)\textsuperscript{1,2}. Where needed, the evidence was complemented with other published systematic reviews and meta-analyses. When such studies were not available, key studies were used.

3.1 Is there a role of IGRAs in the diagnosis of active TB?

**Considerations**

- Following international standards, active TB is diagnosed by evaluating a patient's medical history, conducting a physical examination, chest radiography, and identifying M. tuberculosis bacilli using microbiologic and molecular diagnostic methods (sputum-smear microscopy, M. tuberculosis culture and nucleic acid amplification)\textsuperscript{4}.

- In some instances, the clinical diagnosis of active TB is difficult and results may be inconclusive (e.g. for patients with sputum-smear, acid-fast stain negative, and/or culture-negative results), despite extensive investigation of suspected TB. New useful, sensitive, and rapid tools to detect active TB are clearly needed\textsuperscript{4,29}. Such new tools for the direct detection of M. tuberculosis (or the corresponding genetic material) may be complemented by new indirect test methods, some of which use immunological approaches. The two latter approaches are not mutually exclusive.

- IGRAs have not been developed for the diagnosis of active TB. IGRAs identify the presence of an adaptive immune response (in peripheral blood) directed towards a defined set of M. tuberculosis antigens (ESAT-6, CFP-10 and, for QFT-GIT, TB7.7) and cannot differentiate between active and latent TB infection\textsuperscript{1,9}.

- In September 2010, the WHO Strategic and Technical Advisory Group for Tuberculosis (STAG-TB) reviewed the evidence and recommendations of an expert group on the ‘Use of commercial IGRAs in low-income and middle-income countries’ (typically high TB-settings and/or high HIV-burden settings), following WHO’s standard procedures for policy development. The STAG-TB endorsed the findings of the WHO expert group and supports the strategic approach to develop negative WHO policy recommendations to discourage the use of commercial IGRAs in low-income and middle-income countries (typically high-TB settings and/or high HIV-burden settings)\textsuperscript{30}. This decision was based on a large body of evidence showing the poor performance of current IGRAs and the risk of increased misdiagnosis, as well as the misplacement of resources in the diagnosis of active TB in low-income and middle-income settings (typically high-TB settings and/or high HIV-burden settings)\textsuperscript{30}.

3.1.1 Can IGRAs be used as a stand-alone tool to diagnose active TB disease?

**Expert opinion**

Based on the evidence, IGRAs should not replace the standard diagnostic methods (including microbiology, molecular tests, and clinical and radiological assessment) for diagnosing active TB.

For the evidence on the above expert opinion, please refer to Section 3.1.2 ‘Can IGRAs be used to support the diagnosis of active TB disease? – Evidence’.

Upon considering the evidence, the panel identified the need to express a separate opinion on the use of IGRAs as a stand-alone tool for the diagnosis of active TB disease.
3.1.2 Can IGRAs be used to support the diagnosis of active TB disease?

**Expert opinion**

Based on the evidence, in most clinical situations IGRAs do not have an added value when combined with standard methods for diagnosing active TB.

However, based on limited evidence, in certain clinical situations (e.g. patients with extrapulmonary TB, patients who test negative for acid-fast bacilli in sputum and/or negative for *M. tuberculosis* on culture, TB diagnosis in children, or in the differential diagnosis of infection with NTM) IGRAs could contribute supplementary information as part of the diagnostic work-up.

Please note that a negative IGRA does not rule out active TB.

**Evidence**

**General**

The TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of active TB included studies with specific data on sensitivity and specificity. Indeterminate results were excluded before sensitivity and specificity were calculated.

**Sensitivity**

Sensitivity measures the ability of a test to correctly identify individuals who have a certain disease. In the context of IGRAs and the diagnosis of active TB, sensitivity denotes the proportion of individuals with known active TB who test positive when IGRAs are used, i.e. the ability of IGRAs to correctly diagnose individuals with active TB and classify them as test-positive.

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of active TB, sensitivity was assessed in patients with clinical suspicion of TB disease (*M. tuberculosis* culture-confirmed and non-confirmed cases).

As listed in Table 4, the pooled sensitivity (95% CI) of QFT-GIT, T-SPOT.TB, and TST was: 80% (75-84%), 81% (78-84%) and 65% (61-68%), respectively.

**Table 3. Sensitivity of IGRAs and TST in the diagnosis of active TB in patients with clinical suspicion of TB disease**

<table>
<thead>
<tr>
<th></th>
<th>Pooled sensitivity (%)</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>80*</td>
<td>75-84</td>
<td>45.3</td>
<td>8</td>
<td>348</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>81**</td>
<td>78-84</td>
<td>93.3</td>
<td>15</td>
<td>749</td>
</tr>
<tr>
<td>TST</td>
<td>65***</td>
<td>61-68</td>
<td>89</td>
<td>12</td>
<td>703</td>
</tr>
</tbody>
</table>

* Pooled sensitivity was 81% (95% CI 78-84%; I²=0%) for patients with culture-confirmed TB.
** Pooled sensitivity was 92% (95% CI 90-93%; I²=78%) for patients with culture-confirmed TB.
*** Pooled sensitivity was 68% (95% CI 63-72%; I²=90%) for patients with culture-confirmed TB.

Based on the analysis, the authors of this meta-analysis concluded that the sensitivity of IGRAs was too low to support their use as rule-out tests for active TB.

**Specificity**

Specificity measures the ability of a test to correctly identify individuals who do not have the disease under investigation. In the context of IGRAs and the diagnosis of active TB, specificity denotes the proportion of individuals known not to have active TB and who test negative when the assay is used, i.e. the ability of IGRAs to correctly diagnose individuals who do not have active TB and classify them as test-negative.

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of TB, specificity was assessed in control groups who were considered to have a low risk of being infected with *M. tuberculosis*. Furthermore, studies including patients suspected of active TB (found to have an alternative disease and thus not TB) were also included in the analyses and calculations. Unlike low-risk controls, this group is more representative of patients that would be tested in a routine clinical setting for active TB.

As listed in Table 5, the pooled specificity (95% CI) of QFT-GIT, T-SPOT.TB, and TST was: 79% (75-82%), 59% (56-62%) and 75% (72-78%) respectively.
Table 4: Specificity of IGRAs and TST in the diagnosis of active TB

<table>
<thead>
<tr>
<th></th>
<th>Pooled specificity (%)</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>79</td>
<td>75-82</td>
<td>81.1</td>
<td>8</td>
<td>569</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>59</td>
<td>56-62</td>
<td>84.5</td>
<td>15</td>
<td>1070</td>
</tr>
<tr>
<td>TST</td>
<td>75</td>
<td>72-78</td>
<td>89.2</td>
<td>12</td>
<td>886</td>
</tr>
</tbody>
</table>

As shown in Table 6, the median proportion of indeterminate results with the interquartile range (IQR) of QFT-GIT and T-SPOT.TB was 7% (12.6%) and 3.4% (5%), respectively.

Table 5: Median proportion of indeterminate results of IGRAs in the diagnosis of active TB

<table>
<thead>
<tr>
<th></th>
<th>QFT-GIT</th>
<th>T-SPOT.TB</th>
<th>TST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median proportion of indeterminate results (%)</td>
<td>7</td>
<td>3.4</td>
<td>n/a</td>
</tr>
<tr>
<td>Interquartile range (IQR) (%)</td>
<td>12.6</td>
<td>5</td>
<td>n/a</td>
</tr>
</tbody>
</table>

As listed in Table 7, the pooled diagnostic odds ratio (OR; 95% CI) of QFT-GIT, T-SPOT.TB, and TST was: 11.47 (5.12-25.69), 18.86 (8.72-40.77) and 5.72 (3.78-8.65), respectively.

Table 6: Pooled diagnostic odds ratio (OR) of IGRAs and TST in the diagnosis of active TB

<table>
<thead>
<tr>
<th></th>
<th>Pooled diagnostic OR</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>11.47</td>
<td>5.12-25.69</td>
<td>67.8</td>
<td>8</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>18.86</td>
<td>8.72-40.77</td>
<td>81.2</td>
<td>15</td>
</tr>
<tr>
<td>TST</td>
<td>5.72</td>
<td>3.78-8.65</td>
<td>46.1</td>
<td>12</td>
</tr>
</tbody>
</table>

The low specificity implies that a high proportion of individuals who do not have active TB would test positive were IGRAs to be used to diagnose active TB. The studies included the assessment of IGRA accuracy not only in control groups with low risk of M. tuberculosis infection, but also in suspects of active TB, a group which, although free from active disease, may have LTBI. This may explain the low specificity reported for QFT-GIT and T-SPOT.TB (79% and 59%, respectively). The authors concluded that the low specificity of IGRAs indicated the low value of the assays in the diagnosis of active TB.

Based on the meta-analysis conducted by TBNET for ECDC that assessed the accuracy (sensitivity and specificity) of IGRAs in the diagnosis of active TB, IGRAs have a low value for diagnosing active TB, and IGRAs cannot be used as a rule-out test for active TB. The authors further concluded that the low specificity of IGRAs may indicate that IGRAs are not suitable to differentiate between LTBI and active TB.

Immunocompromised persons

Immunocompromised patients (e.g. those receiving immunosuppressive drugs, patients with human immunodeficiency virus infection, HIV) represent a group that is at higher risk of reactivating a latent TB infection.

In the TBNET/ECDC systematic review and meta-analysis that assessed the accuracy of IGRAs in the diagnosis of active TB, the included studies that covered immunocompromised patients did not report stratified results, which precluded a specific analysis of the accuracy of IGRA tests among this population group.

A limited number of research studies assessing IGRAs in the diagnosis of active TB in immunosuppressed patients were found. Two studies assessed the accuracy of T-SPOT.TB (not QFT-GIT) in the diagnosis of active TB in immunosuppressed patients. Both studies were conducted in countries of higher TB incidence.

Lai et al. assessed the performance of T-SPOT.TB in the diagnosis of active TB in patients undergoing chronic dialysis in Taiwan. As listed in Table 8, the sensitivity and specificity of T-SPOT.TB were 91.7% and 64.7%, respectively.

Table 7: Sensitivity and specificity of T-SPOT.TB in the diagnosis of active TB in patients undergoing chronic dialysis

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-SPOT.TB</td>
<td>91.7</td>
<td>64.7</td>
<td>29</td>
</tr>
</tbody>
</table>

The authors concluded that these results suggest that T-SPOT.TB represents a sensitive tool for the diagnosis of active TB in patients undergoing chronic dialysis.
Kim et al. assessed and compared the performance of T-SPOT.TB in the diagnosis of extrapulmonary TB in immunocompetent and immunocompromised patients (patients with HIV, lung malignancy, liver cirrhosis, chronic renal failure, or receiving immunosuppressive treatment) in South Korea\textsuperscript{32}.

As listed in Table 9, the sensitivity and specificity (95% CI) of T-SPOT.TB were 96\% (87-100) and 64\% (51-76\%), respectively, in immunocompetent patients; and 88\% (68-97\%) and 69\% (51-83\%), respectively, in immunocompromised patients.

**Table 8. Sensitivity and specificity of T-SPOT.TB in the diagnosis of extrapulmonary TB in immunocompetent and immunocompromised patients**\textsuperscript{32}

<table>
<thead>
<tr>
<th>T-SPOT.TB</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunocompetent patients (n=113)</td>
<td>96% (87-100)</td>
<td>64% (51-76)</td>
</tr>
<tr>
<td>Immunocompromised patients (n=56)</td>
<td>88% (68-97)</td>
<td>69% (51-83)</td>
</tr>
</tbody>
</table>

The authors concluded that T-SPOT.TB had the same sensitivity in immunocompetent and immunocompromised patients (no statistical difference, p=0.32).

The number of studies addressing IGRA accuracy in the diagnosis of active TB in immunocompromised patients remains low, and in the studies presented here no conclusions on the performance of IGRAs in the diagnosis of active TB in this risk group could be drawn.

**HIV-infected patients**

HIV-infected individuals represent a group at higher risk of reactivating a latent TB infection. Furthermore, immunosuppression can lower the sputum bacillary load, making the diagnosis of active TB by microscopy more challenging\textsuperscript{33}. New diagnostic tools that aid the diagnosis of active TB in this risk group are therefore urgently needed\textsuperscript{34-35}.

In the TBNET/ECDC systematic review and meta-analysis that assessed the accuracy of IGRAs in the diagnosis of active TB, the included studies did not stratify the results for immunocompetent and immunosuppressed subgroups\textsuperscript{1}. The authors could therefore not perform an analysis of IGRA accuracy in the subgroup of patients with HIV infection.

More studies addressing IGRAs accuracy in the diagnosis of active TB in HIV-positive patients are needed to allow for the analysis of IGRAs’ accuracy in this sub-group.

In a study by Clark et al. (not included in the systematic review by TBNET that assessed the accuracy of IGRAs in the diagnosis of active TB) the accuracy (sensitivity and specificity) of T-SPOT.TB in patients with HIV infection was determined and stratified on the basis of CD4 T-cell counts\textsuperscript{36}.

As listed in Table 10, the sensitivity for T-SPOT.TB in patients with <300, <200, and <100 CD4 T-cells/\mu l was 95.4\%, 92.9\% and 87.5\%, respectively. Specificity was 100\% for all patients groups.

**Table 9: Sensitivity and specificity of T-SPOT.TB in the diagnosis of active TB in patients with HIV infection, stratified by CD4 T-cell count**\textsuperscript{36}

<table>
<thead>
<tr>
<th>CD4 count (cells/\mu l)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Total number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;300</td>
<td>95.4</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td>&lt;200</td>
<td>92.9</td>
<td>100</td>
<td>14</td>
</tr>
<tr>
<td>&lt;100</td>
<td>87.5</td>
<td>100</td>
<td>8</td>
</tr>
</tbody>
</table>

The authors concluded that T-SPOT.TB sensitivity was not affected by CD4 T-cell count.

**Children**

Children, and particularly infants and children under two years of age, exposed to active TB cases are at increased risk of establishing an infection and developing active TB (including TB meningitis). The diagnosis of TB in children is particularly challenging as symptoms can be confused with symptoms of common childhood diseases. Furthermore, sputum samples are more difficult to obtain from children, and only 10 to 15\% of active TB cases in children are diagnosed by acid-fast staining of *M. tuberculosis* bacilli\textsuperscript{37}.

In the TBNET/ECDC systematic review and meta-analysis that assessed the accuracy of IGRAs in the diagnosis of active TB in children, four studies addressed the performance of IGRAs in this patient subgroup\textsuperscript{1, 38-41}.

As listed in Table 11, the mean sensitivity (SD) of QFT-GIT, T-SPOT.TB, and TST was 79.9\% (20.9\%), 42.2\% (11\%) and 65.4\% (21.1\%), respectively.
Table 10: Sensitivity of IGRAs and TST in the diagnosis of active TB in children

<table>
<thead>
<tr>
<th></th>
<th>Mean sensitivity and SD (%)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>79.9 (20.9)</td>
<td>3</td>
<td>491</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>42.2 (11)</td>
<td>3</td>
<td>227</td>
</tr>
<tr>
<td>TST</td>
<td>65.4 (21.1)</td>
<td>3</td>
<td>n/a</td>
</tr>
</tbody>
</table>

As listed in Table 12, the mean specificity of QFT-GIT, T-SPOT.TB, and TST was: 85.8%, 84% and 89.4% respectively (Table 10).

Table 11: Specificity of IGRAs and TST in the diagnosis of active TB in children

<table>
<thead>
<tr>
<th></th>
<th>Mean specificity (%)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>85.8</td>
<td>1</td>
<td>n/a</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>84*</td>
<td>1</td>
<td>n/a</td>
</tr>
<tr>
<td>TST</td>
<td>89.4</td>
<td>2</td>
<td>n/a</td>
</tr>
</tbody>
</table>

*Note: The TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of active TB in children misstates the mean specificity of the T.SPOT-TB for one study. The correct mean specificity in this study is 84.4%.

As listed in Table 13, the median proportion (IQR) of indeterminate results for QFT-GIT and T-SPOT.TB was 6.3% (3.6%) and 8% (2.5%), respectively.

Table 12: Median proportion of indeterminate of IGRAs in the diagnosis of active TB in children

<table>
<thead>
<tr>
<th></th>
<th>Median proportion of indeterminate results (%)</th>
<th>IQR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>6.3</td>
<td>3.6</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>8</td>
<td>2.5</td>
</tr>
<tr>
<td>TST</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

In the TBNET/ECDC systematic review and meta-analysis, the analysis of IGRA accuracy in the testing of children for active TB could not be stratified by age. For instance, there was not differentiation for children under and over the age of five. Infants and children with active TB under the age of five are at increased risk for poor clinical outcome. Due to the low number of studies, data for a meta-analysis of IGRAs in this vulnerable group are sparse.

The authors of the TBNET/ECDC systematic review and meta-analysis underlined the particularly low number of studies addressing the performance of IGRAs in the diagnosis of active TB in children. Authors concluded that the low sensitivity and specificity of IGRAs in the diagnosis of active TB in children does not support the role of IGRAs as rule-out test for active TB.

3.2 Is there a role for IGRAs in the diagnosis of latent TB infection?

Considerations
- IGRA testing should take place in the context of an overall risk assessment for LTBI which should consider the individual’s history of M. tuberculosis exposure, their clinical history, risk factors, chest radiography, and TST (if applicable).
- LTBI should be screened for in individuals who would benefit from preventive treatment. Screening should be conducted with the intent to determine whether preventive treatment is required.
- The clinical/biological status of LTBI varies widely. It includes individuals previously exposed and infected with M. tuberculosis bacilli which are in a persistent latent state (with possible undetected periods of M. tuberculosis reactivation/dormancy) as well as individuals previously exposed and infected with M. tuberculosis bacilli and with primary lesions which have become sterile over the time.
- IGRAs do not directly measure latent infection with M. tuberculosis bacilli. Instead, they measure the presence of an adaptive immune response (in peripheral blood) directed towards a defined set of M. tuberculosis antigens (ESAT-6, CFP-10 and for QFT-GIT, the antigen TB7.7).
- There is currently no gold standard for diagnosing LTBI and thus for assessing new LTBI diagnostic tools. Instead, individuals with active TB are commonly used as surrogates of LTBI to assess the accuracy of IGRAs. This represents a major limitation as the sensitivity and cut-off of IGRAs derived from individuals with active TB may not translate to individuals with LTBI.
• Please note that the studies reviewed in the TBNET meta-analysis and systematic review focused on low-incidence settings: the derived predictive values may differ compared with those obtained from high-incidence settings.
• Please note that the TBNET meta-analysis and systematic review included a limited number of studies and that the follow-up time of only two years for the derived negative predictive value (NPV) represents an additional limitation.

3.2.1 What is the value of IGRA tests in identifying individuals for preventive treatment?

Evidence

Sensitivity

Sensitivity measures the ability of a test to correctly identify individuals who have a certain disease. In the context of IGRA and the diagnosis of LTBI, sensitivity denotes the proportion of individuals with known LTBI who test positive when tested with IGRA, i.e. the ability of IGRA to correctly diagnose individuals with LTBI.

There is currently no gold standard for the diagnosis of LTBI and thus no method to truly confirm LTBI diagnosis. The sensitivity of IGRA for LTBI diagnosis is therefore commonly assessed in patients with active TB, using this group as a surrogate for LTBI.

The TBNET/ECDC systematic review and meta-analysis did not assess the sensitivity of IGRA in the diagnosis of LTBI.

In a meta-analysis, Menzies et al. assessed the accuracy of IGRA in the diagnosis of LTBI. Sensitivity was determined by using patients with newly diagnosed active TB as a surrogate of latent infection. This meta-analysis included studies performed in low- and high-TB-burden countries.

As listed in Table 14, the pooled sensitivity (95% CI) of QFT-GIT, T-SPOT.TB, and TST was 67% (65–78%), 87% (78–95%) and 71% (64–74%), respectively.

Table 13. Sensitivity of IGRA and TST in the diagnosis of LTBI

<table>
<thead>
<tr>
<th></th>
<th>Pooled sensitivity (%)</th>
<th>95% CI</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>67</td>
<td>46-78</td>
<td>3</td>
<td>133</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>87</td>
<td>78-95</td>
<td>8</td>
<td>337</td>
</tr>
<tr>
<td>TST</td>
<td>71</td>
<td>65-74</td>
<td>14</td>
<td>437</td>
</tr>
</tbody>
</table>

The authors of this meta-analysis concluded that IGRA has a suboptimal sensitivity for identifying LTBI. In clinical terms, the measured low sensitivity would imply that a relatively high proportion of individuals with LTBI (33% for QFT-GIT and 13% for T-SPOT.TB) would test negative if tested with IGRA.

Specificity

Specificity measures the ability of a test to correctly identify individuals who do not have the disease under investigation. In the context of IGRA and the diagnosis of LTBI, specificity denotes the proportion of individuals known not to be infected with M. tuberculosis and who test negative when tested with IGRA; i.e. the ability of IGRA to correctly diagnose individuals who do not have LTBI and classify them as test-negative. As there is no gold standard for diagnosing LTBI, the specificity of IGRA is commonly assessed in populations or settings with a known low or minimal risk of M. tuberculosis infection. This population then represents a surrogate for a group free of M. tuberculosis infection.
In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of LTBI, specificity was assessed among individuals at very low risk of TB infection in low-TB-burden countries. It should be noted that the TBNET analysis included only studies with defined M. tuberculosis cases, only covered commercially available IGRAs, and that data were stratified for low- and high-incidence countries.

As listed in Table 15, the pooled specificity (95% CI) of QFT-GIT, T-SPOT.TB, and TST was 99.4% (97.9–99.9%), 98% (86.8–99.9%) and 88.7% (84.6–92%), respectively.

**Table 14: Specificity of IGRAs and TST in the diagnosis of LTBI**

<table>
<thead>
<tr>
<th></th>
<th>Pooled specificity (%)</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>99.4</td>
<td>97.9–99.9</td>
<td>0</td>
<td>4</td>
<td>346</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>98</td>
<td>86.8–99.9</td>
<td>n/a</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>TST</td>
<td>88.7</td>
<td>84.6–92</td>
<td>94.5</td>
<td>3</td>
<td>309</td>
</tr>
</tbody>
</table>

As listed in Table 16, the median proportion of invalid/indeterminate results, as calculated from the data of the TBNET meta-analysis and systematic review of QFT-GIT, T-SPOT.TB and TST was 6.43%, 11.1% and 11.1%, respectively.

**Table 15: Median proportion of indeterminate results, as calculated from the data of the TBNET systematic review and meta-analysis of IGRAs and TST in the diagnosis of LTBI**

<table>
<thead>
<tr>
<th></th>
<th>QFT-GIT</th>
<th>T-SPOT.TB</th>
<th>TST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median proportion of invalid/indeterminate results (%)</td>
<td>6.43</td>
<td>11.1</td>
<td>11.1</td>
</tr>
<tr>
<td>IQR (%)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

The authors of the meta-analysis concluded that IGRAs show a higher specificity than TST in individuals with very low risk of TB infection in low-TB burden countries. The analysis indicates that IGRAs are a better LTBI diagnostic tool in these settings. In clinical terms this would imply that in a low-TB-burden setting the majority of individuals not infected with M. tuberculosis will be correctly diagnosed as ‘healthy’. The authors’ conclusions were provided with the caveat that only a few studies were included in the analysis.

**Positive predictive value (PPV) for progression**

The positive predictive value (PPV) for progression of an LTBI diagnostic test represents the probability that an individual who tests positive is truly at risk of developing active TB disease later in life.

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of LTBI, PPV for progression was assessed among individuals suspected of LTBI (tested positive using IGRAs and refusing preventive treatment) and subsequently followed up for a period of up to two years.

As listed in Table 17, the PPV for progression (95% CI) range for QFT-GIT, T-SPOT.TB and TST was 2.8% (0.9–6.4%) to 14.6% (6–29%), 3.3% (1.2–7%) to 10% (1.2–32%) and 2.3 (0.7–5.2%) to 3.1% (1.4–5.8%), respectively.

**Table 16: PPV for progression of IGRAs and TST**

<table>
<thead>
<tr>
<th></th>
<th>PPV for progression (%)</th>
<th>95% CI</th>
<th>Time of follow-up (months)</th>
<th>number of studies</th>
<th>N. of untreated subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>2.8</td>
<td>0.9–6.4</td>
<td>22 (median)</td>
<td>1</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td>8.3</td>
<td>1.8–22</td>
<td>19 (mean)</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>14.6</td>
<td>6–29</td>
<td>24 (mean)</td>
<td>1</td>
<td>41</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>3.3</td>
<td>1.2–7</td>
<td>22 (median)</td>
<td>1</td>
<td>181</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.2–32</td>
<td>24</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>TST</td>
<td>2.3*</td>
<td>0.7–5.2</td>
<td>24 (mean)</td>
<td>1</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>3.1**</td>
<td>1.4–5.8</td>
<td>22 (median)</td>
<td>1</td>
<td>288</td>
</tr>
</tbody>
</table>

* TST >5mm, ** TST >10mm

The authors of the meta-analysis noted that there are only a few studies assessing the PPV for progression of IGRAs (only four studies were included) and that the study design varied widely, making the presented values uncertain. Because of the insufficient statistical power due to the low number of studies and the small study populations, it was not possible to make a valid general statement on the PPV for progression of IGRAs. The authors highlighted the need for further research in this field.
Negative predictive value (NPV)

The negative predictive value (NPV) refers to the ability of a test to dismiss from suspicion individuals that do not actually suffer from the disease in question. It measures the probability that the patient will not have the disease when restricted to all patients who test negative. With regard to diagnosing LTBI, the NPV represents the extent to which individuals that test negative truly do not have LTBI.

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of LTBI, the NPV was determined by using patients with confirmed active TB, using the proportion of active TB patients with false-negative IGRA results as a surrogate for the proportion of false-negative scores in LTBI suspects (due to the lack of a gold standard for latent infection identification). The studies included in the meta-analysis were based in low-, intermediate- and high-TB burden countries.

As listed in Table 18, the pooled NPV (95% CI) of QFT-GIT and T-SPOT.TB was 88% (85–92%) and 94% (92–96%), respectively.

### Table 17. NPV of IGRAs in the diagnosis of LTBI

<table>
<thead>
<tr>
<th></th>
<th>Pooled NPV (%)</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>88</td>
<td>85-92</td>
<td>85.1</td>
<td>7</td>
<td>362</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>94</td>
<td>92-96</td>
<td>73.3</td>
<td>12</td>
<td>739</td>
</tr>
<tr>
<td>TST</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

NPV for progression

The NPV for progression of an LTBI diagnostic test represents the probability that an individual who tests negative is not at risk of developing active TB later in life.

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of LTBI, the NPV for progression was assessed among healthy individuals in low-incidence countries that were suspected of LTBI, but subsequently tested negative. The individuals were followed for an average of two years to assess whether they remained disease-free. A number of the studies included in the meta-analysis also included subjects at increased risk of developing TB disease, such as close contacts of active TB patients.

As listed in Table 19, the pooled NPV for progression (95% CI) of QFT-GIT, T-SPOT.TB and TST was 99.8% (99.4–100%), 97.8% (94–99%) and 99.7% (98.5–100%), respectively.

### Table 18: NPV for progression of IGRAs and TST

<table>
<thead>
<tr>
<th></th>
<th>Pooled NPV for progression (%)</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>Time of follow-up (months)</th>
<th>number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>99.8</td>
<td>99.4-100</td>
<td>78.1</td>
<td>Up to 24</td>
<td>4</td>
<td>1442</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>97.8</td>
<td>94-99</td>
<td>65.9</td>
<td>Up to 24</td>
<td>3</td>
<td>182</td>
</tr>
<tr>
<td>TST</td>
<td>99.7</td>
<td>98.5-100</td>
<td>n/a</td>
<td>24 (mean)</td>
<td>1</td>
<td>354</td>
</tr>
</tbody>
</table>

The authors of the meta-analysis concluded that the high NPV for progression measured for IGRAs indicates that an individual with a negative IGRA result will most likely not develop TB disease in the future. However, the authors pointed out that the studies included in the meta-analysis only covered a small number of individuals and were restricted to follow-up periods of up to two years. Further studies on the NPV for progression would be of value.

3.2.2 Can IGRAs differentiate LTBI from active TB?

**Considerations**

IGRAs identify the presence of an adaptive immune response (in peripheral blood) directed towards a defined set of *M. tuberculosis* antigens (ESAT-6, CFP-10 and TB7.7). No evidence available at this time supports that IGRAs are able to distinguish individuals with LTBI from individuals with active TB.

**Expert opinion**

Based on the evidence, IGRAs are not able to differentiate LTBI from active TB. An approach relying exclusively on IGRAs should therefore not be used to differentiate LTBI from active TB.
Evidence
The authors of the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of active TB concluded that, based on the evidence, IGRAs (which indirectly diagnose past or present *M. tuberculosis* infection) cannot differentiate LTBI from active TB.

3.2.3 How should IGRAs be used in different population groups and settings?

Immunocompromised persons

Considerations

- Immunocompromised individuals represent a heterogeneous group which includes patients receiving immunosuppressive treatment and patients with immunodeficiency disorders, such as chronic kidney diseases or HIV. Impaired immune-competence can further be due to other factors, including an immature immune system (e.g. children), genetic or acquired immune defects, immunosuppression associated with other infections, malignancies or immunosuppression induced by treatment modalities (particularly treatment interfering with tumour necrosis factor-alpha (TNF-α) activity).
- Recommendations from existing guidelines or from different professional societies for the diagnosis of LTBI and active TB in immunocompromised individuals should be followed.
- It is essential to maximise the sensitivity in immunocompromised individuals in order to correctly identify as many truly infected individuals as possible.

Expert opinion

As it is essential to maximise sensitivity in immune-compromised individuals, the simultaneous use of TST and IGRAs could be beneficial in identifying LTBI. Any TST or IGRA-positive result should be taken into account in the context of an overall risk assessment when considering preventive treatment.

IGRA should thus be used as part of a comprehensive risk assessment in this group of patients in view of the high risk for TB morbidity and mortality; and prevailing national / society guidelines should be maintained and followed.

Please note that in immune-compromised individuals, IGRAs should not be used to exclude LTBI and/or active TB.

Evidence

Immunocompromised patients (e.g. patients receiving immunosuppressive drugs, or individuals with HIV) represent a group at higher risk of reactivating a latent TB infection, and screening for LTBI is therefore often recommended for this group. However, only limited data are available on the accuracy of IGRAs for this high-risk group.

In immunocompromised patients, IGRA responses have been shown to be reduced compared to immunocompetent subjects, with the former group exhibiting a higher proportion of indeterminate results. The presence of positive controls when running IGRAs (T-cell activation induced by PHA) allows for the assessment of test performance by measuring the ability of the sample’s T-cells to produce IFN-γ, a function that may be impaired in immunocompromised patients.

Only a limited number of studies assessing the accuracy of IGRAs in the diagnosis of LTBI in immunocompromised patients were available. This number was reduced even further when the inclusion criteria defined in the TBNET/ECDC systematic review and meta-analysis were applied. The systematic review therefore did not assess the accuracy of IGRAs in the diagnosis of LTBI for this subgroup.

The summaries given below were taken from a selection of reviews and studies assessing the accuracy of IGRAs in the diagnosis of LTBI in immunocompromised patients. None of these studies were selected or assessed in the TBNET/ECDC systematic review and meta-analysis.

Richeldi et al. assessed the performance of IGRAs in the diagnosis of LTBI (Table 20) in different categories of immunocompromised patients (liver transplant candidates, HIV-infected patients, and patients with hematologic malignancies).
Use of interferon-gamma release assays in support of TB diagnosis

Table 19: Results of IGRAs and TST in the diagnosis of LTBI for different categories of immunocompromised patients

<table>
<thead>
<tr>
<th>% of test result</th>
<th>Liver transplant candidates (n=120)</th>
<th>Patients with HIV (n=116)</th>
<th>Patients with hematologic malignancies (n=95)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive (16.7*)</td>
<td>5.2</td>
<td>10.5*</td>
</tr>
<tr>
<td></td>
<td>Negative (83.3)</td>
<td>94.8</td>
<td>89.5</td>
</tr>
<tr>
<td>QFT-GIT</td>
<td>Positive (26.7)</td>
<td>3.5</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>Negative (72.5)</td>
<td>96.5</td>
<td>72.6</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>0.8</td>
<td>0</td>
<td>1.1</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>Positive (23.3)</td>
<td>4.3</td>
<td>17.9</td>
</tr>
<tr>
<td></td>
<td>Negative (66.7)</td>
<td>89.7</td>
<td>76.8</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>10</td>
<td>6</td>
<td>5.3</td>
</tr>
</tbody>
</table>

* The percentage of positive TST results in liver transplant patients and patients with hematologic diseases was significantly different compared with the percentage of positive results of QFT-GIT and T-SPOT.TB (p<0.05).

As listed in Table 21, the concordance (agreement of test results) between the TST and IGRAs ranged from 80.6% (TST vs T-SPOT.TB in liver transplant candidates) to 95.4% (TST vs QFT-GIT in patients with HIV).

Table 20: Diagnostic agreement of TST and IGRAs in the diagnosis of LTBI in different categories of immunocompromised patients

<table>
<thead>
<tr>
<th>Concordance (%)</th>
<th>Liver transplant candidates (n=108)</th>
<th>Patients with HIV (n=109)</th>
<th>Patients with hematologic malignancies (n=89)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TST vs T-SPOT.TB</td>
<td>80.6</td>
<td>92.7</td>
<td>80.9</td>
</tr>
<tr>
<td>TST vs QFT-GIT</td>
<td>85.2</td>
<td>95.4</td>
<td>91</td>
</tr>
</tbody>
</table>

Indeterminate results not included in calculations.

Richeldi et al. concluded that the performance of IGRAs in the diagnosis of LTBI varies between different categories of immunocompromised patients: in order to maximise the accuracy of LTBI diagnosis, a combined approach based on IGRAs and TST may be of value in these high-risk groups.

Segall et al. reviewed studies assessing IGRA test performance in the diagnosis of LTBI in patients undergoing chronic dialysis (LTBI was defined according to established risk factors associated with LTBI).

As listed in Table 22, the sensitivity, specificity, and indeterminate results for QFT-GIT were 71.4%, 100% and 2.6%, respectively. For T-SPOT.TB, the results were 22–78.6%, 41.9–61.2% and 4.8–11% respectively.

Table 21: Sensitivity, specificity and indeterminate results of IGRAs in the diagnosis of LTBI in immunocompromised patients

<table>
<thead>
<tr>
<th>Patients undergoing chronic dialysis</th>
<th>QFT-GIT</th>
<th>T-SPOT.TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of subjects with determinate results</td>
<td>39</td>
<td>432</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>71.4%</td>
<td>22-78.6%</td>
</tr>
<tr>
<td>Specificity</td>
<td>100%</td>
<td>41.9-61.2%</td>
</tr>
<tr>
<td>Indeterminate results</td>
<td>2.6%</td>
<td>4.8-11%</td>
</tr>
</tbody>
</table>

Although the number of available studies was low, Segall et al. concluded that the rate of indeterminate IGRA results for the screening of LTBI in patients undergoing chronic dialysis was also rather low.

Richeldi et al. suggested the tailored use of IGRAs for the diagnosis of LTBI in different categories of immunocompromised patients and further that caution should be taken when interpreting IGRA results in immunosuppressed patients; Segall et al. concluded that – in the context of an overall risk-assessment – IGRAs should be used instead of TST in the diagnosis of LTBI in patients undergoing chronic dialysis.

The different conclusions drawn by the authors of the two studies described above illustrate the complexity of assessing the accuracy of IGRAs in immunocompromised patients (composed of patients with different diseases, and varying degrees of immunosuppression) in the diagnosis of LTBI. This also underlines the need for further studies assessing the accuracy of IGRAs in the different groups of immunocompromised patients.

19
**HIV-infected patients**

**Considerations**
- HIV/TB co-infection increases the risk of developing active TB\(^50\). The risk of developing active TB has been shown to double by the end of the first year of HIV-infection\(^51\).
- It is essential to maximise the sensitivity in immunocompromised individuals in order to correctly identify as many truly infected individuals as possible.

**Evidence**

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of LTBI, two studies included patients with HIV infection for determining the PPV for progression\(^2\). The number of studies addressing the remaining variables (specificity, NPV, and NPV for progression) of IGRA accuracy in the diagnosis of LTBI TB in HIV-positive individuals was very low and no evaluation of IGRA performance in the diagnosis of LTBI in this risk group could be carried out. More studies on IGRA accuracy in the diagnosis of LTBI and active TB in patients with HIV infection are needed to allow an analysis.

As listed in Table 23, the PPV for progression (95% CI) of QFT-GIT and T-SPOT.TB was 8.3% (1.8–22%) and 10% (1.2–32%), respectively.

**Table 22. PPV for progression of IGRAs in patients with HIV infection\(^2\)**

<table>
<thead>
<tr>
<th></th>
<th>PPV for progression (%)</th>
<th>95% CI</th>
<th>Time of follow-up (months)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>8.3</td>
<td>1.8-22</td>
<td>19 (mean)</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>10</td>
<td>1.2-32</td>
<td>24 (mean)</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

In order to assess individual variables on the accuracy of IGRAs in HIV-infected individuals, the guidance panel was provided with additional studies not assessed in the TBNET/ECDC systematic review and meta-analysis. These studies are presented below.

False-negative or indeterminate IGRA results, especially in patients with advanced HIV-infection and low CD4 T-cell counts, are commonly encountered in HIV-infected individuals. Preliminary studies suggest that the T-SPOT.TB test may be more robust as a standardised number of cells per assay is used for lower CD4 T-cell counts, whereas the QFT-GIT test uses a standardised volume of blood per assay\(^52-53\). This may account for the better performance of the T-SPOT.TB assay in patients with low CD4 counts. In a study by Richeldi et al., this trend was not observed in the HIV-positive study population\(^57\).

Cattamanchi et al. observed a significant difference in the proportion of indeterminate results in T-SPOT.TB tests when comparing patients infected with HIV with different CD4 T-cell counts\(^54\).

As listed in Table 24, the proportion of indeterminate T-SPOT.TB results in patients with >200, 51–200 and ≤50 CD4 T-cells/µl was 95.4%, 92.9% and 87.5%, respectively. The proportion of indeterminate T-SPOT.TB results were significantly different when comparing the different patients stratified by CD4 T-cell count (P=0.03).
**Table 23:** Proportion of indeterminate results of T-SPOT.TB tests in patients with HIV infection, stratified by CD4 T-cell count

<table>
<thead>
<tr>
<th>CD4 count (cells/µL)</th>
<th>Indeterminate results (%)</th>
<th>Total number of subjects tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;200</td>
<td>14</td>
<td>43</td>
</tr>
<tr>
<td>51-200</td>
<td>25</td>
<td>60</td>
</tr>
<tr>
<td>≤50</td>
<td>30</td>
<td>109</td>
</tr>
</tbody>
</table>

When testing HIV-infected patients with IGRAs, it is recommended that the assays be performed as early as possible in the course of the infection, before a decline in CD4 counts. Also, IGRA testing should be repeated after the initiation of highly active anti-retroviral therapy (HAART)\(^47\).

**Children Considerations**

- Children, particularly infants and children under two years of age exposed to active infectious cases, are at increased risk of infection and, if they do not receive preventive treatment, are at subsequent risk to develop active disease (including disseminated forms such as TB meningitis or miliary TB)\(^37\). It is therefore vital to diagnose LTBI and provide preventive treatment. In children under two years of age and/or children with immunosuppression, preventive treatment should be offered (after excluding active disease) following recent exposure in order to prevent infection and subsequent development of active disease.

- New tools to diagnose LTBI and active TB in children are urgently needed\(^55\-56\).

- It is essential to achieve the highest sensitivity of detection when diagnosing LTBI and active TB in children, particularly in children under five years of age\(^57\).

- The diagnosis of active TB in children is particularly challenging as signs and symptoms can be confused with symptoms of other childhood diseases and clinical symptoms may be absent. Furthermore, sputum samples are more difficult to obtain from children, and only 10–15% of active TB cases in children are diagnosed by acid-fast smear-staining\(^37\). TB diagnosis therefore usually relies on a composite of different diagnostic tests. TST and IGRAs are sometimes added to this composite diagnosis, and positive test results indicate an increased risk for TB.

- As there is currently very little data assessing the accuracy of IGRAs in children, particularly children under five years of age, more research on the use of IGRAs, such as studies assessing the longitudinal IGRA responses in children given preventive treatment, is urgently needed.

- It is essential to achieve the highest sensitivity of detection when diagnosing LTBI and active TB in children, particularly in children under five years of age\(^57\).

- The diagnosis of active TB in children is particularly challenging as signs and symptoms can be confused with symptoms of other childhood diseases and clinical symptoms may be absent. Furthermore, sputum samples are more difficult to obtain from children, and only 10–15% of active TB cases in children are diagnosed by acid-fast smear-staining\(^37\). TB diagnosis therefore usually relies on a composite of different diagnostic tests. TST and IGRAs are sometimes added to this composite diagnosis, and positive test results indicate an increased risk for TB.

- As there is currently very little data assessing the accuracy of IGRAs in children, particularly children under five years of age, more research on the use of IGRAs, such as studies assessing the longitudinal IGRA responses in children given preventive treatment, is urgently needed.

- There is a need for a more rigorous and comparable methodological approach, including reference standards, when assessing TB diagnostic tools in children.

- When children are exposed to an infectious TB case, active TB in children under five years of age should be promptly ruled out following the diagnosis of the index case. Preventive treatment should be initiated, regardless of TST and/or IGRA results. Results should be re-evaluated after 8 to 12 weeks (with an assessment of symptoms and a TST and/or IGRA if they were initially negative) to exclude progression to active disease despite preventive treatment. In the event of a negative TST result at 8 to 12 weeks, ongoing preventive treatment should be stopped. There is no need to repeat/perform a TST or an IGRA after completion of preventive treatment in children with an initially positive test result. In children with a positive TST, but with a low risk for TB (i.e. immunocompetent child with no known exposure), a subsequent IGRA (two-step approach) may be considered to rule out false-positive reactions caused by BCG vaccination and/or exposure to NTM.

**Expert opinion**

The available evidence on the use of IGRAs in children is not sufficient to change current practices and guidelines on the diagnosis and treatment of LTBI and/or active TB, particularly in children under five years of age.

Regardless of the approach chosen, these three approaches should not be used to rule out LTBI and/or active TB in children under five years: TST alone, IGRAs alone, a two-step approach.

If applied, IGRAs must always be performed in the context of an overall risk assessment, and decision to treat must be based on this overall risk assessment.

**Evidence**

In the meta-analysis by Menzies et al., assessing the accuracy of IGRAs for the diagnosis of LTBI (see Section 3.2.1 ["What is the value of IGRA tests in identifying individuals for preventive treatment?"] for full study description), the authors did not specify whether any, or which, studies included children to assess the sensitivity of IGRAs. They did, however, indicate that they could not specifically address the accuracy of IGRAs for the diagnosis of LTBI in children due to the insufficient number of studies\(^46\).
In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy (specificity, NPV, and PPV) of IGRA in the diagnosis of LTBI, only a limited number of studies specifically addressed the use of IGRA in children.

As listed in Table 25, the specificity (95% CI) for QFT-GIT, T-SPOT.TB and TST was 100% (91–100%), 98% (87–100%) and 55% (38–71%), respectively.

Table 24: Specificity of IGRA and TST in the diagnosis of LTBI in children

<table>
<thead>
<tr>
<th></th>
<th>Specificity %</th>
<th>95% CI</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
<th>Median age (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>100</td>
<td>91-100</td>
<td>1</td>
<td>40</td>
<td>44 52.5</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>98</td>
<td>87-100</td>
<td>1</td>
<td>40</td>
<td>44 52.5</td>
</tr>
<tr>
<td>TST</td>
<td>55</td>
<td>38-71</td>
<td>1</td>
<td>40</td>
<td>44* 52.5</td>
</tr>
</tbody>
</table>

Study included children with NTM and children with other forms of respiratory tract infection.

* 18/23 children with NTM and 0/22 children with other forms of respiratory tract infection had a positive TST result.

As listed in Table 26, the NPV (95% CI) range for QFT-GIT and T-SPOT.TB was 95.2% (84–99%) to 97.4% (93–99%) and 92.3% (86–96%) to 95.1% (83.5–99.4%), respectively.

Table 25: NPV of IGRA and TST in the diagnosis of LTBI in children

<table>
<thead>
<tr>
<th></th>
<th>Range of NPV (%) (95% CI)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>95.2 (84-99) – 97.4 (93-99)</td>
<td>2</td>
<td>259</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>92.3 (86-96) – 95.1 (83.5-99.4)</td>
<td>2</td>
<td>255</td>
</tr>
<tr>
<td>TST</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

As a subgroup of children with confirmed NTM was included in the analyses, the specificity of the TST in the diagnosis of LTBI could not be determined. Furthermore, due to the limited number of studies addressing the accuracy of IGRA in children in the diagnosis of LTBI, it was not possible to draw conclusions from the analysis.

In one of the studies (Diel et al.; included in the TBNET/ECDC systematic review and meta-analysis), the PPV for progression of QFT-GIT in contacts <16 years and ≥ 6 years old was assessed.

Table 26: PPV for progression of QFT-GIT and TST in children <16 years

<table>
<thead>
<tr>
<th></th>
<th>PPV for progression (%)</th>
<th>Number of test-positive subjects developing disease/total number with test-positive result</th>
<th>Follow-up time (year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>28.6</td>
<td>6/21</td>
<td>Up to 4</td>
</tr>
<tr>
<td>TST</td>
<td>15</td>
<td>6/40</td>
<td>Up to 4</td>
</tr>
</tbody>
</table>

Diel et al. concluded that the results suggest the QFT-GIT is more reliable than TST for identifying children who are at higher risk of developing active TB.

The authors of the TBNET/ECDC systematic review and meta-analysis underlined the particularly low number of studies addressing the performance of IGRA in the diagnosis of LTBI in children and the urgent need for large-scale studies assessing IGRA in this vulnerable group.

The IGRA performance in children has also been reviewed by Lewinsohn et al. Indeterminate rates for IGRA were more frequent in immunocompromised children and in young children (under five years of age), which was also observed by Tsolia et al. Lewinsohn et al. concluded that children aged five years and older can be tested with IGRA for the diagnosis of LTBI; alternatively IGRA can be used as an adjunct to other tests for active TB diagnosis in children aged five years and older, accepting a positive result from either test, while combined negative results cannot exclude infection.

Altogether, the few studies addressing IGRA accuracy in the diagnosis of LTBI in children suggest that the specificity of IGRA is superior to that of the TST. In the diagnosis of active TB, the low sensitivity and specificity does not seem to support the role of IGRA as rule-out test for active TB and indicate that IGRA are not suitable to differentiate children with LTBI from children with active TB. In general, more studies in children addressing the accuracy of IGRA in the diagnosis of LTBI and active TB are needed, particularly in children under the age of five years, who are at increased risk of poor clinical outcome upon developing active disease.
High-incidence and low-incidence TB settings/populations

Considerations
- Populations originating from high-TB incidence countries are often BCG vaccinated.
- More studies assessing the difference in IGRAs predictive value within high-TB and low-TB incidence settings are needed.
- In September 2010, the WHO Strategic and Technical Advisory Group for Tuberculosis (STAG-TB) reviewed the evidence and recommendations of a WHO expert group on the ‘Use of commercial IGRAs in low-income and middle-income countries’, based on the Organization’s standard procedures for policy development. The STAG-TB endorsed the findings of the WHO expert group and supports the strategic approach to develop negative WHO policy recommendations to discourage the use of commercial IGRAs in low-income and middle-income countries (typically high-TB settings and/or high HIV-burden settings). Regarding the use of IGRAs in LTBI diagnosis, this decision was based on a the large body of evidence discouraging the use of IGRAs to diagnose LTBI in adults, children, healthcare workers, contacts, and those involved in outbreak investigations in low-income and middle-income countries (typically high-TB settings and/or high HIV-burden settings). The STAG-TB also acknowledged the challenge to obtain high-quality data due to the lack of a reference standard to identify LTBI.
- As to the use of IGRAs in specific risk groups (e.g. immunocompromised persons, HIV-infected persons and/or children), see the specific sections above.

Evidence
In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of LTBI, only studies from low-TB incidence settings were included for the calculation of specificity and NPV for progression. The analysis showed that IGRAs had a higher association with exposure to M. tuberculosis compared with the TST. As listed in Table 28, the range of OR (95% CI) of QFT-GIT, T-SPOT.TB and TST in the multivariate analysis studies presented in the TBNET study, in which exposure status was a predictor of test positivity, was 1.8 (0.88–3.8) to 66.8 (10.1–441), 1.2 (0.3–4.8) to 38.4 (7.59–616.11), and 0.94 (0.46–1.93) to 6.5 (1.1–36.9), respectively.

Table 27: Odds ratio (OR) of IGRAs and TST in multivariate analysis studies in which exposure status was a predictor of test positivity

<table>
<thead>
<tr>
<th></th>
<th>Range of OR (95% CI)</th>
<th>Total number of subjects tested</th>
<th>Total number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>1.82 (0.88–3.8) – 66.8 (10.1–441)</td>
<td>1604</td>
<td>8</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>1.2 (0.3–4.8) – 38.4 (7.59–616.11)</td>
<td>434</td>
<td>5</td>
</tr>
<tr>
<td>TST</td>
<td>0.94 (0.46–1.93) – 6.5 (1.1–36.9)</td>
<td>1230</td>
<td>8</td>
</tr>
</tbody>
</table>

In low-TB-incidence countries, IGRAs used for the diagnosis of LTBI have also been shown not to be affected by BCG vaccination and correlated better along a gradient of exposure to M. tuberculosis than TST.

The accuracy of IGRAs in the diagnosis of LTBI in high-TB incidence settings was not the main focus of the TBNET/ECDC systematic review and meta-analysis. Separate systematic reviews and meta-analyses on such settings are needed to assess the existing evidence. In countries with high-TB incidence, exposure to NTM may be increased, including for instance exposure to the M. leprae homologues of M. tuberculosis ESAT-6 and CFP-10, which could cause positive IGRA results. In high-TB incidence settings, the priority may be diagnosing and treating patients with active TB.
BCG-vaccinated and non-vaccinated individuals

Considerations

- IGRAs identify the presence of an adaptive immune response (in peripheral blood) directed towards a defined set of *M. tuberculosis* antigens (ESAT-6, CFP-10 and TB7.7) that are absent in most NTM (with the exception of *M. flavescentis, M. kansasii, M. marinum* and *M. szulgai*). More importantly, these antigens are not present in any of the BCG vaccine strains, therefore eliminating the consideration of test cross-reactivity in BCG-vaccinated individuals.
- Due to different BCG vaccination policies (BCG vaccination at birth, repeated BCG-vaccination, no vaccination) in EU and non-EU countries, the BCG vaccination situation is not homogenous.
- It has been observed that ten years post-vaccination, BCG received in infancy has no evident effect on TST results, whereas BCG-vaccination in older age groups induces more persistent, more frequent and more pronounced TST responses.
- For the diagnosis of LTBI in specific settings of immunocompetent adults, please refer to Section 3.2.1.
- More studies are needed on the effect of BCG vaccination given in infancy and potential subsequent exposure to NTM, on IGRA results (short-term and long-term effect), and on the PPV of IGRAs in BCG-vaccinated people.
- Cost-effectiveness is a factor when deciding whether to use a single-test or a two-step approach (TST followed by IGRA). Logistical aspects also influence the method of choice. In a BCG-vaccinated population, a two-step approach aims at increasing test specificity (with a higher NPV).

Evidence

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of LTBI, the influence of BCG vaccination on the accuracy of IGRAs was analysed. IGRAs were not affected by prior BCG vaccination and more likely to be associated with exposure to *M. tuberculosis* cases. This was assessed through multivariate analysis studies performed in low-, intermediate- and high-TB burden countries.

As listed in Table 29, the range in OR (95% CI) of QFT-GIT, T-SPOT.TB and TST in the multivariate analysis studies included in the TBNET review, in which BCG vaccination was a predictor of test positivity, was 0 for both QFT-GIT and T-SPOT.TB, and 3.8 (1-13.9) to 24.7 (11.7-52.5) for TST.

Table 28: Odds ratio (OR) of IGRAs and TST in multivariate analysis studies in which BCG vaccination status was a predictor of test positivity

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Range of OR (95% CI)</th>
<th>Total number of subjects tested</th>
<th>Total number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT &amp; T-SPOT.TB</td>
<td>0 (i.e. no correlation)</td>
<td>n/a</td>
<td>9</td>
</tr>
<tr>
<td>TST</td>
<td>3.8 (1-13.9) – 24.7 (11.7-52.5)</td>
<td>n/a</td>
<td>9</td>
</tr>
</tbody>
</table>

In a systematic review and meta-analysis by Pai et al. in 2008, the specificity of IGRAs in BCG-non-vaccinated and BCG-vaccinated populations was compared. In respect to pooled specificity of the QFT-GIT, a high specificity was measured regardless of vaccination status; 99% (CI 98-100%) in BCG-non-vaccinated and 96% (CI 94-98%) in BCG-vaccinated individuals. Data are not shown here as no specific data on the QFT-GIT could be extracted from Pai et al.’s review.

As listed in Table 30, and as reported in the systematic review by Pai et al., the specificity for T-SPOT.TB and the pooled specificity for TST in non-BCG (or predominantly non-vaccinated) individuals is 100% and 97%, respectively. The specificity for T-SPOT.TB and the pooled specificity for TST in BCG-vaccinated (or predominantly vaccinated) individuals was 84.7% and 59%, respectively. As the table shows however, the values for the T-SPOT.TB assay are based on only one study, and the authors pointed out the limited amount of data on this assay.

1 The review also included studies on the predecessor, QuantiFERON-TB Gold.
Table 29: Pooled specificity of T-SPOT.TB and TST for diagnosis of LTBI in BCG-vaccinated and non-vaccinated individuals

<table>
<thead>
<tr>
<th></th>
<th>Not BCG or predominantly non-vaccinated</th>
<th>BCG or predominantly vaccinated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled specificity (%)</td>
<td>Number of studies</td>
</tr>
<tr>
<td>T-SPOT.TB</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>TST</td>
<td>97</td>
<td>6</td>
</tr>
</tbody>
</table>

Contact tracing

Considerations
- The aim of contact tracing is to detect LTBI in exposed individuals. Contact tracing should be conducted with the intent to provide preventive treatment.
- National guidelines for contact tracing should be followed for the most efficient and cost-effective approach.

Expert opinion
Given the available evidence, IGRAs could be used in contact tracing algorithms that use the two-step approach (following TST, in TST-positive subjects).

This combined approach is based on the need to maximise specificity while improving the cost-effectiveness of contact tracing in immunocompetent adult contacts.

Evidence
The risk for progression to active TB is at its highest during the first years following infection. Therefore, the diagnosis of *M. tuberculosis* infection in newly infected individuals who have been in contact with active TB cases is particularly important in order to provide them with appropriate preventive treatment.

The two-step approach in the diagnosis of LTBI commonly consists of a) conducting the TST, followed by b) an IGRA test. This approach is considered to improve the sensitivity and specificity of the TST results, as it may corroborate a positive TST. Conversely, it may consolidate, in certain clinical situations and risk groups, a negative TST result.

A number of expert opinions have been published regarding the use of IGRAs for contact tracing. Overall, most experts appear to consider a two-step approach (TST followed by IGRA) the most promising when screening contacts for LTBI.

In a 2007 workshop on the use of IGRAs in low- and medium-prevalence countries in Europe, experts agreed that applying the two-step approach for diagnosing LTBI is the optimal strategy for contract tracing. This is particularly valid in contact tracing situations in which there is a known index case (i.e. case with active TB).

A TBNET consensus statement on LTBI, published in 2009, gives an overview of the different aspects to consider when conducting contact investigations for LTBI (who to screen; how to provide preventive treatment). In their statement, the experts agreed that IGRAs can be used to confirm a positive TST result in order to prevent unnecessary treatment of contacts that do not have LTBI.

In 2009, a review of national guidelines on the use of IGRAs was presented at the second Global Symposium on IGRAs, giving an overview of the different strategies recommended. National guidelines often recommend the use of IGRAs for diagnosing LTBI during contact tracing, with most countries favouring the two-step approach.

Lastly, the newly updated US CDC guidelines on IGRAs and the detection of *M. tuberculosis* infection state that TST or IGRAs can be used alone when conducting contact investigations.

Screening of occupational healthcare workers

Considerations
- The purpose of screening healthcare workers is to identify LTBI.
- Serial TSTs of BCG-vaccinated individuals can result in boosting and thus cause a false-positive result. It is therefore recommended to take into account the setting/country and its guidelines/policies on healthcare workers, as well as the healthcare workers’ BCG vaccination status.
- Occupational healthcare workers are often BCG vaccinated.
- Healthcare workers may have an increased exposure to NTM in their work settings.
- Practices regarding the use of IGRAs in the screening of healthcare workers vary, depending on the screening objectives. Some guidelines propose IGRAs for the screening of healthcare workers that have been exceptionally exposed to TB or recommend that healthcare workers receive a screening before taking
up their jobs. However, several guidelines do not specifically mention whether IGRAs should be a preferred tool for the screening of healthcare workers\textsuperscript{68}.

- Existing national guidelines for occupational healthcare workers should be followed.
- The total exposure of healthcare workers may have an impact on the outcome or predictive value of IGRA tests.
- IGRAs may be used as a baseline test, but serial testing problems of conversion/reversion (particularly around the cut-off values for the tests) may occur\textsuperscript{68}.
- More research is needed on the use of IGRAs for the screening of occupational healthcare workers, e.g. studies assessing the accuracy of repeated IGRA testing or exploring the issue of conversion/reversion of test results.

**Expert opinion**

There is insufficient evidence on the PPV of IGRAs for the screening of healthcare workers to state an educated opinion on this topic.

However, given the available evidence the use of IGRAs in the two-step approach could increase the specificity depending on the population tested (e.g. BGC vaccination status).

**Evidence**

A review conducted by Swindells et al. concluded that IGRA testing for diagnosing LTBI in healthcare workers was beneficial but that more studies are needed (the heterogeneity of the studies assessing the role of IGRAs in the testing of healthcare workers did not allow to perform a meta-analysis)\textsuperscript{69}. 


4 Future research needs and considerations

4.1 Can IGRAs be used with extrasanguinous fluids to support the diagnosis of active TB?

**Considerations**

- The standard diagnostic methods for active TB diagnosis are described in Section 1. However, in cases that are difficult to diagnose, all available methods should be used for the direct detection of the pathogen and its components. New microbiological and immunological diagnostic tools are needed especially for early confirmation of severe disease, for example TB meningitis.
- IGRAs were developed for blood samples and are not licensed for use with extrasanguinous fluids\textsuperscript{20-21}.
- Findings from the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of active TB warrant continued research, particularly on the use IGRAs with extrasanguinous fluids in order to support the diagnosis of difficult-to-diagnose patients.

**Expert opinion**

There are certain clinical situations that urgently require supplementary tools for the direct or indirect diagnosis of active TB. Research on the use of IGRAs in extrasanguinous fluids is ongoing, but there is currently not enough evidence to support the use of IGRAs with extrasanguinous fluids in the diagnosis of active TB.

**Evidence**

**Sensitivity**

Sensitivity measures the ability of a test to correctly identify individuals who have a certain disease. When applied to IGRAs and the diagnosis of active TB, sensitivity denotes the proportion of individuals with known active TB disease who test positive when IGRAs are used; i.e. the ability of IGRAs to correctly diagnose individuals with active TB and classify them as test-positive.

In the TBNET/ECDC systematic review and meta-analysis that assessed the accuracy of IGRAs in the diagnosis of active TB, sensitivity was assessed with extrasanguinous fluids (pleural fluids, bronchoalveolar lavage or ascetic fluid) in patients with clinical suspicion of TB disease (culture-confirmed and unconfirmed TB cases)\textsuperscript{1}.

As listed in Table 31, the pooled sensitivity (95% CI) of QFT-GIT was 48% (39–58%); the pooled sensitivity of T-SPOT.TB was 88% (82–92%).

<table>
<thead>
<tr>
<th>IGRAs</th>
<th>Pooled sensitivity (%)</th>
<th>95% CI</th>
<th>Inconsistency I\textsuperscript{2} (%)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT*</td>
<td>48</td>
<td>39-58</td>
<td>0</td>
<td>4</td>
<td>116</td>
</tr>
<tr>
<td>T-SPOT.TB**</td>
<td>88</td>
<td>82-92</td>
<td>57.9</td>
<td>7</td>
<td>186</td>
</tr>
</tbody>
</table>

* Pooled sensitivity was 52% (95% CI 39-64%; I\textsuperscript{2}=38%) for patients with culture-confirmed TB.
** Pooled sensitivity was 88% (95% CI 81-93%; I\textsuperscript{2}=22%) for patients with culture-confirmed TB.

The authors of this meta-analysis concluded that the sensitivity of IGRAs was too low to support their role as a rule-out test for active TB.

**Specificity**

Specificity measures the ability of a test to correctly identify individuals who do not have the disease in question. In the context of IGRAs and the diagnosis of active TB, specificity denotes the proportion of individuals known not to have active TB disease and who test negative when IGRAs are used; i.e. the ability of IGRAs to correctly diagnose individuals who do not have active TB and classify them as test-negative.

In the TBNET/ECDC systematic review and meta-analysis assessing the accuracy of IGRAs in the diagnosis of active TB, the specificity was assessed with extrasanguinous fluids (pleural effusion, bronchoalveolar lavage fluids, or ascetic fluids) from patients with clinical suspicion of TB disease\textsuperscript{1}.

As listed in Table 32, the pooled specificity (95% CI) of QFT-GIT was 82% (70–91%); the pooled specificity of T-SPOT.TB was 82% (78–86%).
Table 31: Specificity of IGRAs in the diagnosis of active TB performed in extrasanguinous fluids (pleural fluids, bronchoalveolar lavage, and ascetic fluid)¹

<table>
<thead>
<tr>
<th></th>
<th>Pooled specificity (%)</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>Number of studies</th>
<th>Total number of subjects with determinate results</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>82</td>
<td>70-91</td>
<td>0</td>
<td>4</td>
<td>56</td>
</tr>
<tr>
<td>T-SOT.TB</td>
<td>82</td>
<td>78-86</td>
<td>71.5</td>
<td>7</td>
<td>368</td>
</tr>
</tbody>
</table>

As listed in Table 33, the median proportion of indeterminate results (IQR) for QFT-GIT was 23.1% (40.1%), the median proportion of indeterminate results for T-SOT.TB was 5% (8%).

Table 32: Median proportion of invalid/indeterminate of IGRAs for diagnosis of active TB performed in extrasanguinous fluids¹

<table>
<thead>
<tr>
<th></th>
<th>QFT-GIT</th>
<th>T-SOT.TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median proportion of invalid/ indeterminate results (%)</td>
<td>23.1</td>
<td>5</td>
</tr>
<tr>
<td>IQR (%)</td>
<td>40.1</td>
<td>9.8</td>
</tr>
</tbody>
</table>

As listed in Table 34, the pooled diagnostic OR (95% CI) of QFT-GIT and T-SOT.TB was: 3.84 (1.73–8.51) and 35.83 (15.57–82.43), respectively.

Table 33: Pooled diagnostic odds ratio (OR) of IGRAs for diagnosis of active TB in extrasanguinous fluids¹

<table>
<thead>
<tr>
<th></th>
<th>Pooled diagnostic OR</th>
<th>95% CI</th>
<th>Inconsistency I² (%)</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT</td>
<td>3.84</td>
<td>1.73-8.51</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>T-SOT.TB</td>
<td>35.83</td>
<td>15.57-82.43</td>
<td>30.8</td>
<td>7</td>
</tr>
</tbody>
</table>

When performed with extrasanguinous fluids, the sensitivity of T-SOT.TB was significantly higher (88%) compared with QFT-GIT (48%). Also, the number of indeterminate results was lower. Based on these data, authors suggested that the T-SOT.TB assay is currently the best available extrasanguinous-based immunological method for the diagnosis of active TB. The authors therefore suggested that the T-SOT.TB performed with extrasanguinous fluids could in low-incidence settings represent an improvement for the rapid diagnosis of active TB when combined with the standard methods of diagnosis¹.

The authors also pointed out that data are limited and that the existing data indicate that IGRAs have limited accuracy in diagnosing active TB when used with extrasanguinous samples.

4.2 Large-scale population screening for LTBI

Considerations

- Large-scale screening encompasses mass screenings (large-scale screening of whole population groups) and selective screening (screening of selected high-risk groups in a population performed at a large scale)⁷⁰.
- In EU countries, the aim of large-scale screening usually is to improve active-TB case-finding and subsequently treat active TB in population groups that have a considerably higher prevalence than average.
- In large-scale screening, a population is commonly mixed in terms of age and BCG vaccination status, and may be composed of individuals originating from high- or low-TB incidence setting as well as groups for which the general immune status should be taken into account.
- IGRAs could have an advantage by providing results after only one visit to the healthcare facility; however, a second visit is required in case of a positive test result.
- Limited evidence is available on the predictive value of IGRAs in large-scale screenings in low-TB incidence settings, and studies on specific risk groups with higher statistical power are needed.

Expert opinion

The decision to conduct large-scale screenings should be based on evidence of the cost-effectiveness of screenings in populations with a high risk.

The decision to use IGRAs alone or in combination with TST in the diagnosis of LTBI is based on the evidence and opinions presented in the above sections referring to population groups or situations. The PPV of the test will vary widely according to the risk of LTBI in the tested population.
4.3 Future research needs

Prospective studies on diagnostic accuracy

- More and larger prospective studies assessing the positive and negative predictive values of IGRAs for the diagnosis of LTBI (and, if possible, TST). Further studies are also needed for the diagnosis of active TB in different settings (low-, intermediate- and high-TB incidence settings) and unselected populations.
- More studies assessing IGRA accuracy and predictive value (and possible limitations) in children (and particularly in children less than five years old) and other high-risk groups such as immunocompromised populations.
- Value of IGRAs in BCG-vaccinated individuals and/or exposed to NTM.
- More studies assessing the ability of IGRAs to discriminate recent from remote LTBI.
- Studies assessing the impact of re-infection with *M. tuberculosis* on immune reactivity, as defined by IFN production in IGRAs.
- More studies addressing the reproducibility of IGRAs and the phenomenon of conversion/reversion of IGRA-results over time (serial testing) as well as after treatment for active TB and LTBI.
- Effect of blood incubation delay on IGRA accuracy and extent of indeterminate results.
- More studies assessing the accuracy of IGRAs when used with extrasanguinous fluids in order to diagnose active TB.
- Studies determining the accuracy of IGRAs in the diagnosis of extrapulmonary TB.
- There is the need to provide harmonised guidelines for prospective studies so investigators have clear definitions, including clinical phenotypes for collating results in order to reach meaningful conclusions with adequate statistical power.

Biological/immunological issues

- More studies to identify the biological basis for discordant results between TST and IGRAs.
- Research to develop IGRAs that incorporate new *M. tuberculosis*-specific antigens and alternative cytokines that would enhance sensitivity, allowing LTBI to be distinguished from active TB.
- Studies assessing which cells contribute to IFN-$\gamma$ production once *M. tuberculosis* infection has been cleared, e.g. by appropriate drug treatment.

Programmatic issues

- Studies to evaluate the feasibility and cost of IGRAs in the diagnosis of LTBI and active TB in different settings and for different purposes (e.g. contact screening, serial testing of healthcare workers).
- Studies to evaluate the resources needed for the implementation of IGRAs.
References


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

Policy Statement
Use of tuberculosis interferon-gamma release assays (IGRAs) in low- and middle-income countries: policy statement.

1. Interferon-gamma. 2. Tuberculosis, Pulmonary - diagnosis. 3. Immunoassay - methods. 4. Reagent kits, diagnostic. 5. Developing countries. 1. World Health Organization.

ISBN 978 92 4 150267 2 (NLM classification: WF 220)
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 IGRA; 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 IGRA. 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

This document was prepared by Karin Weyer, Christopher Gilpin, Fuad Mirzayev and Wayne van Gemert (WHO Stop TB Department) on the basis of consensus at an international Expert Group Meeting convened by WHO in Geneva on 20th-21st July 2010.

WHO gratefully acknowledges the contributions of the Chair of the Expert Group (Holger Schünemann) and the members of the Expert Group (Annex 1) who developed the recommendations.

The findings and recommendations from the Expert Group Meeting were presented to the WHO Strategic and Technical Advisory Group for Tuberculosis (STAG-TB, Annex 2), in September 2010 (http://www.who.int/tb/advisory_bodies/stag/en/). STAG-TB acknowledged a compelling evidence base and large body of work demonstrating the poor performance of current commercial IGRAs in low- and middle-income countries (typically high TB and/or HIV burden settings) and the adverse impact of misdiagnosis and wasted resources on patients and health services using these tests for the diagnosis of active TB.

STAG-TB also acknowledged a large body of work and compelling evidence base to discourage the use of IGRAs in low- and middle-income countries for the detection of LTBI, acknowledging the difficulty in obtaining high quality data on the diagnosis of LTBI in the absence of a reference standard.

STAG-TB endorsed the findings of the Expert Group and supported the strategic approach to develop WHO policy recommendations to discourage the use of commercial IGRAs over the TST in low- and middle-income countries. This document was finalized following consideration of all comments and suggestions from the participants of the Expert Group and STAG-TB.

USAID is acknowledged for funding the development of these guidelines through USAID-WHO Consolidated Grant No. GHA-G-00-09-00003. TDR and TREAT-TB/The Union are acknowledged for sponsoring the systematic reviews commissioned in advance of the Expert Group meeting.
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 (QFT-GT, 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-GT: 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.¹

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

¹GRC 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 IGRA 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 IGRA for diagnosis of active TB beyond readily available patient data and conventional tests;

• The systematic review focused on the use of IGRA 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 IGRA 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 IGRA 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 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

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 (LMIC).

Table 2. GRADE Summary of Findings – Role of IGRAs for evaluation of patients with pulmonary TB in low- and middle-income countries

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

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

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

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

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

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

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

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

Table 11. GRADE Evidence Profile: Performance of IGRAs for the diagnosis of LTBI in contacts of active TB in low-and middle-income countries.
Table 12. GRADE Evidence Profile: Predictive value of commercial IGRA for incident active TB in low and middle-income countries

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


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

<table>
<thead>
<tr>
<th>No of Participants (studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Outcome: Diagnostic accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Positives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19) &lt;sup&gt;A1&lt;/sup&gt;</td>
<td>Cross-sectional</td>
<td>No Serious Limitation &lt;sup&gt;A2&lt;/sup&gt;</td>
<td>No Serious Indirectness &lt;sup&gt;A3&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;A4&lt;/sup&gt; (-1)</td>
<td>Serious &lt;sup&gt;A5&lt;/sup&gt; (-1)</td>
<td>Likely &lt;sup&gt;A6&lt;/sup&gt;</td>
<td>Low &lt;sup&gt;⊗⊗⊗⊗&lt;/sup&gt;</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>True Negatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19) &lt;sup&gt;A1&lt;/sup&gt;</td>
<td>Cross-sectional</td>
<td>No Serious Limitation &lt;sup&gt;A2&lt;/sup&gt;</td>
<td>No Serious Indirectness &lt;sup&gt;A3&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;A4&lt;/sup&gt; (-1)</td>
<td>Serious &lt;sup&gt;A5&lt;/sup&gt; (-1)</td>
<td>Likely &lt;sup&gt;A6&lt;/sup&gt;</td>
<td>Low &lt;sup&gt;⊗⊗⊗⊗&lt;/sup&gt;</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>False Positives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19) &lt;sup&gt;A1&lt;/sup&gt;</td>
<td>Cross-sectional</td>
<td>No Serious Limitation &lt;sup&gt;A2&lt;/sup&gt;</td>
<td>No Serious Indirectness &lt;sup&gt;A3&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;A4&lt;/sup&gt; (-1)</td>
<td>Serious &lt;sup&gt;A5&lt;/sup&gt; (-1)</td>
<td>Likely &lt;sup&gt;A6&lt;/sup&gt;</td>
<td>Low &lt;sup&gt;⊗⊗⊗⊗&lt;/sup&gt;</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>False Negatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067 (19) &lt;sup&gt;A1&lt;/sup&gt;</td>
<td>Cross-sectional</td>
<td>No Serious Limitation &lt;sup&gt;A2&lt;/sup&gt;</td>
<td>No Serious Indirectness &lt;sup&gt;A3&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;A4&lt;/sup&gt; (-1)</td>
<td>Serious &lt;sup&gt;A5&lt;/sup&gt; (-1)</td>
<td>Likely &lt;sup&gt;A6&lt;/sup&gt;</td>
<td>Low &lt;sup&gt;⊗⊗⊗⊗&lt;/sup&gt;</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>B. Outcome: Proportion indeterminate tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2872 (33) &lt;sup&gt;B1&lt;/sup&gt;</td>
<td>Cross-sectional</td>
<td>Serious &lt;sup&gt;B2&lt;/sup&gt; (-1)</td>
<td>No Serious Indirectness &lt;sup&gt;B3&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;B4&lt;/sup&gt; (-1)</td>
<td>No Serious Imprecision &lt;sup&gt;B5&lt;/sup&gt;</td>
<td>Likely &lt;sup&gt;B6&lt;/sup&gt;</td>
<td>Low &lt;sup&gt;⊗⊗⊗⊗&lt;/sup&gt;</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>C. Outcome: Incremental value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>943 (2) &lt;sup&gt;C1&lt;/sup&gt;</td>
<td>Cohort</td>
<td>Serious &lt;sup&gt;C2&lt;/sup&gt; (-1)</td>
<td>No Serious Indirectness &lt;sup&gt;C3&lt;/sup&gt;</td>
<td>No Serious Inconsistency &lt;sup&gt;C4&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;C5&lt;/sup&gt; (-1)</td>
<td>Unlikely &lt;sup&gt;C6&lt;/sup&gt;</td>
<td>Low &lt;sup&gt;⊗⊗⊗⊗&lt;/sup&gt;</td>
<td>Critical (7-9)</td>
</tr>
</tbody>
</table>
Quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high quality observational studies (cross-sectional or cohort studies enrolling patients with diagnostic uncertainty) and at moderate when these types of studies were absent. One point was then subtracted when there was a serious issue identified or two points when there was a very serious issue identified in any of the criteria used to judge the quality of evidence. The evidence rankings were considered to be the same for consideration of true positives, false positives, false negatives, and true negatives.

Footnotes:

1 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 A1.

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

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

B6 Please see footnote A6.

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

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

C3 See footnote A1. In addition, area under the receiver-operating-characteristic curve (AUC) may be a less clinically interpretable measure of risk assessment than risk-reclassification statistics.

C4 Only two studies were available; effect estimates for both studies were in the same direction and consistent.

C5 Imprecision, as evaluated by 95% confidence intervals of the area under the receiver-operating-characteristic curves (AUC), was reasonable for both studies.

C6 Because of the relative novelty of these methods, at this time it is unlikely that studies of IGRA incremental value have been unpublished due to publication bias.
### 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

<table>
<thead>
<tr>
<th>Subgroups</th>
<th>Effect % (95% CI)</th>
<th>No. of participants (studies)</th>
<th>What do these results mean given 10% prevalence among suspects being screened for TB?</th>
<th>What do these results mean given 30% prevalence among suspects being screened for TB?</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T-SPOT.TB; HIV-infected</strong></td>
<td>Sensitivity 78% (56, 91) Specificity 55% (45, 64)</td>
<td>549 (5)</td>
<td>With a prevalence of 10%, 100/1000 will have TB. Of these, 78 (TP) will be identified; 22 (FN) will be missed by T-SPOT.TB. Of the 900 patients without TB, 495 (TN) will not be treated; 405 (FP) will be unnecessarily treated.</td>
<td>With a prevalence of 30%, 300/1000 will have TB. Of these, 234 (TP) will be identified; 66 (FN) will be missed by T-SPOT.TB. Of the 700 patients without TB, 385 (TN) will not be treated; 315 (FP) will be unnecessarily treated.</td>
<td>Low</td>
</tr>
<tr>
<td><strong>T-SPOT.TB; HIV-uninfected</strong></td>
<td>Insufficient data for pooled estimates</td>
<td>364 (3)</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td><strong>QuantiFERON-TB Gold In-Tube; HIV-infected</strong></td>
<td>Sensitivity 62% (41, 79) Specificity 51% (39, 64)</td>
<td>469 (6)</td>
<td>With a prevalence of 10%, 100/1000 will have TB. Of these, 62 (TP) will be identified; 38 (FN) will be missed by QFT-GIT. Of the 900 patients without TB, 459 (TN) will not be treated; 441 (FP) will be unnecessarily treated.</td>
<td>With a prevalence of 30%, 300/1000 will have TB. Of these, 186 (TP) will be identified; 114 (FN) will be missed by QFT-GIT. Of the 700 patients without TB, 357 (TN) will not be treated; 343 (FP) will be unnecessarily treated.</td>
<td>Low</td>
</tr>
<tr>
<td>Outcome</td>
<td>Subgroup</td>
<td>Effect % (95% CI)</td>
<td>No. of participants (studies)</td>
<td>What do these findings mean?</td>
<td>Quality of Evidence</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>IGRA-TST sensitivity difference*</td>
<td>QuantiFERON-TB Gold In-Tube</td>
<td>1% (-11 to 13%)*</td>
<td>475 (10)</td>
<td>This evidence suggests that QFT-GIT is no more sensitive than TST for active TB diagnosis in low- and middle-income countries.</td>
<td>Low ☓☐☐</td>
</tr>
<tr>
<td></td>
<td>T-SPOT.TB</td>
<td>9% (-10% to 28%)*</td>
<td>206 (5)</td>
<td>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.</td>
<td>Low ☓☐☐</td>
</tr>
<tr>
<td>Proportion indeterminate tests</td>
<td>QuantiFERON-TB Gold In-Tube, HIV-uninfected Subjects</td>
<td>4% (1-7%)</td>
<td>1603 (11)</td>
<td>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.</td>
<td>Low ☓☐☐</td>
</tr>
<tr>
<td></td>
<td>T-SPOT.TB, HIV-uninfected Subjects</td>
<td>3% (1-4%)</td>
<td>494 (5)</td>
<td>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.</td>
<td>Low ☓☐☐</td>
</tr>
<tr>
<td></td>
<td>QuantiFERON-TB Gold</td>
<td>16% (10-21%)</td>
<td>728 (10)</td>
<td>In low- and middle-income</td>
<td>Low</td>
</tr>
<tr>
<td>Study</td>
<td>Subjects</td>
<td>Proportion</td>
<td>Number of Participants</td>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>------------</td>
<td>------------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>In-Tube, HIV-infected Subjects</td>
<td></td>
<td></td>
<td></td>
<td>In low- and middle-income 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.</td>
<td></td>
</tr>
<tr>
<td>T-SPOT.TB, HIV-infected Subjects</td>
<td>8% (1-15%)</td>
<td>666 (7)</td>
<td></td>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

**Incremental value**

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

* 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

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

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

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

Footnotes

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

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

Study limitations were assessed using the QUADAS tool. Two (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.

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

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)

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)

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)
**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)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. Participants</th>
<th>Principal Findings</th>
<th>What do these findings mean?</th>
<th>Quality of Evidence</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive value for active TB</td>
<td></td>
<td>No studies in LMIC</td>
<td></td>
<td>Critical (7-9)</td>
<td></td>
</tr>
</tbody>
</table>
| Performance of IGRAs against dichotomous measure of exposure | QFT-G/QFT-GIT: 59 (3) | Pooled Odds ratios  
• QFT-G/QFT-GIT: OR 1.30 (95% CI 0.20-8.32)  
• T-SPOT: OR 2.24 (95% CI 0.88-5.64)  
• TST (10mm): OR 0.81 (95% CI 0.38-1.74) | Children exposed to TB have a higher risk of LTBI, expressed by a higher probability of a positive test for LTBI (QFT, T-SPOT or TST) than in unexposed children. Wide and overlapping confidence intervals indicate similar performance of all three tests. | Very Low (3-5) | Critical (7-9) |
| Performance of IGRAs against exposure gradient | QFT-G/QFT-GIT: 773 (5) | 1. Pooled correlation between test and exposure gradient:  
• QFT-G/QFT-GIT: 0.28 (95%CI 0.06-0.86, 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 (3-5) | 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

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: What is the sensitivity of IGRAs in children with active TB?</td>
<td>207 (6)</td>
<td>Mainly cross-sectional</td>
<td>Serious (2)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Very serious (2)</td>
<td>Likely</td>
<td>Very Low&lt;br&gt;⊕〇〇〇</td>
</tr>
<tr>
<td>B: What is the specificity of IGRAs in children without TB?</td>
<td>519 (4)</td>
<td>Mainly cross-sectional</td>
<td>Serious (2)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not applicable</td>
<td>Likely</td>
<td>Very Low&lt;br&gt;⊕〇〇〇</td>
</tr>
<tr>
<td>C: What is the proportion of indeterminate IGRA results among children assessed for active TB?</td>
<td>656 (5)</td>
<td>Mainly cross-sectional</td>
<td>Serious (2)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Serious</td>
<td>Likely</td>
<td>Very Low&lt;br&gt;⊕〇〇〇</td>
</tr>
<tr>
<td>D: What is the diagnostic accuracy of IGRAs in HIV-infected children?</td>
<td>36 (1)</td>
<td>Cross-sectional</td>
<td>Serious (2)</td>
<td>Not serious</td>
<td>Not applicable</td>
<td>Very serious</td>
<td>Likely</td>
<td>Very Low&lt;br&gt;⊕〇〇〇</td>
</tr>
<tr>
<td>E: What is the diagnostic accuracy of IGRAs in children &lt; 5 years?</td>
<td>471 (2)</td>
<td>Cross-sectional</td>
<td>Serious (2)</td>
<td>Not applicable</td>
<td>Not serious</td>
<td>Very serious</td>
<td>Likely</td>
<td>Very Low&lt;br&gt;⊕〇〇〇</td>
</tr>
</tbody>
</table>

Footnotes

1 The quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence was rated 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.
A1. 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 I² 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)

A4. The I² 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.

The I² 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.

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)

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 I² 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%, I² 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)
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)

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

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

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

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

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

Only one study – inconsistency therefore cannot be assessed.

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

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

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

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

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

The confidence intervals for sensitivity and specificity of QFT-GIT were small, but wide for T-SPOT. The data from LMIC to address this objective was extremely limited. (2 points subtracted)
### Table 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

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Index test</th>
<th>No. of Participants (Studies)</th>
<th>Effect % (95% CI) Main findings</th>
<th>What do these results mean given a 10% prevalence among suspects being screened for TB?</th>
<th>What do these results mean given a 30% prevalence among suspects being screened for TB?</th>
<th>Quality of Evidence</th>
</tr>
</thead>
</table>
| What is the diagnostic accuracy of IGRAs for active TB? | T-SPOT | Sensitivity 143 (3) | **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 ☹☹☹☹ |
| | | Specificity 97(2) | | | | |
| | | QuantiFERON-G/QuantiFERON-GIT | Sensitivity 84 (3) | **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 ☹☹☹☹ |
<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Pooled sensitivity</th>
<th>Pooled specificity</th>
<th>What do these results mean?</th>
</tr>
</thead>
<tbody>
<tr>
<td>TST</td>
<td>168 (5)</td>
<td>490 (3)</td>
<td>65% (31-99)</td>
<td>90% (82-98)</td>
<td>IGRAs do not perform significantly different from TST</td>
</tr>
<tr>
<td>T-SPOT</td>
<td>132 (2)</td>
<td></td>
<td>Indeterminates/total number of tests</td>
<td>6.82% (0-8%)</td>
<td>What is the proportion of indeterminate IGRAs among children assessed for active TB?</td>
</tr>
<tr>
<td>QFT-G/ QFT-GIT</td>
<td>36 (1)</td>
<td>QFT-GIT 488 (2)</td>
<td>Indeterminates/total number of tests</td>
<td>25% (9/36)</td>
<td>Very Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QFT-G: 9/36 = 25%</td>
<td>QFT-GIT: 20/488=4.1%</td>
<td>(Range of % indeterminates across studies 0-5%)</td>
</tr>
<tr>
<td>Performance of IGRAs in HIV-infected children</td>
<td>T-SPOT</td>
<td>No studies</td>
<td>No studies</td>
<td>TST</td>
<td>No studies</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
<td>-----</td>
<td>------------</td>
</tr>
<tr>
<td>QFT-G</td>
<td>Sensitivity</td>
<td>36 (1)</td>
<td>Sensitivity</td>
<td>QFT-G: 63% (16-100)</td>
<td>• 47% (0-100) if indeterminates counted as FN</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>No studies</td>
<td>Specificity</td>
<td>39% (0-100)</td>
<td>Specificity</td>
</tr>
<tr>
<td>TST</td>
<td>Sensitivity</td>
<td>36 (1)</td>
<td>Sensitivity</td>
<td>39% (0-100)</td>
<td>Specificity</td>
</tr>
<tr>
<td>Performance in children &lt;5yrs</td>
<td>T-SPOT</td>
<td>Sensitivity</td>
<td>134 (3)</td>
<td>Pooled sensitivity</td>
<td>77% (23-100)</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>97 (2)</td>
<td>Pooled specificity</td>
<td>93% (83-100)</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Pooled sensitivity</td>
<td>Pooled specificity</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td><strong>QFT-GIT</strong></td>
<td>36 (1)</td>
<td>327 (1)</td>
<td>35% (30-40)</td>
<td>85% (81-89)</td>
<td></td>
</tr>
<tr>
<td><strong>TST</strong></td>
<td>99 (2)</td>
<td>395 (2)</td>
<td>41% (0-85)</td>
<td>83% (81-86)</td>
<td></td>
</tr>
</tbody>
</table>

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.

Sensitivity and specificity of T-SPOT are higher than of QFT-GIT or TST, but the difference is not significant (overlapping confidence intervals).
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

<table>
<thead>
<tr>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Outcome: Predictive value of IGRAs for active TB</td>
<td>Prospective cohort</td>
<td>Serious (-1)</td>
<td>None (-1)</td>
<td>Serious (-1)</td>
<td>Likely</td>
<td>Very Low</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>(1100 (3)) LMIC: 306 (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Outcome: Sensitivity for active TB (as a surrogate reference standard for LTBI)</td>
<td>Mainly cross-sectional</td>
<td>No serious limitations (-1)</td>
<td>Very Serious (-2)</td>
<td>Serious (-1)</td>
<td>Likely</td>
<td>Very Low</td>
<td>Important (4-6)</td>
</tr>
<tr>
<td>(1523 (18)) LMIC: 1056 (16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Outcome: Concordance with TST</td>
<td>Cross-sectional</td>
<td>No serious limitations (-2)</td>
<td>Serious (-1)</td>
<td>None (-1)</td>
<td>Likely</td>
<td>Very Low</td>
<td>Important (4-6)</td>
</tr>
<tr>
<td>(2158 (15)) LMIC: 401 (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Footnotes

1Quality 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.

81 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).

82 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.

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

84 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.

85 The number of incident TB cases was small in all studies, leading to wide confidence intervals for risk estimates. In the two studies that reported cumulative incidence of TB, the difference in cumulative incidence of TB between IGRA positive and IGRA negative persons was not statistically significant.
Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could not be ruled out. Some degree of publication bias was assumed likely because: 1) literature on 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.

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

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

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

There was significant heterogeneity in sensitivity estimates for both TSPOT (range 54-100%, I² 73%, p<0.002) and QFT-GIT (range 20-92%, I² 78%, p<0.001) in low/middle income countries.

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.

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

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

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

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

Among studies conducted in low/middle income countries, there was significant heterogeneity in estimates of percent concordance between IGRA and TST for TSPOT (range 70-90%, I² 63%, p=0.04). There was only 1 study of QFT-GIT (concordance 91%).

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

Data included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias could be ruled out. Some degree of publication bias was assumed likely because: 1) literature on 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)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>N</th>
<th>Principal Findings</th>
<th>What do these findings mean?</th>
<th>Quality of Evidence</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive value for active TB</td>
<td>1100 (3 studies)</td>
<td>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) QFT-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%).</td>
<td>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.</td>
<td>Very Low  ✦✦✦✦</td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>Sensitivity for active TB (a surrogate reference standard for LTBI)</td>
<td>1523 (18 studies)</td>
<td>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.</td>
<td>In low- and middle-income countries, IGRAs have suboptimal sensitivity for active TB and do not consistently have higher sensitivity than TST.</td>
<td>Very Low  ✦✦✦✦</td>
<td>Important (4-6)</td>
</tr>
<tr>
<td>Concordance with TST</td>
<td>1822 (14 studies)</td>
<td>1) TSPOT: Pooled concordance 77% (95% CI 67-88%). 2) QFT-GIT: 1 study; concordance 91%.</td>
<td>In low- and middle-income countries, IGRAs have moderate concordance with TST.</td>
<td>Very Low  ✦✦✦✦</td>
<td>Important (4-6)</td>
</tr>
</tbody>
</table>
Table 9.  GRADE Evidence Profile: IGRAs for tuberculosis screening of healthcare workers in low- and middle-income countries

<table>
<thead>
<tr>
<th>No of participants (studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Efficacy of preventive therapy based on IGRA test results</td>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>B. Predictive value of IGRA for active TB</td>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
</tr>
<tr>
<td>C. Outcome: Correlation of IGRA results with occupational TB exposure</td>
<td>991 (2) &lt;sup&gt;A1&lt;/sup&gt;</td>
<td>Cross-sectional</td>
<td>No serious limitations &lt;sup&gt;A2&lt;/sup&gt;</td>
<td>No serious Indirectness &lt;sup&gt;A3&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;A4&lt;/sup&gt; (-1)</td>
<td>Serious &lt;sup&gt;A5&lt;/sup&gt; (-1)</td>
<td>Likely &lt;sup&gt;A6&lt;/sup&gt;</td>
<td>Low &lt;sup&gt;⊕⊕〇〇&lt;/sup&gt;</td>
</tr>
<tr>
<td>D. Outcome: Correlation between IGRA conversions and occupational TB exposure</td>
<td>No studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Outcome: Sensitivity for active TB (as a surrogate reference standard for LTBI)</td>
<td>No Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Important (4-6)</td>
</tr>
<tr>
<td>F. Outcome: Concordance between IGRAs and TST (cross-sectional)</td>
<td>1,357 (4) &lt;sup&gt;B1&lt;/sup&gt;</td>
<td>Cross-sectional</td>
<td>No serious limitations &lt;sup&gt;B2&lt;/sup&gt;</td>
<td>Serious &lt;sup&gt;B3&lt;/sup&gt; (-1)</td>
<td>Serious &lt;sup&gt;B4&lt;/sup&gt; (-1)</td>
<td>Serious &lt;sup&gt;B5&lt;/sup&gt; (-1)</td>
<td>Likely &lt;sup&gt;B6&lt;/sup&gt;</td>
<td>Very Low &lt;sup&gt;⊕〇〇〇&lt;/sup&gt;</td>
</tr>
<tr>
<td>G. Outcome: concordance between IGRA and TST conversions (longitudinal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Page | 42
<table>
<thead>
<tr>
<th>216 (1)</th>
<th>C1</th>
<th>Longitudinal</th>
<th>No serious limitations</th>
<th>C2</th>
<th>C3</th>
<th>No serious inconsistency</th>
<th>C4</th>
<th>C5</th>
<th>Very Serious</th>
<th>C6</th>
<th>Likely</th>
<th>C6</th>
<th>Very Low</th>
<th>Important (4-6)</th>
</tr>
</thead>
</table>

Footnotes:

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

A1 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.
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 healthcare 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)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. Participants</th>
<th>Principal Findings</th>
<th>What do these findings mean?</th>
<th>Quality of Evidence</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficacy of preventive therapy based on IGRA test results</td>
<td>No Studies in HCWs</td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
<td></td>
</tr>
<tr>
<td>Predictive value of IGRA for active TB</td>
<td>No Studies in HCWs</td>
<td></td>
<td></td>
<td>Critical (7-9)</td>
<td></td>
</tr>
<tr>
<td>Correlation between IGRA positivity and occupational TB exposure</td>
<td>991 (2 studies)</td>
<td>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.</td>
<td>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.</td>
<td>Critical (7-9)</td>
<td></td>
</tr>
<tr>
<td>Correlation between IGRA conversions and occupational TB exposure</td>
<td>No Studies in HCWs</td>
<td></td>
<td>Critical (7-9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity for active TB (as a surrogate reference standard for LTBI)</td>
<td>No Studies in HCWs</td>
<td></td>
<td>Important (4-6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concordance between TST and IGRAs</td>
<td>1,357 (4 studies)</td>
<td>In low and middle income studies, agreement between IGRA and TST results ranged from 50.2%-81.4%. While IGRA consistently estimated a lower rate, this difference was significant in only 2/4 cases.</td>
<td>Concurdance was fair to poor in low and middle income settings. Both tests provide similar estimates of prevalence in low and middle income countries.</td>
<td>Very Low ⊕⊕⊕</td>
<td>Important (4-6)</td>
</tr>
<tr>
<td>Concordance between IGRA and TST conversions</td>
<td>216 (1 study)</td>
<td>This study found 96% agreement between test conversions (QFT &amp; TST).</td>
<td>IGRA and TST conversions show moderate concordance. Data are limited in all settings.</td>
<td>Very Low ⊕⊕⊕</td>
<td>Important (4-6)</td>
</tr>
</tbody>
</table>
Table 11. GRADE Evidence Profile: Performance of IGRAs for the diagnosis of LTBI in contacts of active TB in low- and middle-income countries

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

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

A1 9 studies were included: 1 study evaluated both T-SPOT.TB and QFT-GIT, 2 studies evaluated T-SPOT.TB, the remaining 6 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.

B1 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 IGRA 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 IGRA; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

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 IGRA 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 IGRA; 3) because a sizeable proportion of IGRA studies have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

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)
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 IGRAs have some level of industry involvement or support, the risk of unpublished negative studies (or delayed publication of negative studies) is not trivial.

2 studies included both IGRAs, 3 studies evaluated only T-SPOT.TB, while the rest evaluated a version of the QFT.

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.

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

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.

Due to heterogeneity in effect estimates concordance could not be pooled. However, effects estimated for individual studies were frequently not significant.
Table 12. GRADE Evidence Profile: Predictive value of commercial IGRA for incident active TB in low- and middle-Income countries

<table>
<thead>
<tr>
<th>No of Participants (Studies)</th>
<th>Study design</th>
<th>Limitations</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication Bias</th>
<th>Quality of Evidence (GRADE)</th>
<th>Importance</th>
</tr>
</thead>
</table>
| 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 compared to IGRA negative results?  
7,392 (3) | Cohort studies | Serious (-1) | Serious (-1) | No serious inconsistency | Very Serious (-2) | Likely | 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) | Cohort studies | Serious (-1) | Serious (-1) | No serious inconsistency | Very Serious (-2) | Likely | 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) | Cohort of TB case-contacts | Serious (-1) | Serious (-1) | Serious (-1) | Very Serious (-2) | Likely | Very low | Important (4-6) |

F. Outcome: Immunological phenotypes of discordant-concordant TST/IGRA pairs and subsequent rates of TB  
5,861 (2) | Cohort studies | Serious (-1) | Serious (-1) | Serious (-1) | Very Serious (-2) | Likely | 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) | Cohort studies | Serious (-1) | Serious (-1) | Serious (-1) | Very Serious (-2) | Likely | 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) |
Quality of evidence was rated as high (no points subtracted), moderate (1 point subtracted), low (2 points subtracted), or very low (>2 points subtracted) based on five criteria: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high quality observational studies and at moderate when these types of studies were absent. We then subtracted one point when there was a serious issue identified or two points when there was a very serious issue identified in any of the criteria used to judge the quality of evidence.

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

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

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

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

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

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

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

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

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

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

Publication bias was not formally assessed, but is deemed likely. See 86.
Only the Zambian study examined if there was an exposure-gradient relationship between baseline quantitative IGRA levels and subsequent rates of TB in those levels. (N refers to numbers included in this stratified analysis)

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

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

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

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

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

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)

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

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

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.

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

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

All 3 studies were included in this evaluation of patient-relevant outcomes. The diagnostic accuracy estimates of sensitivity and specificity etc are surrogates of patient-relevant outcomes important for assessing the frequency and impact of either a false negative or false positive IGRA result at baseline. A falsely positive outcome may result in possible isoniazid preventive therapy (IPT) prescription for a period of 6-9 months, depending on country guidelines. IPT, although safe, is not without serious adverse effects, notably, clinical 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.

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

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

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

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

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, regardless of HIV infection status.

**Setting:** Community-based cohort in a high-burden country, high-risk for TB individuals attending outpatient clinics and school-going adolescents residing in a high-burden country.

**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.

<table>
<thead>
<tr>
<th>Principal Findings</th>
<th>Outcome</th>
<th>N (No. of studies)</th>
<th>Efficacy of preventive therapy based on IGRA results</th>
<th>Prospective predictive value of IGRA for the development of active TB (Do IGRA positive results have a stronger association with subsequent development of active TB compared to IGRA negative results?)</th>
</tr>
</thead>
</table>
| N                  | Principal Findings | 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 1.4-7.4), I² = 50% (p = 0.005). This estimate is not statistically significant—the confidence interval includes the null. Furthermore, the small number of studies, the heterogeneity of populations studied, and the potential for bias in the study design result in a high risk of bias.  
2) IGRA positives results appear to have higher rates of incident TB than IGRA negatives. A pooled IR (IGRA+) = 16.5 (95% CI 11.2-24.7), I² = 98%, p < 0.001. The 95% CI do not overlap suggesting the difference may be significant. However, even in those with positive IGRA results, the vast majority of individuals did not progress to TB disease during follow-up. |
| What do these findings mean? | Quality of Evidence | Very low (SH)  | Critical (7-9) | What do these findings mean? | Importance | Quality of Evidence | Critical (7-9) |

**Importance:** The predictive value of IGRA 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 IGRA in targeting chemoprophylaxis for IGRA-positive individuals.

**Quality of Evidence:** The predictive value of IGRA 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 IGRA in targeting chemoprophylaxis for IGRA-positive individuals.
Interpretation of statistical heterogeneity

<table>
<thead>
<tr>
<th>Predictive value of IGRA for the development of active incident TB compared to the TST (Do IGRA (positive vs. negative) have a stronger statistical association with subsequent active TB than the TST (positive vs. negative))?</th>
<th>7,392 (3)</th>
<th>IGRA+ and TST+ may have a similar strength of association with subsequent TB compared to test negative individuals.</th>
<th>Very low</th>
<th>Critical (7-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) IGRA+: Pooled IRR=3.24 (0.62-4.69); I²=0%, p=0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) TST+: Pooled IRR=2.3 (0.83-3.73); I²=0%, p=0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 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) | 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

<table>
<thead>
<tr>
<th>Studies</th>
<th>Rates of TB during follow-up may be higher in those with double positive TST+/IGRA+ results than in those with double negative results.</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity, Specificity, False positive rates etc for active TB (as surrogates of patient relevant outcomes)

<table>
<thead>
<tr>
<th>Studies</th>
<th>IGRA have moderate sensitivity for subsequent TB in keeping with observed moderate rates. This is not different from the TST.</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>IGRA+ is higher than TST+</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>

TST sensitivity for incident TB was similar at 76% (50-93) and 73% (59-84) for the China and South Africa studies, respectively. Specificity for those studies was 35% (29-41) and 58% (57-58). The proportions that would unnecessarily receive IPT are lower than those for IGRA. 

<table>
<thead>
<tr>
<th>IGRA have moderate sensitivity for subsequent TB in keeping with observed moderate rates. This is not different from the TST.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>IGRA have moderate sensitivity for subsequent TB in keeping with observed moderate rates. This is not different from the TST.</th>
</tr>
</thead>
</table>

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+ and does not necessarily imply a higher risk of TB in those with IGRA+. 

<table>
<thead>
<tr>
<th>IGRA have moderate sensitivity for subsequent TB in keeping with observed moderate rates. This is not different from the TST.</th>
</tr>
</thead>
</table>

Very low

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.

| 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 10006 USA
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, Ikorodu 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
ddggb@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
kulier@who.int

Dr Christian Lienhardt, TBP
lienhardtc@who.int

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

Dr Mario Raviglione, STB
raviglionem@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
Noncommercial culture and drug-susceptibility testing methods for screening patients at risk for multidrug-resistant tuberculosis

Policy statement
Abbreviations

CI        confidence interval
CRI       colorimetric redox indicator
DST       drug-susceptibility testing
GRADE     grades of recommendation assessment, development and evaluation
LJ        Lowenstein-Jensen
MDR       multidrug-resistant
MODS      microscopic observation of drug susceptibility
NRA       nitrate reductase assay
STAG-TB   Strategic and Technical Advisory Group for Tuberculosis
TB        tuberculosis
WHO       World Health Organization
Executive summary

Commercial liquid culture systems and molecular line-probe assays have been endorsed by the World Health Organization (WHO) as gold standards for rapid detection of multidrug-resistant (MDR) tuberculosis (TB); however, because of technical complexity, cost and the requirement for sophisticated laboratory infrastructure, use of these techniques has been limited in many resource-constrained settings. Several noncommercial culture and drug-susceptibility testing (DST) methods have been developed specifically for settings with limited access to sophisticated laboratory infrastructure and technical expertise. Several rapid, inexpensive methods have shown initial promise. The most advanced are microscopic observation of drug susceptibility (MODS), colorimetric redox indicator (CRI) methods, thin-layer agar methods, the nitrate reductase assay (NRA) and mycobacteriophage-based assays.

- **MODS**: a microcolony direct method in liquid culture, based on inoculation of specimens into drug-free and drug-containing media, followed by microscopic examination of early growth;
- **thin-layer agar**: a microcolony direct method on solid culture, based on inoculation of specimens into drug-free and drug-containing media, followed by microscopic examination of early growth;
- **CRI methods**: indirect methods based on the reduction of a coloured indicator added to liquid culture medium on a microtitre plate after exposure of *Mycobacterium tuberculosis* strains to anti-TB drugs in vitro;
- **NRA**: a direct or indirect method based on the ability of *M. tuberculosis* to reduce nitrate, which is detected by a colour reaction; and
- **phage-based assays**: assays in which bacteriophages are used to infect and detect the presence of viable *M. tuberculosis* in clinical specimens and culture isolates.

In 2009, the strength of the evidence for these noncommercial methods was assessed by WHO, following standards appropriate for evaluating both the accuracy of new TB diagnostics and their effect on patients and public health. The results showed that the current evidence is insufficient to recommend the use of thin-layer agar or phage-based assays. There was considered to be sufficient evidence for the use of CRI methods, MODS and NRA under clearly defined programme and operational conditions, in reference laboratories and under strict laboratory protocols, and as an interim solution while capacity for genotypic or automated liquid culture and DST is being developed.

Under these conditions, MODS and NRA are recommended for direct testing of sputum specimens. Together with CRI methods, MODS and NRA are also recommended for indirect DST of *M. tuberculosis* isolates grown in conventional culture. The time to detection of MDR-TB may not necessarily be faster with indirect testing, and none of these methods can detect extensively drug-resistant TB. Conventional culture and DST capacity are therefore still required in all settings.
Policy statement

Noncommercial culture and drug-susceptibility testing methods for screening patients at risk for multidrug-resistant tuberculosis

1. Background

Early detection of drug resistance in TB ensures appropriate treatment regimens for patients and therefore better TB control. The spread of drug-resistant strains of *M. tuberculosis* and the management of patients with drug-resistant disease are formidable obstacles faced by national TB control programmes, which are compounded by a lack of appropriate diagnostic tools and vastly inadequate laboratory capacity.

Rapid methods for DST are crucial, in view of the increasing rates of MDR-TB worldwide and the emergence of extensively drug-resistant TB, with high HIV-associated mortality. Conventional culture and DST methods entail long delays for confirmation of mycobacterial growth and for detection of drug resistance, during which time patients may be inappropriately treated, drug-resistant strains may continue to spread, and resistance may be amplified. Rapid diagnosis of TB and drug resistance therefore has obvious benefits for both patients and public health, including better prognosis, increased survival, prevention of acquisition of further drug resistance and reduced spread of drug-resistant strains to vulnerable populations.

No single test currently satisfies all the demands for ‘quick’, ‘cheap’, and ‘easy’ kits. Commercially available liquid culture systems and molecular line-probe assays for rapid detection of MDR-TB have been endorsed by WHO; however, because of their complexity and cost and the requirement for sophisticated laboratory infrastructure, uptake has been limited in many resource-constrained settings.

Several noncommercial culture and DST methods have been developed, specifically for use in laboratories that lack access to more sophisticated infrastructure and techniques. Of these, MODS, CRI methods, thin-layer agar methods, the NRA and mycobacteriophage-based assays have shown initial promise as being rapid and inexpensive. These tests were recently assessed by WHO, and the results are summarized below.

- **MODS**: a microcolony direct method in liquid culture, based on inoculation of specimens into drug-free and drug-containing media, followed by microscopic examination of early growth;
- **thin-layer agar**: a microcolony direct method on solid culture, based on inoculation of specimens into drug-free and drug-containing media, followed by microscopic examination of early growth;
- **CRI methods**: indirect methods based on the reduction of a coloured indicator added to liquid culture medium on a microtitre plate after exposure of *M. tuberculosis* strains to anti-TB drugs in vitro;
- **NRA**: a direct or indirect method based on the ability of *M. tuberculosis* to reduce nitrate, which is detected by a colour reaction; and
- **phage-based assays**: assays in which bacteriophages are used to infect and detect the presence of viable *M. tuberculosis* in clinical specimens and culture isolates.

There was considered to be sufficient evidence for the use of CRI methods, MODS and NRA under clearly defined conditions, as outlined below.
2. Evidence for policy formulation

2.1 Synthesis of evidence

In September 2009, WHO assessed the evidence base for selected noncommercial culture and DST methods in a systematic, structured way. The first step was a systematic review and meta-analysis of published and unpublished data with standard methods appropriate for studies of diagnostic accuracy. The second step was the convening of an expert group to evaluate the strength of the evidence, recommend operational and logistical considerations for using noncommercial culture and DST methods within national TB control programmes and identify gaps to be addressed by future research. The third step was presentation of draft recommendations to the WHO Strategic and Technical Advisory Group for Tuberculosis (STAG-TB) for endorsement.

In accordance with current WHO standards for evidence assessment in the formulation of policy recommendations, the grades of recommendation assessment, development and evaluation (GRADE) system (1) was used by the Expert Group to assess the findings of the systematic review. This approach provides a systematic, structured framework for evaluating the accuracy of new interventions and their impact on patients and public health.

The Expert Group’s findings and the final GRADE evaluation (2) were presented to STAG-TB in November 2009. STAG-TB recognized the evidence base on CRI methods, MODS and NRA and advised WHO to proceed with policy recommendations under clearly defined premises and conditions. STAG-TB also asked WHO to prepare an overarching policy framework to guide the use of new TB diagnostics, methods and approaches at country level (3).

This document provides a pragmatic summary of the evidence and recommendations related to CRI methods, MODS and NRA, and should be read in conjunction with the detailed findings from the Expert Group report (which include the GRADE tables) and the WHO framework for using TB diagnostics (2). The framework gives the context for use of one or more of the currently approved WHO diagnostic tools and methods in relation to country infrastructure, resources, TB epidemiology and TB policy reform.

The existing TB diagnostic tools are not mutually exclusive: they can be used in various combinations in country screening and diagnostic algorithms, which are highly setting- and resource-specific. Expert laboratory input is therefore needed to define the most cost-effective and efficient algorithms for individual countries, guided by WHO standards (e.g. for laboratory biosafety) and procedures and in the context of overall, integrated, laboratory strengthening.

2.2 Management of declarations of interest

Expert Group members were asked to submit completed declaration of interest forms, which were reviewed by the WHO secretariat before the Expert Group meeting. None of the members declared any conflict of interest. The declaration of interest statements were summarized by the co-chair of the Expert Group meeting at the start of the meeting. No additional declarations were made.

Selected individuals with intellectual or research involvement in the methods reviewed were invited as observers to provide technical input and answer technical questions. These individuals did not participate in the GRADE evaluation and were asked to leave the meeting during the final discussions, when the recommendations were developed. They were also not involved in writing the final meeting report, nor in preparation of the STAG-TB documentation or the final WHO policy statements.

The process for evidence synthesis and policy development was reviewed by the WHO Guidelines Review Committee, and the policy recommendations were approved in June 2010. The target date for review is 2015.
2.3 Premises for adopting new diagnostic tools and methods

- The current gold standards for culture and DST (conventional solid and automated liquid culture and DST systems, molecular line-probe assays) should be phased in and scaled up as a matter of urgency and priority.

- Genotypic (molecular) methods have considerable advantages in scaling up programme management and surveillance of drug-resistant TB because of the speed of diagnosis, standardized testing, potentially high through-put and biosafety.

- Rapid DST methods applied directly on sputum specimens are of greatest benefit for patients and public health. Rapid DST is essential to identify patients at risk for MDR-TB and should be a first priority in screening strategies.

- Rifampicin resistance is a reliable proxy for MDR-TB. Once MDR-TB has been confirmed, additional first- and second-line DST should be performed on the basis of current WHO recommendations and available laboratory capacity.

- Noncommercial methods are less expensive, make laboratories independent of single-test commercial providers and may be an incentive to commercial providers to lower prices. Noncommercial methods are, however, prone to error due to lack of standardization and local variations in methodology.

- The evidence base for selected noncommercial culture and DST methods has been reviewed and the performance of these methods found to be acceptable for use under stringent laboratory protocols in reference or national laboratories in selected settings.

- Techniques and methods for culture and DST are not mutually exclusive. Molecular line-probe assays and the selected noncommercial culture and DST methods are suitable for direct application on smear-positive specimens only. Conventional culture is still required for smear-negative specimens, while conventional DST is needed to detect extensively drug-resistant TB.

3. Summary of results

3.1 Colorimetric redox indicator methods

CRI methods are indirect tests, done on \( M. tuberculosis \) isolates grown from conventional culture. The time to diagnosis of MDR is therefore not faster than with conventional phenotypic DST in liquid culture or genotypic testing with line-probe assays.

CRI methods are highly sensitive (pooled estimate, 98%; 95% confidence interval [CI], 96–99%) and specific (pooled estimate, 99%; 95% CI, 99–100%) for the detection of rifampicin resistance and also isoniazid resistance (pooled sensitivity, 97%; 95% CI, 96–98%; pooled specificity, 98%; 95% CI, 97–99%).

In comparison with the conventional indirect proportion DST method on Lowenstein-Jensen (LJ) medium, CRI methods require additional staff skills, similar equipment but additional consumables that may be difficult to obtain. In comparison with the conventional indirect proportion method in liquid culture medium, CRI methods require similar staff skills, less equipment and consumables that may be readily available.

CRI methods have been standardized, and testing protocols are available (4). The methods require manipulation of concentrated suspensions of mycobacteria, with a high risk for aerosol creation. CRI methods should therefore be performed under laboratory biosafety level 3 conditions.

CRI methods are suitable for use at reference laboratory level; scale-up and decentralization to lower-level laboratories are not recommended.
3.2 Microscopically observed drug susceptibility

MODS can be performed as a direct or an indirect test, by observing microcolony growth and typical cord formation of *M. tuberculosis* in sealed microtitre plates containing liquid culture medium, through an inverted microscope.

Studies on the accuracy of combined (direct and indirect) use showed that MODS is highly sensitive (pooled estimate, 98%; 95% CI, 95–99%) and specific (pooled estimate, 99%; 95% CI, 96–100%) for the detection of rifampicin resistance and slightly less so for isoniazid (pooled sensitivity, 91%; 95% CI, 87–95%). High sensitivity and specificity are retained in direct MODS testing.

Initial concerns about microscopic differentiation of *M. tuberculosis* from nontuberculous mycobacteria were addressed by a revised MODS platform that includes a microtitre well containing *p*-nitrobenzoic acid. *M. tuberculosis* fails to grow in the presence of this compound. Absence of growth combined with cord formation in wells that do not contain *p*-nitrobenzoic acid are therefore indicative of *M. tuberculosis* (similar to current WHO recommendations for use of *p*-nitrobenzoic acid in conventional solid culture and DST methods). Addition of a *p*-nitrobenzoic acid-containing well to the microtitre plate also obviates reopening of the plate and consequently reduces the biosafety risk.

In comparison with the conventional indirect proportion DST method on LJ medium, MODS requires additional staff skills, an additional inverted microscope and additional consumables that may be difficult to obtain. In comparison with the conventional indirect proportion method in liquid culture medium, MODS requires additional staff skills, less equipment and consumables that may be readily available.

MODS has been standardized, with testing protocols and online support available through a dedicated website. The biosafety risk associated with use of the revised MODS platform is considered to be similar to that of conventional culture on solid medium and therefore requires biosafety level 2 precautions.

MODS is suitable for use at reference laboratory level; scaling-up and decentralization to lower-level laboratories is not recommended.

3.3 Nitrate reductase assay

The NRA can be used as a direct test on smear-positive sputum specimens or as an indirect test on *M. tuberculosis* isolates grown from conventional solid culture. Indirect testing with NRA is therefore not faster than conventional phenotypic DST with liquid media.

Studies on combined (direct and indirect) use showed that NRA is highly sensitive (pooled estimate, 97%; 95% CI, 95–98%) and specific (pooled estimate, 100%; 95% CI, 99–100%) for the detection of rifampicin resistance and for isoniazid resistance (pooled sensitivity, 97%; 95% CI, 95–98%; pooled specificity, 99%; 95% CI, 99–100%). The diagnostic accuracy of NRA by direct testing alone did not differ significantly from that of combined testing, although the sensitivity values in individual studies showed wider variation (range, 85–100%), and the data were limited.

The reagents for NRA are nonproprietary and relatively inexpensive. In comparison with the conventional indirect proportion DST method on LJ medium, NRA requires similar staff skills, similar equipment and no additional consumables; in comparison with the conventional indirect proportion method in liquid culture medium, NRA requires fewer staff, equipment and consumables.

Procedures for NRA have been standardized, and testing protocols are available (4). As NRA involves solid culture media, the biosafety requirements are similar to those for conventional solid culture (biosafety level 2); however, addition of NRA reagent requires regular opening of tubes, which poses a significant risk for aerosol generation. This should therefore be done inside an appropriate biological safety cabinet.
NRA is suitable for use at reference laboratory level; scaling-up and decentralization to lower-level laboratories should not be considered until those laboratories have demonstrated proficiency in performing solid culture.

4. Policy recommendations

The GRADE process confirmed that there is sufficient generalizable evidence to recommend the use of selected noncommercial culture and DST methods as an interim solution in resource-constrained settings, under clearly defined programme and operational conditions, while capacity for genotypic or automated liquid culture and DST is being developed.

With due consideration of the issues raised under section 2.2 above, WHO recommends the selective use of one or more of the following noncommercial culture and DST methods in reference laboratories, under strict laboratory protocols:

- **CRI methods**, as indirect tests on *M. tuberculosis* isolates from patients suspected of having MDR-TB, recognizing that the time to detection of MDR-TB is not faster (but less expensive) than with conventional DST methods with commercial liquid culture or molecular line-probe assays;

- **MODS**, as direct or indirect tests for rapid screening of patients suspected of having MDR-TB; and

- **NRA**, as direct or indirect tests for screening patients suspected of having MDR-TB, recognizing that the time to detection of MDR-TB in indirect application is not faster than with conventional DST methods with liquid culture.

To ensure testing standards and consistency, WHO will:

- review existing documents on technical procedures, standard operating procedures and biosafety requirements for each method; and

- prepare and disseminate procedures for internal quality control and external quality assurance for each method.

5. Intended audience

This policy statement should be used to guide use of noncommercial culture and DST methods for TB diagnosis in national TB control programmes. It is intended for use by national TB control programme managers and laboratory directors, in coordination with external laboratory consultants, donor agencies, technical advisors, laboratory technicians, laboratory equipment procurement officers, warehouse managers, other service providers, other relevant government officials and implementing partners involved in country-level TB laboratory strengthening. People responsible for programme planning, budgeting, resource mobilization and training for TB diagnostic services may also benefit from reading this document.

6. References

7. Annexes

Annex 1 : Expert Group Members

**Dr Lucia Barrera**
Head of WHO/IUATLD Supranational Reference Laboratory for Tuberculosis
Jefe Servicio Micobacterias
INEI ANLIS Dr CG Malbran
Velez Sarsfield 563
1281 Buenos Aires
Argentina
e-mail: lbarrera@anlis.gov.ar

**Dr Cesar Bonilla Asalde**
TB Strategy for National Control and Prevention
TB/Leprosy Control Programme
Ministry of Health, Casillia 2117
Los Opalos 225 - Urbanizyacion Benjamin Doig - La Perla Callao
Salaverry Avenue 801 - Jesus Maria
Lima 11, Peru
e-mail: cesarbon@yahoo.es; cbonilla@minsa.gob.pe

**Dr Maryline Bonnet**
Epicentre, c/o MSFCH
78 Rue de Lausanne
116 - CH, 1211 Geneve
Switzerland
e-mail: maryline.bonnet@geneva.msf.org

**Dr Frank J Cobelens**
Deputy director, AMC CPCD Foundation
Centre for Poverty-related Communicable Diseases
affiliated to the Academic Medical Center,
University of Amsterdam
Meibergdreef 9, room T.0.-126
1105 AZ Amsterdam
The Netherlands
e-mail: CobelensF@knvtbc.nl

**Prof Francis Drobniewski**
Head of WHO-IUATLD Supranational Reference Laboratory for Tuberculosis
Health Protection Agency
National Mycobacterium Reference Unit
Head: Clinical TB and HIV Group
Clinical Sciences Research Centre
2 Newark Street, E1 AT-London
United Kingdom
e-mail: f.drobniewski@qmul.ac.uk

**Dr Knut Feldman**
Head: WHO-IUATLD Supranational Reference Laboratory for Tuberculosis
Kuratorium Tuberkulose in der Welt e.V
Robert Koch Allee 2
82131 - Gauting
Germany
e-mail: knut.feldmann@alice-dsl.net

**Dr Christopher Gilpin**
Global Laboratory Manager
International Organization for Migration
17 route des Morillons
CH-1211 Geneva
Switzerland
Email: cgilpin@iom.int

**Dr Willie A Githui**
Kenya Medical Research Institute,
KEMRI
Centre for Respiratory Diseases
Research CRDR
P.O. Box 47855
Nairobi
Kenya
Email: wgithui@yahoo.com
Dr Christy Hanson  
Chair: STP Retooling Task Force  
US Agency for International Development (USAID)  
USAID/BGH/HIDN/ID  
3.7.23, 3rd Floor  
Ronald Reagan Building  
20523-5900 - Washington, DC  
USA  
e-mail: chanson@usaid.gov

Dr Moses Joloba  
Head of National TB Reference Laboratory  
Department of Medical Microbiology  
Microbiology-Pathology Building  
Uganda  
e-mail: moses.joloba@case.edu

Dr Anandi Martin  
Institute of Tropical Medicine  
Mycobacteriology Unit  
Nationalestraat 155  
2000 Antwerp  
Belgium  
Email: amartin@itg.be

Dr Madhukar Pai  
Co-Chair of Subgroup on Evidence Synthesis  
For TB Diagnostics, STP New Diagnostics Working Group, Assistant Professor  
Department of Epidemiology, Biostatistics & Occupational Health  
McGill University  
1020 Pine Avenue West  
Montreal, Canada  
e-mail: madhukar.pai@gill.ca

Dr John Ridderhof  
Chair of STP Global Laboratory Initiative Working Group  
Associate Director for Laboratory Science  
National Center for Preparedness, Detection and Control of Infectious Diseases, CCID  
Centers for Disease Control and Prevention  
1600 Clifton Rd NE, MS-C12  
Atlanta, Georgia 30333  
USA  
e-mail: jridderhof@cdc.gov

Dr Jaime Robledo Restrepo  
Corporacion para Investigaciones Biologicas  
Cra 72A No. 78B-141  
Medellin  
Colombia  
e-mail: jrobledo@cib.org.co

Dr Karen Steingart  
Francis J. Curry National Tuberculosis Center  
University of California, San Francisco  
3180 18th Street, Suite 101  
San Francisco, CA 94110-2028  
USA  
e-mail: karenst@u.washington.edu

Mr Javid Syed  
TB/HIV Project Director  
Treatment Action Group  
611 Broadway, Suite 308  
New York, NY 10012  
USA  
e-mail: javid.syed@treatmentactiongroup.org

Dr Armand van Deun  
Bacteriology Consultant  
International Union Against Tuberculosis and Lung Disease  
Mycobacteriology Unit  
Institute of Tropical Medicine  
Nationalestraat 155  
B-2000 - Antwerpen  
Belgium  
e-mail: avandeun@iuatld.org

Dr Adithya Cattamanchi  
Assistant Professor of Medicine  
San Francisco General Hospital  
Division of Pulmonary and Critical Care Medicine  
Room 5K1  
1001 Potrero Avenue  
San Francisco, California 94110  
e-mail:cattamanchi@medsfgh.ucsf.edu

Dr Jessica Minion  
Medical Microbiology  
MSc Epidemiology  
McGill University  
1020 Pine Avenue West  
Montreal, H3A 1A2, Canada  
e-mail:jessica.minion@mail.mcgill.ca
Dr Juan Carlos Palomino  
Institute of Tropical Medicine  
Mycobacteriology Unit  
Nationalestraat 156  
2000 - Antwerp  
Belgium  
e-mail: palomino@itg.be

Dr J Lucian Davis  
Assistant Professor of Medicine  
San Francisco General Hospital  
Division of Pulmonary and Critical Care Medicine  
Room 5K1, 1001 Potrero Avenue  
San Francisco, California 94110  
Email: lucian.davis@ucsf.edu

Dr C N Paramasivan  
Foundation for Innovative New Diagnostics (FIND)  
71 Avenue Louis-Casai  
1216 Geneva  
Switzerland  
e-mail: CN.paramasivan@finddiagnostics.org

Dr Mark Perkins  
Foundation for Innovative New Diagnostics (FIND)  
71 Avenue Louis-Casai  
1216 Geneva  
Switzerland  
e-mail: mark.perkins@finddiagnostics.org

Dr Eric Adam  
Foundation for Innovative New Diagnostics (FIND)  
71 Avenue Louis-Casai  
1216 Geneva  
Switzerland  
e-mail: eric.adam@finddiagnostics.org

Dr D Moore  
Heredia  
Laboratorio de Investigacion de Enfermedades Infecciosas (UPCH)  
San Martin de Porras  
Lima 31  
Peru  
e-mail: davidajmoore@msn.com
Annex 2: WHO staff members

WHO-STB

Dennis Falzon: falzond@who.int
Jean Iragena: iragenaj@who.int
Paul Nunn: nunnp@who.int
Karin Weyer: weyerk@who.int
Matteo Zignol: zignolm@who.int

WHO-TDR

L Cuevas: cuevasl@who.int
F Moussy: moussyl@who.int
Andrew Ramsay: ramsaya@who.int
S Swaminathan: swaminathans@who.int
Sanne Van Kampen: vankampens@who.int

WHO-OTHER

WHO Guidelines Review Committee Secretariat: grcsecretariat@who.int
Annex 3 : STAG/TB members

**Mr Faruque Ahmed**  
Director, Health Program  
BRAC  
Dhaka  
Bangladesh

**Ms Olayide Akanni**  
Executive Director  
Journalists Against AIDS (JAAIDS) in Nigeria  
Abuja  
Nigeria

**Dr Salah Al Awaidy**  
Director  
Department of Communicable Disease Surveillance & Control  
DGHA, HQ, Ministry of Health  
Muscat  
Oman

**Dr R V Asokan**  
RNTCP Project Coordinator  
Indian Medical Association  
Punalur - Kerala  
India

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

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

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

**Prof Francis Drobniewski**  
Director, Health Protection Agency  
National Mycobacterium Reference Unit  
Institute for Cell and Molecular Sciences, Barts and the London School of Medicine  
London  
United Kingdom

**Dr Wafaa El-Sadr**  
Director, CIDER  
Mailman School of Public Health  
Columbia University  
New York, NY  
USA

**Dr Paula I Fujiwara**  
Director, Dept of HIV and Senior Advisor  
The Union (Paris, France)  
USA

**Mr Case Gordon**  
World Care Council  
Viols en Laval  
France

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

**Dr Mao Tan Eang**  
Advisor to the Minister of Health  
Director, National Center for Tuberculosis and Leprosy Control  
Ministry of Health  
Phnom Penh  
Cambodia  
USA
Dr Yogan Pillay  
Deputy Director General  
Strategic Health Programmes  
Pretoria  
South Africa

Dr Minghui Ren  
Director-General  
Department of International Cooperation  
Ministry of Health  
Beijing  
China

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

Dr Rajendra Shukla  
Joint Secretary  
Ministry of Health & Family Welfare  
New Delhi  
India

Dr Pedro Guillermo Saurez  
TB & HIV/AIDS  
Center for Health Services  
Management Sciences for Health  
Arlington, VA USA

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

Prof Wang Longde  
Dean, School of Public Health  
Peking University  
Beijing  
China

Dr Yuthichai Kasetjaroen  
Director  
Bureau of Tuberculosis  
Ministry of Health  
Bangkok  
Thailand
For further information please contact:

Stop TB Department
World Health Organization
20 Avenue Appia
1211 Geneva 27
Switzerland

Telephone: + 41 22 791 21 11
Facsimile (fax): + 41 22 791 31 11

Website: http://www.who.int/topics/tuberculosis/en/
Same-day diagnosis of tuberculosis by microscopy

Policy statement
Abbreviations

CI confidence interval
GRADE grades of recommendation assessment, development and evaluation
LED light-emitting diode
STAG-TB Strategic and Technical Advisory Group for Tuberculosis
TB tuberculosis
WHO World Health Organization
Executive summary

Direct sputum smear microscopy is the most widely used means for the diagnosis of pulmonary tuberculosis (TB) and is available in most primary health-care laboratories at health-centre level. Smear microscopy may, however, be costly and inconvenient for patients, who have to make multiple visits to health facilities to submit multiple sputum specimens over several days. A number of TB control programmes have found high rates of initial patient default as a result, with high mortality recorded in several resource-limited settings.

It has been shown conclusively that good quality microscopy of two consecutive sputum specimens identifies the vast majority (95–98%) of smear-positive TB patients. Conventional case-finding approaches usually involve microscopic examination of ‘spot-morning’ sputum specimens (in countries with a two-specimen system) or examination of ‘spot-morning-spot’ sputa (in those with a three-specimen approach). The majority of sputum results are therefore available only on the second or third day after the patient presents to a health service.

In 2009, the strength of the evidence for a ‘same-day diagnosis’ approach (microscopy of two consecutive sputum specimens on the same day) was assessed by the World Health Organization (WHO), following standards appropriate for evaluating both the accuracy and the effect of new interventions on patients and public health. It was found that there was sufficient generalizable evidence that a same-day diagnosis approach is equivalent, in terms of diagnostic accuracy, to conventional microscopy case-finding strategies. Nevertheless, significant organizational and programme changes would be required to optimize the advantages of same-day diagnosis, to ensure that laboratory results are received back at the health facility and that patients start treatment on the same day. In addition, there is currently no evidence that early diagnosis of TB results in better uptake of treatment or improved treatment outcomes, so that programmes must closely monitor the effect of revised case-finding strategies.

On the basis of these findings, WHO recommends that countries that have successfully implemented current WHO policy for a two-specimen case-finding strategy consider switching to same-day diagnosis, especially in settings where patients are likely to default from the diagnostic process. Countries that are still using the three-specimen case-finding strategy should consider a gradual change to same-day diagnosis, once WHO-recommended external microscopy quality assurance systems are in place and good-quality microscopy results have been documented. It is essential that programmatic, logistic and operational implications at country level be taken into account in implementation of same-day diagnosis.
Policy statement

Same-day diagnosis of tuberculosis by microscopy

1. Background

Direct sputum smear microscopy is the most widely used means for diagnosing pulmonary TB and is available in most primary health-care laboratories at health-centre level. Smear microscopy may, however, be costly and inconvenient for patients, who have to make multiple visits to health facilities to submit multiple sputum specimens over several days. A number of TB control programmes have found high rates of initial patient default as a result, with high mortality recorded in several resource-limited settings.

It has been shown conclusively that good-quality microscopy of two consecutive sputum specimens identifies the vast majority (95–98%) of smear-positive TB patients. WHO policy on case detection by microscopy was therefore revised in 2007 (1) to recommend a reduction in the number of specimens examined, from three to two in settings with appropriate external quality assurance and documented good-quality microscopy. The case definition was also revised for these settings (1), to one positive smear, defined as one or more acid-fast bacillus in at least 100 microscopic fields. This approach greatly reduces the workload of laboratories, a considerable advantage in countries with a high proportion of smear-negative TB patients due to HIV and/or extrapulmonary disease.

Conventional case-finding approaches usually involve microscopic examination of 'spot-morning' sputum specimens (in countries with a two-specimen system) or examination of 'spot-morning-spot' sputa (in those with a three-specimen approach). Most sputum results are therefore available only on the second or third day after a patient presents to a health service.

Recent research has addressed the diagnostic accuracy of conventional case-finding strategies in comparison with an approach in which two consecutive sputum specimens ('spot-spot') are examined on the same day (so-called 'front-loaded' or 'same-day') and whether patient drop-out from the diagnostic pathway can be reduced as a result.

2. Evidence for policy formulation

2.1 Synthesis of evidence

In September 2009, WHO assessed the evidence for a 'same-day' diagnostic approach in a systematic, structured way. The first step was a systematic review and meta-analysis of published and unpublished data with standard methods appropriate for studies of diagnostic accuracy. The second step was the convening of an expert group to evaluate the strength of the evidence, recommend operational and logistical considerations for use of same-day diagnosis in national TB control programmes and identify gaps to be addressed by future research. The third step was presentation of draft recommendations to the WHO Strategic and Technical Advisory group for Tuberculosis (STAG-TB) for endorsement.

In accordance with current WHO standards for evidence assessment in the formulation of policy recommendations, the grades of recommendation assessment, development and evaluation (GRADE) system (2) was used by the Expert Group to assess the findings of the systematic reviews. This approach provides a systematic, structured framework for evaluating both the accuracy of new interventions and their impact on patients and public health.
The Expert Group’s findings and the final GRADE evaluation (1) were presented to STAG-TB in November 2009. STAG-TB recognized that the evidence showed that examining two specimens in 1 day was equivalent, in terms of diagnostic accuracy, to existing case-finding strategies but acknowledged that significant organizational and programme changes would be required to optimize the advantages of same-day diagnosis (3). STAG-TB subsequently advised WHO to proceed with policy recommendations on same-day diagnosis and asked WHO to prepare an overarching policy framework to guide the use of new TB diagnostics, methods and approaches at country level (3).

This document provides a pragmatic summary of the evidence and recommendations for same-day diagnosis. It should be read in conjunction with the detailed findings of the Expert Group (which include the GRADE tables) and the WHO framework for using TB diagnostics (1). The framework gives the context for use of one or more of the currently approved WHO diagnostic tools and methods in relation to country infrastructure, resources, TB epidemiology and TB policy reform.

The existing TB diagnostic tools are not mutually exclusive: they can be used in various combinations in country screening and diagnostic algorithms, which are highly setting- and resource-specific. Expert laboratory input is therefore needed to define the most cost-effective and efficient algorithms for individual countries, guided by WHO standards (e.g. for laboratory biosafety) and procedures and in the context of overall, integrated, laboratory strengthening.

2.2 Management of declarations of interest

Expert Group members were asked to submit completed declaration of interest forms, which were reviewed by the WHO secretariat before the Expert Group meeting. None of the members declared any conflict of interest. The declaration of interest statements were summarized by the co-chair of the Expert Group meeting at the start of the meeting. No additional declarations were made.

Selected individuals with intellectual or research involvement in same-day diagnosis were invited as observers to provide technical input and answer technical questions on the methods. These individuals did not participate in the GRADE evaluation and were asked to leave the meeting during the final discussions, when the recommendations were developed. They were also not involved in writing the final meeting report, nor in preparation of the STAG-TB documentation or the final WHO policy statements.

The process for evidence synthesis and policy development was reviewed by the WHO Guidelines Review Committee, and the policy recommendations were approved in June 2010. The target date for review is 2015.

3. Summary of results

The results of seven studies involving 7308 patients were reviewed.

*Same-day diagnosis* (*‘spot-spot’*) *versus the conventional strategy* (*‘spot-morning’*), *with two specimens and direct Ziehl-Neelsen microscopy*

Same-day diagnosis was on average 2.8% less sensitive than the conventional approach (95% confidence interval [CI], –5.2% – +0.3%), indicating that this strategy would be no more than 5% worse than the conventional approach. The specificity of the two approaches (with culture as the reference standard) was identical (98%; 95% CI, 97–99%).

As expected, spot specimens were more likely than morning specimens to give low-positivity results, indicating the need for strict internal quality control during smear preparation and meticulous examination of smears.
One large randomized controlled trial (6068 patients at four geographical sites) included data on patient loss to follow-up. Patients assigned to same-day diagnosis were more likely to submit both specimens (drop-out, 2%) than patients screened conventionally (drop-out, 5.8%).

**Same-day diagnosis ('spot-spot morning') versus the conventional strategy ('spot-morning-spot') with three specimens and direct Ziehl-Neelsen microscopy**

The 'spot-spot-morning' strategy was 3% more sensitive (71%; 95% CI, 65–77%) than the 'spot-morning-spot' approach (68%; 95% CI, 63–73%), although this difference was not statistically significant. The specificity (with culture as reference standard) was also similar, at 98% (95% CI, 96–99%) and 99% (95% CI, 97–99%), respectively.

In the same randomized controlled trial described above, patients assigned to the 'spot-spot-morning' approach were more likely to submit the third specimen (drop-out, 5.9%) than those assigned to the spot-morning-spot strategy (drop-out, 6.7%).

**Same-day diagnosis versus conventional strategies in HIV-infected patients**

The above-mentioned randomized controlled trial included data on the performance of the two strategies in a subset of 586 HIV-infected patients. The study was underpowered for this subanalysis, and the results should therefore be interpreted with caution.

Overall, HIV coinfection seemed to reduce the sensitivity of microscopy, independently of the approach used. In the three-specimen strategy, the sensitivity decreased from 81.3% for HIV-negative to 71.4% for HIV-positive patients screened with the 'spot-spot-morning' approach, and from 68.4% to 51.9% for those screened with the 'spot-morning-spot' strategy. These differences were not statistically significant.

In the two-specimen strategy, the sensitivity decreased from 76.7% for HIV-negative patients to 66.7% for HIV-positive patients screened with the 'spot-spot' approach, and from 68.4% to 50.0% for those screened with the 'spot-morning' approach. These differences were not statistically significant.

**Same-day diagnosis versus conventional strategies with light-emitting diode fluorescence microscopy**

The randomized controlled trial mentioned above compared light-emitting diode (LED) microscopy with conventional fluorescence microscopy in a subset of 2303 patients. The study was underpowered for this subanalysis, and the results should therefore be interpreted with caution.

Overall, LED microscopy performed as well as conventional fluorescence microscopy, irrespective of the case-finding approach. With the two-specimen strategy, the sensitivity of LED microscopy (68%; 95% CI, 62–74%) did not differ significantly from that of conventional fluorescence microscopy (72%; 95% CI, 66–77%). The specificity of LED microscopy (95%; 95% CI, 93–96%) was also not statistically different from that of conventional fluorescence microscopy (94%; 95% CI, 92–95%).

In the three-specimen strategy, the sensitivity of LED microscopy (75%; 95% CI, 69–80%) did not differ significantly from that of conventional fluorescence microscopy (74%; 95% CI, 68–79%). The specificity was also similar, at 92% (95% CI, 91–94%) for LED and 93% (95% CI, 91–94%) for conventional fluorescence microscopy.
**Same-day diagnosis with two smears from a single sputum specimen**

Two studies included secondary analyses of the yield from two smears prepared from the same sputum specimen (n = 1849). Overall, the quality of evidence from both studies was rated as low, and the data were therefore excluded.

### 4. Policy recommendations

The GRADE process showed that there was sufficient generalizable evidence that same-day diagnosis (microscopy of two consecutive spot-spot sputum specimens) is equivalent, in terms of diagnostic accuracy, to conventional case-finding strategies by microscopy.

As stated in previous WHO policy guidance, most patients with smear-positive TB are identified by examination of the first two sputum specimens. The proposed same-day diagnostic approach would allow initiation of anti-TB treatment on the same day, which would contribute to lowering patient-related costs and might reduce patient loss in the diagnostic pathway.

Significant organizational and programme changes would, however, be required in order to optimize the advantages of same-day diagnosis, i.e. ensuring that laboratory results are received back at the health facility and that patients start treatment on the same day. In addition, there is currently no evidence that early diagnosis of TB results in better uptake of treatment or improved treatment outcomes, so that programmes must closely monitor the effect of revised case-finding strategies.

The available evidence arose from carefully conducted studies, and replication of their findings under routine programme conditions will depend heavily on key health service and operational considerations, as for any new diagnostic approach or tool.

**WHO therefore recommends that:**

- countries that have implemented the current WHO policy for two-specimen case-finding consider switching to same-day diagnosis, especially in settings where patients are likely to default from the diagnostic pathway;

- countries that are still using the three-specimen case-finding strategy consider a gradual change to same-day diagnosis, once WHO-recommended external microscopy quality assurance systems are in place and good-quality microscopy results have been documented;

- the change to same-day diagnosis be preceded by a detailed assessment of the programme, logistical and operational implications at country level and supported by a carefully phased implementation plan that includes the following:

  - Service providers should be able to initiate or refer patients for treatment on the same day of consultation. This will require training of health staff responsible for requesting sputum smear microscopy and instructing patients on sputum collection and people responsible for registering patients and initiating treatment.

  - Laboratory operations and procedures should be realigned with sputum collection and reporting of results on the same day, within the constraints of existing human resources and laboratory workload. Particular attention must be given to internal quality control and external quality assurance of microscopy procedures.
- The contact time between infectious patients and vulnerable groups attending the same facility should be minimized, especially in settings with a high HIV prevalence or a high burden of drug-resistant TB. Separation and rapid triage of coughing patients is especially important to reduce the risk for TB transmission in health-care settings.
- Monitoring of patient drop-out between laboratory and patient registers and of trends in case detection and treatment outcomes is essential.

WHO will assist countries in implementing same-day diagnosis by facilitating, with partners and technical agencies, a coordinated approach to revised case-finding strategies at country level.

5. Intended audience
This policy statement should be used to guide implementation of same-day diagnosis in TB case-finding by microscopy within national TB control programmes. It is intended to be used by national TB control programme managers and laboratory directors, in coordination with external programme consultants, donor agencies, technical advisors, health-care and laboratory staff, other service providers, other relevant government officials and individuals responsible for TB training activities.

6. References
7. Annexes

Annex 1: Expert Group Members

Dr Maryline Bonnet
Epicentre, c/o MSFCH
78 Rue de Lausanne
116 - CH, 1211 Geneve
Switzerland
e-mail: maryline.bonnet@geneva.msf.org

Dr Saidi Egwaga
National TB/Leprosy Programme
Ministry of Health, TB/Leprosy Unit
PO Box 9083
Dar Es Salaam
Tanzania
e-mail: tantci@intafrica.com

Dr Bernard Fourie
Chief Scientific Officer and Director of South African Operations
Medicine in Need Inc. USA and Medicine in Need South Africa (Pty) Ltd
PO Box 12660, Queenswood 0121
Pretoria
South Africa
e-mail: bfourie@medicineinneed.org

Dr Christy Hanson
Chair of STP Retooling Task Force
US Agency for International Development (USAID), USAID/BGH/HIDN/ID
3.7.23, 3rd Floor, Ronald Reagan Bldg
20523-5900 - Washington, DC
USA
e-mail: chanson@usaid.gov

Dr Moses Joloba
Head of National TB Reference Laboratory
Department of Medical Microbiology
Microbiology-Pathology Building
Uganda
e-mail: moses.joloba@case.edu

Dr Marija Joncevska
Regional laboratory specialist
Project HOPE Central Asia
162 Kunaeva st.
050010 Almaty
Kazakhstan
e-mail: mjoncevska@projecthope.kz

Prof Paul R Klatser
Co-Chair of Subgroup on Optimizing Smear Microscopy, STP New Diagnostics Working Group
Head of Department
KIT Biomedical Research
Royal Tropical Institute
Meibergdreef 39
1105 AZ Amsterdam
The Netherlands
e-mail: p.klatser@kit.nl

Dr Madhukar Pai
Co-Chair of Subgroup on Evidence Synthesis for TB Diagnostics, STP New Diagnostics Working Group
Assistant Professor
Department of Epidemiology, Biostatistics & Occupational Health
McGill University
1020 Pine Avenue West
Montreal, H3A 1A2
Canada
e-mail: madhukar.pai@mcgill.ca

Dr John Ridderhof
Chair of STP Global Laboratory Initiative Working Group
Associate Director for Laboratory Science
National Center for Preparedness, Detection and Control of Infectious Diseases, CCID Centers for Disease Control and Prevention
1600 Clifton Rd NE,
MS-C12
Atlanta, Georgia 30333
USA
e-mail: jridderhof@cdc.gov

Dr Bertie Squire
Co-Chair of Subgroup on TB Diagnostics and Poverty, STP New Diagnostics Working Group
Liverpool School of Tropical Medicine
Pembroke Place
Liverpool
UNITED KINGDOM
Tel: 0044 151 705 3101
e-mail: s.b.squire@liverpool.ac.uk
Dr Karen Steingart  
Francis J. Curry National Tuberculosis Center 
University of California, San Francisco 
3180 18th Street, Suite 101 
San Francisco, CA 94110-2028 
USA 
e-mail: karenst@u.washington.edu

Mr Javid Syed  
TB/HIV Project Director 
Treatment Action Group 
611 Broadway, Suite 308 
New York, NY 10012 
USA 
e-mail: javid.syed@treatmentactiongroup.org

Dr Armand van Deun  
Bacteriology Consultant 
International Union Against Tuberculosis and Lung Disease 
Mycobacteriology Unit 
Institute of Tropical Medicine 
Nationalestraat 155 
B-2000 - Antwerpen 
Belgium 
e-mail: avandeun@iuatld.org

Dr Adithya Cattamanchi  
Assistant Professor of Medicine 
San Francisco General Hospital 
Division of Pulmonary and Critical Care Medicine 
Room 5K1 
1001 Potrero Avenue 
San Francisco, California 94110 
Email: acattamanchi@medsfgh.ucsf.edu

Dr Jessica Minion  
Medical Microbiology 
MSC Epidemiology 
McGill University 
1020 Pine Avenue West 
Montreal, H3A 1A2 
Canada 
e-mail: jessica.minion@mail.mcgill.ca

Dr Catharina Boehme  
Foundation for Innovative New Diagnostics (FIND) 
71 Avenue Louis-Casai 
1216 Geneva 
Switzerland 
e-mail: catharina.boehme@finddiagnostics.org

Dr C N Paramasivan  
Foundation for Innovative New Diagnostics (FIND) 
71 Avenue Louis-Casai 
1216 Geneva 
Switzerland 
e-mail: CN.paramasivan@finddiagnostics.org

Dr Eric Adam  
Foundation for Innovative New Diagnostics (FIND) 
71 Avenue Louis-Casai 
1216 Geneva 
Switzerland 
e-mail: eric.adam@finddiagnostics.org

Dr Richard Bumgarner  
Independent Consultant 
Health Economics and Finance 
Program and Institutional Evaluations 
Tuberculosis Control Adviser 
1715 Abbey Oak Drive 
Vienna, VA 22182 
USA 
e-mail: richardbumgarner@gmail.com
Annex 2 : WHO staff members

WHO-STB

Leopold Blanc: blancl@who.int
Jean Iragena: iragenaj@who.int
Paul Nunn: nunnp@who.int
Karin Weyer: weyerk@who.int
Matteo Zignol: zignolm@who.int
Sanne Van Kampen: vankampens@who.int

WHO-TDR

L Cuevas: cuevasl@who.int
F Moussy: moussyf@who.int
Andrew Ramsay: ramsaya@who.int
S Swaminathan: swaminathans@who.int

WHO-OTHER

WHO Guidelines Review Committee Secretariat:
grcsecretariat@who
Annex 3: STAG-TB members

Mr Faruque Ahmed
Director, Health Program
BRAC
Dhaka
Bangladesh

Ms Olayide Akanni
Executive Director
Journalists Against AIDS (JAAIDS) in Nigeria
Abuja
Nigeria

Dr Salah Al Awaidy
Director
Department of Communicable Disease Surveillance & Control
DGHA, HQ, Ministry of Health
Muscat
Oman

Dr R. V. Asokan
RNTCP Project Coordinator
Indian Medical Association
Punalur - Kerala
India

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

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

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

Prof Francis Drobniewski
Director, Health Protection Agency
National Mycobacterium Reference Unit
Institute for Cell and Molecular Sciences,
Barts and the London School of Medicine
London
United Kingdom

Dr Wafaa El-Sadr
Director, CIDER
Mailman School of Public Health
Columbia University
New York, NY
USA

Dr Paula I Fujiwara
Director, Dept of HIV and Senior Advisor
The Union (Paris, France)
USA

Mr Case Gordon
World Care Council
Viols en Laval
France

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

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

Dr Megan Murray
Associate Professor of Epidemiology
Harvard University School of Public Health
Department of Epidemiology
Boston, MA
USA
Dr Yogan Pillay
Deputy Director General
Strategic Health Programmes
Pretoria
South Africa

Dr Minghui Ren
Director-General
Department of International Cooperation
Ministry of Health
Beijing
China

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

Dr Rajendra Shukla
Joint Secretary
Ministry of Health & Family Welfare
New Delhi
India

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

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

Prof Wang Longde
Dean, School of Public Health
Peking University
Beijing
China

Dr Yuthichai Kasetjaroen
Director
Bureau of Tuberculosis
Ministry of Health
Bangkok
Thailand
For further information please contact:

**Stop TB Department**
World Health Organization
20 Avenue Appia
1211 Geneva 27
Switzerland

Telephone: + 41 22 791 21 11
Facsimile (fax): + 41 22 791 31 11

Website: [http://www.who.int/topics/tuberculosis/en/](http://www.who.int/topics/tuberculosis/en/)