Critical Evaluation of Tuberculosis Diagnostic Tests in Low- and High-Burden Settings

by

John Metcalfe

Dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the Graduate Division

of the

University of California, Berkeley

Committee in charge:

Arthur Reingold, Chair
Lee Riley
Maya Petersen
Eva Harris

Fall 2012

Abstract

Critical Evaluation of Tuberculosis Diagnostic Tests in Low- and High-Burden Settings

by

John Metcalfe

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Arthur Reingold, Chair

Tuberculosis (TB) is the second leading cause of death from an infectious disease worldwide and remains a major public health challenge, particularly in resource limited settings. Given the effectiveness of current treatment regimens, enhanced detection of disease in both low- and high-burden settings will be critical in making progress towards TB elimination.

As TB case rates have declined in high-income settings, TB control has centered on finding and treating individuals with non-infectious latent TB infection (LTBI) in order to prevent reactivation infectious TB disease. Since 2005, the Centers for Disease Control and Prevention has recommended use of interferon-γ release assays (IGRAs), *in vitro* immuno-diagnostic tests that measure effector T-cell mediated interferon-gamma (IFN-γ) response to *M. tuberculosis* specific antigens, could be used for targeted screening of LTBI in all circumstances in which the tuberculin skin test (TST) is used. We found that higher quantitative IFN-γ results were associated with active tuberculosis and added clinical value to a prediction model incorporating conventional risk factors; however, in all settings and especially within low- and middle-income countries, IGRAs are inadequate rule-out or rule-in tests for active TB. Although IGRAs are widely used in high-income countries and numerous studies have evaluated their diagnostic performance for detection of LTBI, there is limited data on the precision of IGRA results. In the largest precision study of an IGRA to date, we found considerable variability in TB response measured by QuantiFERON-TB Gold In-Tube (QFT-GIT, Cellestis, Australia); test results should be interpreted cautiously among low-risk individuals with positive TB response less than 0.59 IU/ml.

In contrast to low TB burden, high income settings, TB control in low income settings focuses on early detection and treatment of individuals with active, infectious TB. In the WHO African Region, the incidence of multidrug resistant TB (MDR-TB) has tripled in the past 20 years and poses a major risk to regional TB control programs. Accurate, timely, and affordable drug susceptibility testing for patient management and in support of surveillance programs is urgently needed. In a politically unstable, high HIV prevalence region of southern Africa, we validated use of a low-cost, accelerated phenotypic method for MDR-TB detection, and provide the first report of the prevalence of MDR-TB from this country in 17 years.

Table of Contents

Introduction and Rationale……………………………………………………………..1

Use of Interferon-γ Release Assays in the Diagnosis of Active Tuberculosis

Chapter 1. Evaluation of Quantitative Interferon-γ Response for Risk Stratification of Patients Suspected of Having Active Tuberculosis……………..3

Chapter 2. Interferon-Gamma Release Assays for Diagnosis of Active Pulmonary Tuberculosis in Low- and Middle-Income Countries: Systematic Review and Meta-Analysis…………………………………………………………………..17

Use of interferon-γ release assays for latent *M. tuberculosis* infection in low-incidence settings

Chapter 3. Test Variability of the Quantiferon-TB Gold In-Tube Assay in Clinical Practice…………………………………………………………………………38

Detection of Multidrug Resistant Tuberculosis in High-burden, High HIV-prevalence Regions

Chapter 4. Use of Microscopic-observation Drug Susceptibility Assay among Patients Suspected of Having Drug-Resistant Tuberculosis in Harare, Zimbabwe……………………………………………………………………51

Chapter 5. Evaluation of Patients Suspected of Having Multidrug Resistant Tuberculosis in Zimbabwe………………………………………………………62

Concluding Remarks……………………………………………..………………...67

References…………………………………………………………………………...68

Introduction and Rationale

Tuberculosis (TB) is an infectious airborne disease caused by the bacillus *Mycobacterium tuberculosis*, one of the oldest known pathogens in humans. TB is the second leading cause of death from an infectious disease worldwide (after the human immunodeficiency virus (HIV)), and remains a major public health problem, particularly in resource limited settings. A six-month regimen of four first-line drugs (isoniazid, rifampicin, ethambutol and pyrazinamide) cures ~90% of cases and has been available since the 1980s; yet, in 2011 the global burden of TB remained substantial, with nearly 9 million new cases and 1.4 million deaths from TB.

In high-income settings such as the United States, the incidence of TB has declined steadily since the early 1990s. However, declines have occurred disproportionately among U.S.-born individuals, so that the majority of TB (and TB/HIV) cases in the U.S. now occur among immigrants from high TB burden countries. As TB case rates have declined, a major goal of TB control in high-income settings has centered on finding and treating individuals with non-infectious latent TB infection (LTBI), in order to prevent reactivation infectious TB disease. Traditionally, the tuberculin skin test (TST), the oldest clinical diagnostic test still in use today has been used to identify individuals with LTBI. Since 2005, the Centers for Disease Control and Prevention (CDC) has endorsed use of interferon-γ release assays (IGRAs), *in vitro* immunodiagnostic tests that measure effector T cell-mediated interferon-gamma response to synthetic *Mycobacterium tuberculosis*–specific polypeptides. In this dissertation, I examine two major issues arising out of the widespread use of IGRAs. First, IGRAs, as with the TST, have been inappropriately employed by clinicians in algorithms for diagnosis of active TB. Commercial IGRAs have been a lucrative market for the private sector in countries with a high burden of TB disease and emerging economies (e.g., India, South Africa, Brazil, and China). Two reports in this dissertation (Chapters 1 and 3) address this issue in low- and high-TB burden settings, respectively. I presented this work at a World Health Organization (WHO) Expert Group meeting, resulting in a recommendation by the WHO against the use of IGRAs for active TB diagnosis in high burden settings. Second, when appropriately used for LTBI diagnosis, there is lack of convincing evidence in support of defining test cut points for positivity (in low-risk populations) or test "conversion" (in any population) to guide initiation of treatment for LTBI, an important issue given the non-trivial nature of LTBI treatment (the most common treatment regimen entails nine months of a potentially hepatotoxic medication). Using a linear mixed effects model fit to the numerical IFN-γ values, Chapter 3 seeks to provide data concerning possible cut points that can be used in clinical practice.

In contrast to high income settings, TB control in low income settings focuses predominantly on early detection and treatment of individuals with active, infectious TB; in these settings, despite high-quality evidence supporting the use of isoniazid preventive therapy among persons with HIV co-infection, treating LTBI is considered low-priority. Although new cases of TB globally have been falling for several years at a slow but steady rate (i.e., a decrease of 2.2% between 2010 and 2011), multidrug resistant (MDR) TB is emerging as a major challenge. The number of

cases of MDR-TB worldwide is increasing, surpassing 300,000 incident cases in 2011. Yet, less than one in five of these cases are reported to national TB programs, and fewer than 0.5% are treated according to standards of care in the United States.

In the WHO African Region, the incidence of MDR has tripled in the past 20 years and poses a major risk to regional TB control programs. Yet, fewer than half of the 46 countries in the WHO African Region have provided representative data concerning the prevalence of drug resistance among *M. tuberculosis* strains, and only ten have reported such data since 2000. Accurate, timely, and affordable drug susceptibility testing (DST) of *M. tuberculosis* in support of surveillance and patient management is urgently needed in high burden countries. Chapter 4 of this dissertation deals with identification of MDR-TB using low-cost, accelerated phenotypic methods in a politically unstable region of southern Africa with a high prevalence of HIV infection, and Chapter 5 presents the first report of the prevalence of MDR-TB from this country in 17 years.

Chapter 1: Evaluation of Quantitative Interferon-γ Response for Risk Stratification of Patients Suspected of Having Active Tuberculosis

Scientific Knowledge on the Subject

The role of interferon-γ release assays (IGRAs) in the evaluation of patients suspected of having active tuberculosis is controversial. Whether IGRAs improve classification of individuals who are suspected of having tuberculosis and whose sputum smear for acid fast bacilli is negative into clinically relevant risk categories has not been examined.

What This Study Adds to the Field

Quantitative interferon-γ levels measured by QuantiFERON®-TB Gold improves risk stratification of smear-negative active tuberculosis suspects when added to objective clinical and demographic risk factors. However, this benefit is attenuated when the judgment of experienced clinicians is also taken into account.

Abstract

Rationale: The contribution of interferon-γ release assays (IGRAs) to appropriate risk stratification of patients suspected of having tuberculosis has not been studied.

Objective: To determine whether addition of quantitative IGRA results to a prediction model incorporating clinical criteria improves risk stratification of smear-negative tuberculosis suspects.

Methods: Clinical data from tuberculosis suspects evaluated by the San Francisco Department of Public Health Tuberculosis Control Clinic from March 2005 to February 2008 were reviewed. We excluded tuberculosis suspects who were acid fast bacilli smear-positive, HIV-infected, or under 10 years of age. We developed a clinical prediction model for culture-positive disease and examined the benefit of adding quantitative interferon-γ results obtained using QuantiFERON®-TB Gold.

Main Results: Of 660 patients meeting eligibility criteria, 65 (10%) had culture-proven tuberculosis. The odds of active tuberculosis increased by 7% (95% CI 3-11%) for each doubling of interferon-γ level. The addition of quantitative interferon-γ results to objective clinical data significantly improved model performance (c-statistic 0.71 vs. 0.78, $p<0.001$) and correctly reclassified 32% of tuberculosis suspects (95% CI 11-52%, $p<0.001$) into higher or lower risk categories. However, quantitative interferon-γ results did not significantly improve appropriate risk reclassification beyond that provided by clinician assessment of risk (5%, 95% CI -7 to +22%, $p=0.14$).

Conclusion: Higher quantitative interferon-γ results were associated with active tuberculosis, and added clinical value to a prediction model incorporating conventional risk factors. While this

benefit may be attenuated within centers with highly experienced clinicians, the predictive accuracy of quantitative interferon-γ levels should be evaluated in other settings.

Introduction

Interferon-γ release assays (IGRAs) are *in vitro* immuno-diagnostic tests that measure effector T-cell mediated interferon-gamma (IFN-γ) response to *M. tuberculosis* specific antigens. IGRAs are as sensitive and more specific than the tuberculin skin test for detecting latent tuberculosis infection (LTBI) (1, 2) and have better correlation with gradient of *M. tuberculosis* exposure (3-8). In 2005, the Centers for Disease Control and Prevention recommended that QuantiFERON TB-Gold (QFT-G, Cellestis, Carnegie, Australia) - the first FDA-approved, commercially available IGRA in widespread use - could be used for targeted screening of LTBI in all circumstances in which the tuberculin skin test (TST) is used (9).

While the advantages of IGRAs in diagnosing LTBI are well established, their role in evaluating patients suspected of having tuberculosis suspects remains unclear. IGRAs have variable, though often suboptimal, sensitivity and specificity for diagnosing active tuberculosis (1, 2, 10-16). To date, with the exception of studies examining these assays in parallel with the TST (11, 17), IGRAs have not been considered in light of conventional risk factors for active disease. In addition, whether IGRAs improve prediction of an individual patient's likelihood of having active tuberculosis has not been examined.

Acid fast bacilli (AFB) smear-positive tuberculosis suspects can often be triaged with relative ease. However, in suspects whose sputa or other tissue are smear-negative for AFB, clinicians use demographic and clinical risk factors, symptoms, and chest radiograph findings to classify patients into low, intermediate, or high risk categories for active tuberculosis. Patients classified as being at high risk are typically initiated on anti-tuberculosis therapy, whereas treatment is withheld for low risk patients. In this study, we use novel risk reclassification methods (18) to assess whether addition of quantitative IFN-γ response measured by QuantiFERON TB-Gold (QFT-G, Cellestis, Carnegie, Australia) to routine clinical evaluation improves risk stratification of smear-negative pulmonary and extrapulmonary tuberculosis suspects.
Some of the results of these studies have been previously reported in abstract form (19).

Methods

Study Population

The San Francisco Department of Public Health (SFDPH) operates a central Tuberculosis Control Clinic that routinely screens contacts, immigrants and refugees, as well as hospitalized, private, and community health center patients for LTBI and active tuberculosis in accordance with American Thoracic Society (ATS), Centers for Disease Control and Prevention (CDC) and Infectious Diseases Society of America (IDSA) guidelines (20). The target population for this study includes AFB smear negative pulmonary or extrapulmonary tuberculosis suspects who presented to the SFDPH Tuberculosis Control Clinic between March 2005 and February 2008 and had QFT-G performed as part of their initial evaluation. Patients with QFT-G results that were (1) indeterminate; (2) performed greater than 14 days prior to or 14 days following their initial clinic visit; or (3) performed more than 7 days into a course of tuberculosis treatment were

excluded. In addition, patients aged less than 10 years (in whom adult-type, non-paucibacillary disease is uncommon) (21, 22); with a positive AFB smear examination; known diagnosis of active tuberculosis at presentation; known HIV-infection; or with a final diagnosis of culture-negative tuberculosis were excluded. Demographic and clinical information was extracted from the SFDPH Tuberculosis Control Clinic electronic database. QFT-G assays were performed at the SFDPH laboratory according to the manufacturer's instructions (23). Patients were considered to have active tuberculosis only when there was culture confirmation of *M. tuberculosis*. The study protocol was approved by the Committee for Human Research at the University of California, San Francisco.

Statistical Methods

The analysis included the following steps. *First*, a novel model selection procedure, the Deletion/Substitution/Addition (DSA) algorithm (24), was used to select the optimal prediction model for culture-confirmed tuberculosis using standard clinical and demographic variables; the following limits for the DSA algorithm were set: third order polynomials, second order interaction terms, and maximum model size of 10 variables. Covariates were considered for inclusion in the model based on previous studies of risk factors for active tuberculosis and included the following: age, gender, foreign birth, homelessness, contact with an active tuberculosis case, previous history of active tuberculosis, predisposing medical condition (e.g., diabetes mellitus, silicosis, cancer, or condition requiring use of immunosuppressive medications), symptoms of active tuberculosis (e.g., night sweats, weight loss, or cough), and findings on initial chest radiograph. We also performed a secondary analysis in which clinician suspicion for active disease at the time of patient evaluation (classified as low, intermediate, or high) was added to the baseline clinical prediction model generated by DSA.

Second, patients were classified as low (< 5%), intermediate (5-20%), or high risk (>20%) for active tuberculosis based on the probability assigned by the baseline clinical prediction model (this classification was distinct from clinician suspicion for active disease described above). The lower and upper probability cut-points for tuberculosis risk categories were selected based on the assumption that empiric tuberculosis treatment would be withheld when the probability of active tuberculosis was below the lower risk threshold (low risk) and prescribed when the probability was above the higher risk threshold (high risk). Sensitivity analyses were performed using alternate low and high risk thresholds of 2.5% and 10%, and 10% and 30%.

Third, quantitative IFN-γ results were added to the clinical prediction model. Performance of the prediction models with and without quantitative IFN-γ results were then compared using receiver-operator characteristic (ROC) analysis (25) and net reclassification index (NRI) (18). Based upon the pre-specified risk thresholds, the NRI reflects the net proportion of patients with culture-positive tuberculosis reclassified into a higher risk category, plus the net proportion of patients without culture-positive tuberculosis reclassified into a lower risk category (NRI = $[P(\text{up}|D=1) - P(\text{down}|D=1)] - [P(\text{up}|D=0) - P(\text{down}|D=0)]$). Final estimates of NRI and AUC were obtained using 10-fold cross-validation. Bootstrap confidence intervals for the NRI estimate are reported based on 1000 re-sampling iterations.

All p-values were two-sided with $\alpha=0.05$ as the significance level. All analyses were performed using Stata 10 (Stata Corporation, College Station, Texas) and R, version 2.8.1 (R Project for Statistical Computing).

Results

0f 1000 active tuberculosis suspects who had a QFT-G performed as part of their evaluation, 660 were included in the analysis (**Figure 1**). Of the 660 suspects, 630 (95%) had sputa and 30 (5%) had other tissue sent for AFB smear and culture as part of their diagnostic evaluation. Sixty-five (10%) patients were ultimately diagnosed with culture confirmed tuberculosis, of whom 14 (22%) had extrapulmonary tuberculosis. Median IFN-γ level was similar in patients with pulmonary and extrapulmonary disease (1.1 IU/ml vs. 1.0 IU/ml, p=0.89). The study population was predominantly male and foreign-born. Cases were more likely than non-cases to have weight loss, night sweats, and chest radiographs with evidence of active disease on presentation, and less likely to have a history of prior active tuberculosis (**Table 1**). Median IFN-γ level was significantly higher in patients with tuberculosis compared to those without (1.1 IU/ml vs. 0.37 IU/ml, p<0.001), and higher IFN-γ levels were associated with increased odds of active tuberculosis (OR 1.07 (95% CI 1.03-1.11) for each doubling of IFN-γ level). For example, a patient with a quantitative IFN-γ result of 10 IU/ml had a 41% (95% CI 16-66%) increased odds of active tuberculosis relative to a patient with test results at the manufacturer-recommended cut-point of 0.35 IU/ml. 85% of cases had quantitative IFN-γ results in the upper three quintiles of IFN-γ concentration (≥ 0.23 IU/ml), while only 6% of cases were in the lowest quintile (< 0.04 IU/ml) (**Table 1**). Sensitivity and specificity of QFT-G for active tuberculosis at the manufacturer-recommended cut-point were 72% and 47%, and positive and negative predictive values were 13% and 89%, respectively.

A tuberculin skin test was performed in 117 (18%) patients prior to QFT-G measurement. There was no difference in the proportion of patients with culture-confirmed tuberculosis among those who did and did not have a tuberculin skin test performed prior to QFT-G (p=0.41).

Clinical Prediction Model

The baseline prediction model including objective demographic and clinical predictors classified 182 (28%) patients into low risk, 407 (62%) into intermediate risk, and 71 (11%) into high risk categories. The presence of new infiltrate, pleural effusion, or lymphadenopathy on chest radiograph was most predictive of active tuberculosis (**Table 2**).

Quantitative IFN-γ Results and Risk Reclassification

The addition of quantitative IFN-γ results to the baseline prediction model including demographic and clinical predictors significantly improved model accuracy (AUC 0.71 (0.64-0.77) vs. 0.78 (0.73-0.84), p<0.001) (**Table 2**) and 32% (95% CI 11-52%, p<0.001) of tuberculosis suspects were correctly reclassified into higher or lower risk categories (**Table 3a**). In comparison to the clinical model alone, both case reclassification (14 more cases classified as high risk and 5 fewer as low risk) and non-case reclassification (90 more non-cases designated as low-risk and only 7 more classified as high-risk) were improved. Results were similar when alternate thresholds were used to define risk categories (**Supplementary Table 1**). Findings on chest radiograph remained the strongest predictor of active tuberculosis.

Secondary Analysis

We performed a secondary analysis to determine whether quantitative IFN-γ levels improved risk reclassification beyond a prediction model including clinician suspicion. First, we evaluated whether a similar benefit in risk reclassification occurred when clinician suspicion, rather than quantitative IFN-γ level, was added to the baseline prediction model including objective demographic and clinical data. When clinician suspicion was added to the baseline model, accuracy increased (AUC 0.71 (95% CI 0.64-0.77) vs. 0.82 (95% CI 0.77-0.88), $p<0.001$) and 45% of tuberculosis suspects (95% CI 23-80%, $p<0.001$) were appropriately reclassified into higher or lower risk categories (data not shown). Next, addition of quantitative IFN-γ results to this expanded model including clinician suspicion significantly increased accuracy (AUC 0.82 (0.77-0.88) vs. 0.86 (0.81-0.91), $p=0.02$), but not net reclassification index (NRI 4%, 95% CI -0.07-0.22, $p=0.14$). Improved prediction among tuberculosis cases was outweighed by worse performance among non-cases (**Table 3b**). The addition of QFT-G results at the manufacturer-recommended cut-point of 0.35 IU/mL in place of quantitative IFN-γ levels did not materially affect results obtained in either the primary or secondary analysis. To further explore performance in cases and non-cases, we examined individual patients' risk before and after quantitative IFN- γ level was added to the model. The majority of culture-proven cases showed an appropriate increase in predicted risk with addition of quantitative IFN- γ results (**Figure 2a**). However, both decreased (appropriate) and increased (inappropriate) risk prediction was common among non-cases (**Figure 2b**).

Discussion

In this study, we found that quantitative IFN-γ results significantly improved risk stratification of smear-negative pulmonary and extrapulmonary tuberculosis suspects when added to objective clinical and demographic risk factors. However, this benefit in prediction became attenuated when clinician suspicion was taken into account. These findings indicate that IFN-γ levels obtained from QFT-G, at either the manufacturer-recommended cut-point or as a quantitative measure, are unlikely to influence clinical management of active tuberculosis suspects attending centers with highly experienced clinicians in low incidence settings.

Risk prediction has long been used in the cardiovascular (26, 27) and cancer (28) fields to improve precision of diagnoses and inform decisions about treatment. Published literature to date assessing IGRA performance has been limited to considerations of sensitivity, specificity, and predictive value, though these measures alone do not describe the predictive accuracy of these assays or the extent to which they improve upon readily available clinical information (29). In the absence of an established risk prediction model for AFB smear-negative tuberculosis, we utilized the Deletion/Substitution/Addition (DSA) routine (24) to identify the optimal prediction model. This state-of-the-art procedure considers non-linear terms and all possible interactions between predictors. Simultaneously, DSA avoids model overfitting through repeated cross-validation. The models generated in this study demonstrate moderate to good discrimination, similar to the Framingham Risk Score for prediction of mortality from coronary heart disease (27, 30).

Previous studies examining quantitative QFT-G results have shown improved sensitivity when using cut-points lower than that suggested by the manufacturer (14, 31, 32). However, cut-points selected from AUC analysis are influenced by disease prevalence in the population being

studied, give equal weight to false positive and false negative test results, and may misclassify individuals whose test result falls near the selected cut-points (33). Our analyses incorporated IFN- γ levels as a continuous measure, reported diagnostic benefit in light of conventional risk factors, and used novel reclassification methods that allow QFT-G results to be considered in the context of standard clinical decision-making. Our overall conclusions weight the net reclassification results more heavily than improvements in discrimination represented by increases in AUC. While broadly used as a summary measure of test performance, the area under the receiver-operator characteristic curve (AUC) does not focus on actual risk probabilities and their relation to clinical decision-making, and is thus limited in its clinical relevance and utility for evaluating risk prediction models (29, 34).

The changes in predicted risk of active tuberculosis following consideration of quantitative IFN-γ results were not uniform. Among intermediate and high risk patients in whom active TB was eventually ruled out, the addition of quantitative interferon-γ results led to clinically significant decreases in risk probabilities whether or not clinical suspicion was also included in the prediction model. These findings support previous work emphasizing a high negative predictive value for QFT-G (11). However, approximately one-quarter of low risk suspects in whom tuberculosis was eventually ruled out were inappropriately reclassified as intermediate risk after consideration of quantitative IFN-γ results. The possibility that quantitative IFN-γ results have increased clinical utility in intermediate and high risk tuberculosis suspects warrants further study.

Our study has several limitations. First, net reclassification index results depend heavily upon both the base prediction model and choice of risk categories. We recognize that addition of IFN-γ to suboptimal base models could produce large improvements in both discrimination and risk reclassification. We utilized novel methods to optimize our prediction models and their performance compares well with other well-accepted risk prediction models (27, 30). In addition, our risk cut-points were pre-specified, and sensitivity analyses of alternate cut-points did not modify our findings. Second, clinician suspicion, as used in our expanded clinical model, could have been influenced in some cases by QFT-G results. This is unlikely to have materially affected our analysis, as 85% of all QFT–G results were not available at the time of clinical evaluation, and quantitative interferon-gamma results are not reported by the SFDPH laboratory. The dramatic improvement in model performance with addition of clinician suspicion, however, indicates that crucial information is obtained in the work up process beyond our measured covariates. Future prospective studies should attempt to better define these factors. Third, the test characteristics of QuantiFERON-TB Gold In-Tube (QFT-G-IT), the most recent generation of this assay, may differ from QFT-G as used in this study. Last, our analysis is most relevant to tuberculosis referral centers with experienced clinicians operating in low incidence settings.

In conclusion, quantitative IFN-γ results obtained from QFT-G improved clinical evaluation of tuberculosis suspects compared to objective criteria. But in our highly experienced tuberculosis control clinic, subjective assessment of risk by clinicians performed even better. Further studies are needed to examine whether quantitative IGRA results have benefit beyond routine clinician evaluation in other settings.

ACKNOWLEDGEMENTS The authors would like to thank the staff at San Francisco Department of Public Health, Tuberculosis Control Section.

Table 1. Description of AFB Smear Negative Active Tuberculosis Suspects

Characteristic	Total	Non-Cases (n=595)	Culture-Positive Cases (n=65)	p-value
Age (yrs); median (IQR)	54 (43-64)	54 (44-65)	50 (36-62)	0.06
Male, %	61	61	69	0.17
Race/Ethnicity, %				
White	6	6	9	0.38
African American	6	6	8	
Asian	79	80	71	
Hispanic	9	8	12	
US Born, %	10	9	18	0.02
Positive QFT-G Result,* %	55	52	72	<0.01
IFN-γ Concentration, IU/ml (IQR)	0.46 (0.06-2.42)	0.37 (0.05-2.38)	1.1 (0.36-3.77)	<0.01
Quintiles of IFN-γ Concentration (IU/ml), %				
<0.04		22	6	<0.001
0.04 - 0.23		20	9	
0.23 – 0.9		19	34	
0.9 – 3.12		20	22	
3.12 - 10		19	29	
Active Disease on CXR, %	19	16	43	<0.001
Clinical Symptoms, %				
Night sweats or weight loss	17	15	32	<0.001
Cough	29	28	35	0.20
Hemoptysis	5	5	8	0.33
Previous Active TB, %	15	16	5	0.01
Contact with Active TB Case, %	2	2	6	0.03
Predisposing Medical Condition,† %	13	13	11	0.60
Homelessness, %	8	8	12	0.18
BCG Vaccination, %	20	20	18	0.79
Diabetes Mellitus, %	9	10	5	0.19
Clinician Suspicion for Active TB at Initial Evaluation, %				
Low	73	79	26	<0.001
Intermediate	13	12	20	
High	14	9	54	

Values are expressed as percentages unless otherwise stated.
* Positive at the manufacturer-recommended cut-point of 0.35 IU/ml.
† Predisposing medical condition: diabetes mellitus, silicosis, cancer, or condition requiring use of immunosuppressive medications.
Definition of abbreviations: IFN-γ = Interferon-gamma; IU/ml = International units per milliliter; IQR = Interquartile range; US = United States; QFT-G = QuantiFERON®-TB Gold; BCG = Bacillus Calmette-Guérin vaccine; TB = tuberculosis; CXR = chest radiograph.

Table 2. Coefficients and Summary Statistics for Prediction Models

	Baseline Clinical Prediction Model	Baseline Prediction Model with IFN-γ Results	Baseline Prediction Model with Clinician Suspicion	Baseline Prediction Model with Clinician Suspicion and IFN-γ Results
CXR, Active Disease*	2.92	3.66	0.92	1.18
Night sweats or weight loss	1.60	2.22	1.12	1.45
Previous Active Disease	0.29	0.27	0.24	.0 23
US Birth†	1.80	2.85	2.01	2.95
Foreign Birth, ≤ 2 years in US †	1.41	1.58	2.33	2.45
Foreign Born, 3-12 years in US †	2.71	3.37	2.09	2.65
Contact to Active Case	2.43	2.11	3.69	3.09
High Clinical Suspicion‡			19.43	19.31
Intermediate Clinical Suspicion‡			5.53	4.83
Quantitative IFN-γ Result (effect size per each doubling, IU/ml)		1.08		1.08
AIC	400	374	346	323
AUC	0.71 (0.64-0.77)	0.79 § (0.73-0.84)	0.82 (0.77-0.88)	0.86 ll (0.81-0.91)

Note: all results displayed in Table 3 are cross-validated.
* Reference category: inactive disease or normal CXR
† Reference category: Foreign born, >12 years in U.S.
‡ Reference category: low clinical suspicion
§ Significant difference (p<.001) between this model and previous model without quantitative IFN-γ results.
ll Significant difference (p=.02) between this model and previous model without quantitative IFN-γ results.
Definition of abbreviations: IFN-γ = Interferon-gamma; IU/ml = International units per milliliter; US = United States; CXR = chest radiograph; AUC = Area under the receiver operating curve, the probability that a randomly selected case will have a higher test value than a randomly selected noncase; a perfect test has an area under the curve of 1.0, while a worthless test has an area of 0.5; AIC = Akaike information criterion, a measure of the goodness of fit of a statistical model with lower values indicating better fit.

Table 3: Risk Reclassification Following Incorporation of IFN-γ Results

A. Comparison to Baseline Clinical Prediction Model

Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN-γ Results				
In 65 patients who developed culture-positive disease:	**≤5% risk**	**5-20% risk**	**>20% risk**	**Total No.**	**Per cent Appropriately Reclassified**
≤5% risk	7	3	0	10	30%
5-20% risk	1	27	12	40	28%
>20% risk	1	0	14	15	7%
Total No.	9	30	26	65	
In 595 patients who ruled out for active tuberculosis:					
≤5% risk	158	18	0	176	-10%
5-20% risk	89	241	27	357	17%
>20% risk	17	10	35	62	44%
Total No.	264	269	62	595	

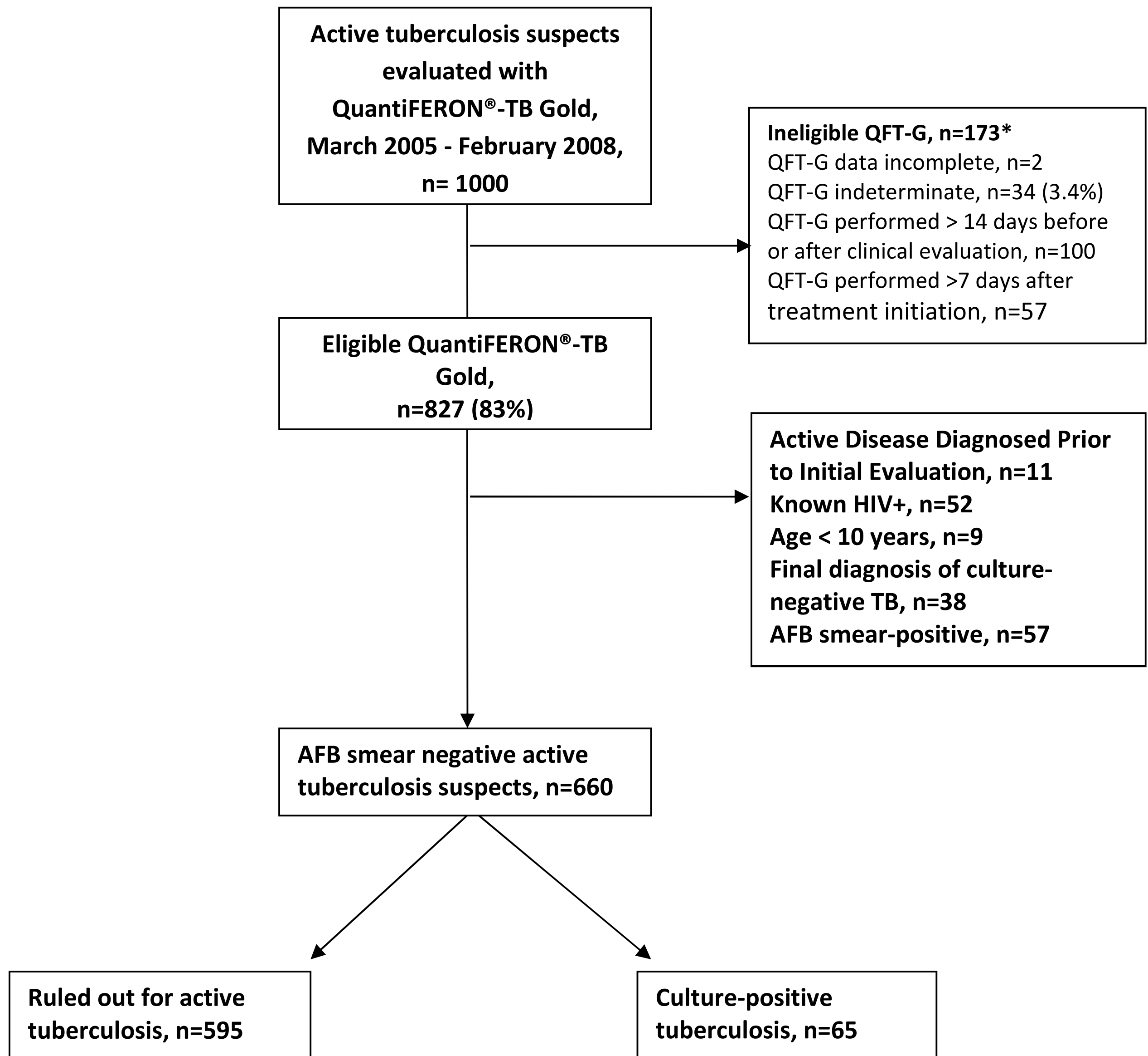
Net reclassification improvement = 31.9% ($p < .001$) [Reclassification among patients who developed culture-positive disease: 20% ($p < .01$), reclassification among patients who ruled out for active tuberculosis: 11.9% ($p < .001$)].

B. Comparison to Expanded Clinical Prediction Model

Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN-γ Results				
In 65 patients who developed culture-positive disease:	**≤5% risk**	**5-20% risk**	**>20% risk**	**Total No.**	**Per cent Appropriately Reclassified**
≤5% risk	7	9	0	16	56%
5-20% risk	3	6	7	16	25%
>20% risk	1	0	32	33	-3%
Total No.	11	15	39	65	
In 595 patients who ruled out for active tuberculosis:					
≤5% risk	334	121	0	455	-27%
5-20% risk	20	34	14	68	9%
>20% risk	9	18	45	72	38%
Total No.	363	173	59	595	

Net reclassification improvement = 3.7% ($p = .31$) [Reclassification among patients who developed culture-positive disease: 18.5% ($p < .01$), reclassification among patients who ruled out for active tuberculosis: -14.8% ($p = 1$)].

Figure 1. Study Flow Diagram



*** Some patients were excluded for more than one reason**

Figure 2a. Changes in Predicted Risk of Active Disease following Incorporation of Quantitative IFN-γ Results, Cases.

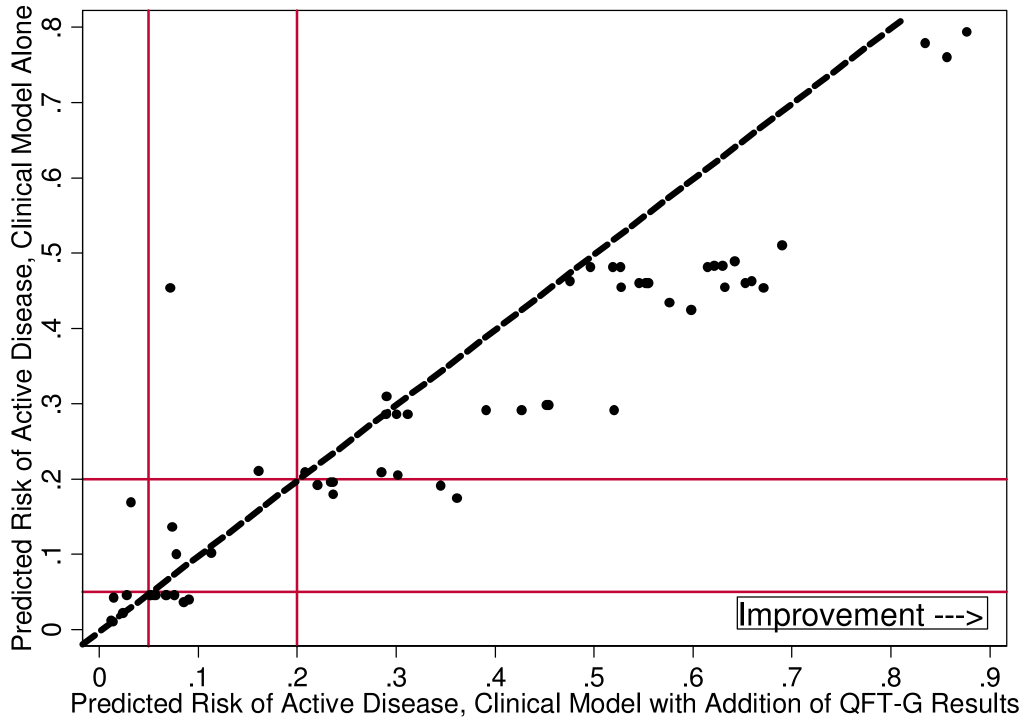
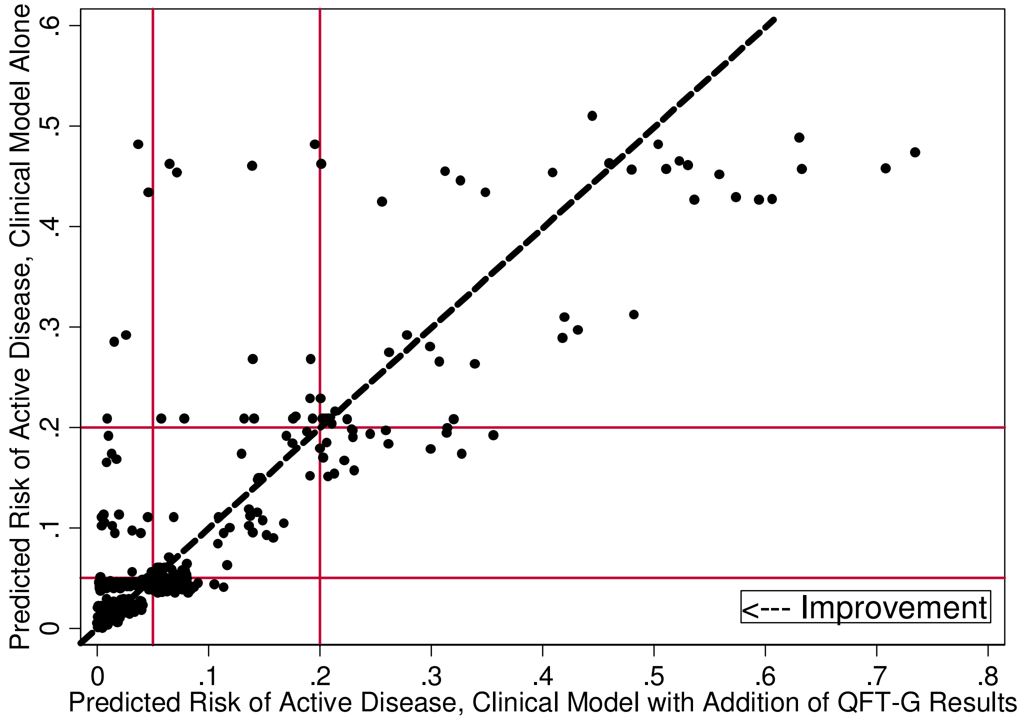


Figure 2b. Changes in Predicted Risk of Active Disease following Incorporation of Quantitative IFN-γ Results, Non-cases.



Dashed diagonal lines represent no change in risk prediction with addition of quantitative IFN-γ results. Among culture-proven cases, individuals to the right of the dashed diagonal line indicate higher (and therefore improved) risk prediction, and those to the left indicate lower (and therefore worse) risk prediction; these criteria are reversed for non-cases. Horizontal and vertical solid lines indicate lower (5%) and upper (20%) risk cut points at ordinate and abscissa, respectively.

Chapter 2. Interferon-gamma Release Assays for Diagnosis of Active Pulmonary TB Diagnosis in Adults in Low- and Middle-Income Countries: Systematic Review and Meta-Analysis

ABSTRACT

Background: The value of IGRAs in the diagnosis of active TB in low- and middle-income countries is unclear.
Methods: We searched multiple databases for studies published through May 2010 evaluating the diagnostic performance of QuantiFERON-TB Gold In-Tube (QFT-GIT) and T-SPOT.TB (T-SPOT) among adults with or suspected of having active pulmonary TB in low- and middle-income countries. We summarized test performance characteristics using forest plots, hierarchical summary ROC (HSROC) curves, and bivariate random effects models.

Results: Our search identified 789 citations, of which 27 observational studies (17 QFT-GIT, 10 T-SPOT) evaluating 590 HIV-uninfected and 844 HIV-infected individuals met inclusion criteria. Among HIV-infected patients, HSROC/bivariate pooled sensitivity estimates (highest quality data) were 76% (95% CI 45-92%) for T-SPOT and 60% (95% CI 34-82%) for QFT-GIT. HSROC/bivariate pooled specificity estimates were low for both IGRA platforms among all subjects (T-SPOT (61%, 95% CI 40-79%), QFT-GIT (52%, 95% CI 41-62%)), and among HIV-infected subjects (T-SPOT (52%, 95% CI 40-63%), QFT-GIT (50%, 95% CI 35-65%)). There was no consistent evidence that either IGRA was more sensitive than the TST for diagnosis of active TB.
Conclusions: In low- and middle-income countries, IGRAs are inadequate for ruling out or ruling in active TB, especially in the setting of HIV co-infection.

BACKGROUND

Interferon-gamma release assays (IGRAs) are the first new diagnostic test for latent tuberculosis infection (LTBI) in over 100 years. Newest generation IGRAs measure interferon-gamma (IFN-γ) secretion after exposure of whole blood (QuantiFERON-TB Gold In-Tube® [QFT-GIT], Cellestis, Carnegie, Australia) or peripheral blood mononuclear cells (T-SPOT.TB® [T-SPOT], Oxford Immunotec, Abingdon, UK) to antigens encoded within the region of difference-1 (RD1), a portion of the MTB genome absent among all BCG strains and most nontuberculous mycobacteria. [1] We have shown in previous systematic reviews that, compared to the tuberculin skin test (TST), IGRAs have higher specificity for LTBI in low tuberculosis (TB) incidence settings, better correlation with surrogate measures of *M. tuberculosis* exposure, and less cross reactivity with response to BCG vaccine. [2-4] In recent years, IGRAs have become widely endorsed in high income-countries for diagnosis of LTBI. [5-7]

IGRAs were explicitly designed to replace the tuberculin skin test (TST) in the diagnosis of LTBI, and were not intended to be used in the diagnosis of active TB, which is a diagnosis based on microbiological tests (e.g., culture and microscopic examination of clinical specimens). Furthermore, the diagnosis and treatment of LTBI remains a low priority in most low- and middle-income countries, where detection and management of active TB is the highest priority for national TB control programs. Because IGRAs, like the TST, cannot distinguish LTBI from active TB, [8-10] these tests can be expected to have poor specificity for diagnosis of active TB in all high burden settings, due to a high background prevalence of LTBI. [11] Other differences in patient characteristics, 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 the observed poorer performance of IGRAs in these settings. [12] Yet, private sector laboratories in high burden countries increasingly employ IGRAs to diagnose active TB, [13] and many investigators continue to recommend the use of IGRAs for this purpose. [14-17]

Because of uncertainty of the benefits and costs to patients and national TB programs, we conducted a systemic review and meta-analysis to determine IGRA test performance in patients suspected of having active pulmonary TB and in patients with confirmed TB living in low- and middle-income settings.

METHODS

Overview. Given the absence of studies evaluating patient-important outcomes among TB suspects randomized to treatment based on IGRA results, we focused our review on the accuracy of IGRAs in diagnosing active TB. We followed standard guidelines and methods for systematic reviews and meta-analyses of diagnostic tests.[18-21]

Search methods. We have previously published systematic and narrative reviews on the accuracy and performance of IGRAs in various subgroups.[2-4, 10, 22] We updated the previous literature searches to identify all studies evaluating IGRAs published through May 2010 searching PubMed, Embase, Biosis and Web of Science for studies in all languages. The search terms used included: ((interferon-gamma release assay*) OR (T-cell-based assay*) OR (antigen-specific T cell*) OR (T cell response*) OR (T-cell response*) OR (interferon*) OR (interferon-

gamma) OR (gamma-interferon) OR (IFN) OR (elispot) OR (ESAT-6) OR (CFP-10) OR (culture filtrate protein) OR (Enzyme Linked Immunosorbent Spot) OR (Quantiferon* OR Quantiferon-TB)) AND (tuberculosis OR mycobacterium tuberculosis). In addition to database searches, we reviewed bibliographies of reviews and guidelines, screened citations of all included studies, searched clinicaltrials.gov for ongoing studies, and contacted both experts in the field and IGRA manufacturers to identify additional published and unpublished studies. We requested pertinent information not reported in the original publication from the primary authors of all studies included in the review.

Study selection and data collection. We included studies that evaluated the performance of the most recent generation of commercial, RD1 antigen-based IGRAs (QuantiFERON-TB Gold In-Tube (QFT-GIT) (Cellestis, Victoria, Australia) and T-SPOT (Oxford Immunotec, Oxford, United Kingdom) among adults (aged ≥ 15 years) suspected of or having active pulmonary TB in low- and middle-income countries, [23] using the World Bank Country Classification as a surrogate for national TB incidence. HIV infection was established either by documented serological testing or self-report. We excluded: (1) studies that evaluated non-commercial (in-house) IGRAs, purified protein derivative (PPD)-based IGRAs, QuantiFERON-TB Gold (2G), and IGRAs performed in specimens other than blood; (2) longitudinal data focused on the effect of anti-TB treatment on IGRA response; (3) studies including < 10 eligible individuals; (4) studies focused on extrapulmonary tuberculosis or children (<15 years); (5) studies reporting insufficient data to determine diagnostic accuracy measures; and (6) conference abstracts, letters without original data, and reviews.

At least two reviewers (JZM, CE, KRS, AC) independently screened each of the accumulated citations for relevance, reviewed full-text articles using the pre-specified eligibility criteria, and extracted data using a standardized form. The reviewers resolved disagreements about study selection and data extraction by consensus.

Assessment of study quality. Because primary outcomes for this systematic review focus on test accuracy, we evaluated study quality using a subset of relevant criteria from QUADAS (a validated tool for diagnostic accuracy studies). [24] Because of growing concerns about conflicts of interest in diagnostic studies and guidelines, [25, 26] we also report whether IGRA manufacturers had any involvement with the design or conduct of each study, including donation of materials, monetary support, work/financial relationships with study authors, and participation in data analysis.

Outcome Definitions. Well-designed diagnostic accuracy studies focus on a representative target population in whom genuine diagnostic uncertainty exists (i.e., patients in whom clinicians would apply the test in the course of regular clinical practice). [27] There is evidence that diagnostic studies that include only known cases with the condition of interest and healthy controls without this condition tend to overestimate test accuracy. [28] Therefore, we considered studies simultaneously evaluating IGRA sensitivity and specificity among active TB suspects to represent the highest quality evidence, while studies evaluating IGRA performance among patients with known active TB (for sensitivity) to be of lesser quality. Because of our focus on the diagnostic accuracy for active TB and the high prevalence of LTBI in high TB burden settings, IGRA specificity was estimated exclusively among studies enrolling active TB suspects where the diagnostic workup ultimately showed no evidence of active disease.

A hierarchy of reference standards for active TB was developed a priori to judge the quality of each individual assessment of IGRA diagnostic accuracy. From most to least favorable, these reference standards included 1) culture-confirmation or sputum smear-positivity in high TB incidence settings (≥50/100,000), where sputum smear microscopy has been shown to have high specificity; [29] 2) sputum smear-positivity without culture in low or intermediate TB incidence settings (<50/100,000); and 3) clinical diagnosis based upon presenting symptoms, radiologic findings and/or response to TB treatment without microbiological confirmation. Because the TST remains in widespread use, and because indeterminate IGRA results may affect assay performance in low income settings, we also evaluated (1) observed differences in sensitivity for active TB diagnosis between IGRA and TST, and (2) the proportion of IGRA results among patients with active TB that are indeterminate.

We used the following definitions for primary outcomes: (1) Sensitivity – the proportion of individuals with a positive IGRA result among those with culture-positive TB (we included indeterminate IGRA results in the denominator if they occurred in individuals with culture-positive TB); and (2) Specificity – the proportion of individuals with a negative IGRA result among those ruled out for active TB disease (indeterminate IGRA results were excluded from analysis). Using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework,[27] these measures can be interpreted as surrogates for patient-important outcomes.

Data synthesis and meta-analysis. Multiple sources of heterogeneity commonly exist when summarizing estimates from studies of diagnostic tests. [30] We adopted the following approach to account for expected heterogeneity. First, when possible, we separately synthesized data for each commercial IGRA and by HIV status. The pre-specified sub-groups minimize heterogeneity related to differences in testing platform (ELISA vs. ELISPOT), antigens used to elicit IFN-g release (ESAT-6/CFP-10 vs. ESAT-6/CFP-10/TB 7.7), and test performance related to HIV-associated host immunosuppression. Second, we visually assessed heterogeneity using forest plots, characterized the variation in study results attributable to heterogeneity (I-squared value), and statistically tested for heterogeneity (chi-squared test). [30] Third, we calculated pooled sensitivity and specificity estimates using random effects modeling, which provide more conservative estimates than fixed effects modeling when heterogeneity is a concern. [19, 31]

For each individual study, we assessed all outcomes for which data were available. First, we generated forest plots to display the individual study estimates and their 95% confidence intervals. Second, we used bivariate random effects regression models [32] when both sensitivity and specificity could be reported from the same TB suspect population. Because pooling sensitivity and specificity separately can produce biased estimates of test accuracy [19], we preferred to generate pooled estimates when both sensitivity and specificity were reported within a study and ranked this as higher quality evidence. Third, we generated hierarchical summary receiver operating characteristic (HSROC) curves to summarize the global test performance.[31] Because of the need to summarize two correlated measures (e.g., sensitivity and specificity), and because substantial between-study heterogeneity is common, meta-analysis of diagnostic accuracy requires different and more complex methods than traditional meta-analytic techniques. Graphically illustrating the trade-off between sensitivity and specificity, HSROC curves differ from traditional Receiver Operating Characteristics (ROC) curves in

allowing accuracy to vary by each individual study (i.e., allowing for random effects, and thus allowing asymmetry in the plotted curve), and by discouraging extrapolation beyond the available data by plotting the curve only over the observed range of test characteristics. The HSROC approach is closely related to the bivariate random effects regression model [33]. These two methods generally produce similar results and are both recommended by the Cochrane Diagnostic Test Accuracy Working Methods group. [20] We calculated pooled estimates when at least four studies were available in any sub-group and summarized individual study results when fewer than four studies were available. We performed all analyses using Stata 11 (Stata Corporation, College Station, Texas, USA). For bivariate random effects regression and HSROC analyses, we used the user-written "metandi" program for Stata. [32]

RESULTS

Search results. The initial search yielded 789 citations (Figure 1). After full-text review of 168 papers, 19 papers [15, 17, 34-50] were determined to meet eligibility criteria for IGRA evaluation of active TB in low- and middle-income settings. Because some papers included more than one commercial IGRA, there were 27 unique evaluations (referred to as studies) – 17 of QFT-GIT, and 10 of T-SPOT – that included a total of 590 HIV-uninfected and 844 HIV-infected individuals.

Study characteristics. Of the total studies, seven (26%) were from low-income countries, and 20 (74%) were from middle-income countries. Fourteen studies (52%) included HIV-infected individuals, and 21 (78%) studies involved ambulatory subjects (i.e., outpatients as well as hospitalized patients) (Table 1). IGRAs were performed in persons suspected of having active TB in 14 (52%) studies, [35, 37-39, 41, 42, 47, 48, 50] and in persons with known active TB in 13 (48%) studies. [15, 17, 34, 36, 40, 44-46, 49, 51] A list of excluded studies and reasons for exclusion is available from the authors upon request.

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

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

Sensitivity. With the exception of two studies, [37, 48] the sensitivity of IGRAs was assessed based on a positive culture result (21 studies, 78%) or a positive sputum AFB-smear result in a high TB incidence setting (4 studies, 15%). Among studies performed in patients with known active TB, 6 (46%) included patients who had been treated for greater than one week.

HIV-infected

Nine studies assessed IGRA sensitivity among HIV-infected active TB suspects. HSROC/bivariate pooled sensitivity estimates were higher for T-SPOT (76%, 95% CI 45-92%, 4 studies [35, 38, 41, 42]) than for QFT-GIT (60%, 95% CI 34-82%, 5 studies [38, 39, 41, 42, 50] (Figure 3). Pooled sensitivity estimates did not change appreciably for either T-SPOT (68%, 95% CI 56-80%, 5 studies [15, 35, 41-43]) or QFT-GIT (65%, 95% CI 52-77%, 7 studies [34, 39, 41, 42, 49, 50]) when studies evaluating patients with known active TB were included in the analysis (Figure 4). Pooled sensitivity estimates for both T-SPOT (I-squared 72%, p<0.01) and QFT-GIT (I-squared 76%, p<0.001) showed significant heterogeneity.

HIV-uninfected

Five studies assessed IGRA sensitivity among HIV-uninfected active TB suspects; data were insufficient to report HSROC/bivariate pooled sensitivity estimates for either QFT-GIT [37, 38, 48] or T-SPOT [38, 47]. Pooled sensitivity estimates were similar for T-SPOT (88%, 95% CI 81-95%, 4 studies [17, 38, 44, 47]) and QFT-GIT (84%, 95% CI 78-91%, 9 studies [10, 34, 36-38, 40, 46, 48, 49]) when studies evaluating patients with known active TB were included in the analysis (Figure 5). Pooled sensitivity estimates showed significant heterogeneity for QFT-GIT (I-squared 60%, p=0.01), but not for T-SPOT (I-squared 28%, p=0.25).

Head-to-head comparisons of QFT and T-SPOT sensitivity

Overall, four studies (three involving HIV-infected subjects [38, 41, 42] and one involving HIV-uninfected subjects [38]) reported head-to-head comparisons of T-SPOT and QFT-GIT sensitivity. T-SPOT sensitivity was higher but not significantly different from QFT-GIT sensitivity (sensitivity difference 19%, 95% CI -17% to 56%, p=0.3) (Table 2). Results were similar when restricted to HIV-infected individuals.

Head-to-head comparison of TST and IGRA sensitivity

Overall, nine studies reported head-to-head comparisons of TST and IGRA (3 T-SPOT and 6 QFT-GIT) sensitivity. TST sensitivity in the five studies [17, 40, 44, 46, 49] involving HIV-uninfected patients was higher (78%, 95% CI 71-86%) than in the four studies [15, 39, 46, 49] involving HIV-infected patients (45%, 95% CI 15-75%). IGRA sensitivity was not statistically different than TST sensitivity for either T-SPOT (sensitivity difference 23%, 95% CI 0% to 45%, p=0.05) or QFT-GIT (sensitivity difference 7%, 95% CI -9% to 23%, p=0.37) (Figure 6). There was significant heterogeneity for both estimates (I-squared >75%, p<0.001). Data were insufficient to form HIV-stratified pooled sensitivity difference estimates for either IGRA.

Specificity. All specificity estimates were determined in TB suspects using HSROC/bivariate techniques. Overall, pooled specificity was low for both T-SPOT (61%, 95% CI 40-79%, 6 studies) and QFT-GIT (52%, 95% CI 41-62%, 8 studies). When restricted to HIV-infected active TB suspects, pooled specificity for T-SPOT (52%, 95% CI 40-63%, 4 studies [35, 38, 41, 42]) was similar to QFT-GIT (50%, 95% CI 35-65%, 5 studies [38, 39, 41, 42, 50]) (Figure 3). Too few studies were available to estimate pooled specificity for HIV-uninfected patients.

Proportion of Indeterminate IGRA Results. The proportion of indeterminate IGRA results among patients with suspected or confirmed active TB varied considerably (range 0-26% among studies enrolling 50 or more subjects). The proportion of indeterminate results was low (4%, 95% CI 1-7%) among HIV-uninfected patients, regardless of IGRA platform (Supplementary Figure 1). However, the proportion of indeterminate results was considerably higher among HIV-infected subjects for both QFT-GIT (15%, 95% CI 9-21%, 8 studies) and T-SPOT (9%, 95% CI 0-17%, 6 studies) (Supplementary Figure 2). Results were similar for HIV-infected subjects when stratified by TB suspects versus known TB cases.

DISCUSSION

The vast majority of the estimated annual 9.3 million new cases of active TB and 1.3 million TB-related deaths occur in low- and middle-income countries. [52] Due to resource constraints, public health policies have appropriately placed limited emphasis on diagnosis and treatment of LTBI in these settings. Clinical use of IGRAs, however, has expanded dramatically in recent years, especially in the private sector. [13] Because of their high burden of disease and emerging economies, such countries (e.g., India, South Africa, Brazil and China) represent a potentially lucrative market for manufacturers of commercial IGRAs. Even though IGRAs are intended to be used to diagnose LTBI and not to diagnose active TB, and even though they cannot distinguish between latent infection and active disease, there is increasing use of IGRAs to diagnose active TB in high burden countries. In this systematic review focused on individuals living in low- and middle-income countries, the highest quality evidence concerning the accuracy of IGRAs among TB suspects demonstrated sensitivity ranging from 69-83% and specificity ranging from 52-61%. Further, there was no consistent evidence that either IGRA was more sensitive than the TST for active TB diagnosis.

The majority of evidence on the diagnostic accuracy of IGRAs to date has come from high income settings, where active TB has been used as a surrogate reference standard when estimating the accuracy for LTBI diagnosis. [4, 14] Yet, diagnostic test performance (e.g., sensitivity and specificity) can be expected to vary according to disease prevalence and other population characteristics. [53, 54] Likewise, clinicians have been advised to base their decision-making on studies that most closely match their own clinical circumstances. [55]

IGRAs were designed to be used as diagnostic tests of LTBI, a setting in which the lack of an accepted gold standard has been a significant limitation in establishing test performance. By contrast, adequate and commonly used reference standards exist for diagnosing active TB. Among studies that enrolled active TB suspects (i.e., patients with diagnostic uncertainty), both IGRAs demonstrated suboptimal ‘rule-out’ value for active TB. In other words, one in four patients, on average, with culture-confirmed active TB can be expected to be IGRA-negative in low and middle income countries - this has consequences for patients in terms of morbidity and mortality if treatment decision sare made based on such results. Although high quality data were limited, sensitivity of both IGRAs was lower among HIV-infected patients (~60-70%), suggesting that nearly one in three HIV-infected patients with active TB will be IGRA-negative. 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. Lastly, comparisons with pooled estimates of TST sensitivity were difficult to

interpret due to substantial heterogeneity. Our results, however, suggest that neither IGRA platform may be more sensitive than the TST in diagnosis of active TB in low- and middle-income countries.

The specificity of IGRAs in diagnosing LTBI, estimated among individuals at low risk for TB exposure in low TB incidence (high income) settings, is known to be high (≥ 98%). [4] In contrast, specificity for active TB diagnosis is best estimated only within studies evaluating TB suspects. As expected, due to the higher background prevalence of LTBI and the known inability of IGRAs to differentiate LTBI from active TB, [10] the specificity of both IGRAs for active TB was low, regardless of HIV infection status. These data suggest that one in two patients without active TB will be IGRA-positive - this has consequences for patients because of unnecessary therapy for TB and its attendant risks. Our findings are intuitive and biologically plausible, because it is well known that T-cell IFN-γ responses are activated in nearly the entire spectrum of TB infection, from latency to active disease, [56] and currently available tests cannot distinguish LTBI from active TB disease. Even within the spectrum of latent TB infection itself, [57] actuated T-cell IFN- γ responses occur throughout each phase, with the possible exception of the innate immune response (which eliminates *M. tuberculosis* without priming a T-cell immune response).

The goal of our systematic review was to critically evaluate the diagnostic accuracy of IGRAs in the diagnosis of active TB in low and middle income settings. Yet, there are inherent limitations to sensitivity, specificity, and predictive values as measures of test performance. These measures are unable to determine either the extent to which a test may improve on readily available clinical information [58] or the degree to which patient-important outcomes are improved by test results. [27] Although limited, available data suggest that IGRAs may add little information to the conventional diagnostic work-up for active TB in low [59] and high TB incidence settings [60]. Further work is necessary to confirm this finding.

Our meta-analysis had several limitations. First, as with previous systematic reviews, [4, 14] heterogeneity was substantial for the primary outcomes of sensitivity and specificity. We utilized empirical random effects weighting, excluded all studies contributing fewer than ten eligible individuals, and separately synthesized data for currently manufactured IGRAs in order to minimize heterogeneity. Second, World Bank income classification is an imperfect surrogate for national TB incidence. Although no standard criteria currently exist for defining high TB incidence countries, our results were fundamentally unchanged when restricted to nations with an arbitrarily chosen annual TB incidence of greater than or equal to 50/100,000. [52] Third, it is likely that unpublished data and ongoing studies were missed. 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 our estimates due to publication bias. Fourth, our review did not include evidence on utility of IGRAs in two patient subgroups where conventional tests for active TB perform poorly - children and patients with suspected extra-pulmonary TB. Lastly, we did not identify any studies directly measuring the impact of IGRA test results on patient-important outcomes.

In conclusion, as in the case of the TST, the data suggest no role for using IGRAs in the diagnosis of active TB among adults living in low- and middle-income countries. These data should help inform evidence-based policies on the role of IGRAs in active TB diagnosis in low- and middle-income settings. Indeed, a World Health Organization (WHO) Expert Group considering this evidence recently recommended that IGRAs not be used as a replacement for conventional microbiological diagnosis of pulmonary and extra-pulmonary TB in low-and middle-income countries.[61]

TABLES AND FIGURES

Table 1. Characteristics of Included Studies.

Study, Year	Country	Income	Setting		Total Patients, *n*	Active TB *n* (%)	Indeterminate[†] *n* (%)	Industry Involvement[‡]
QFT-GIT								
Aabye 2009	Tanzania	Low	Inpatient/Outpatient	HIV-	93	93 (100)	8 (9)	Work relationship
Aabye 2009	Tanzania	Low	Inpatient/Outpatient	HIV+	68	68 (100)	15 (22)	Work relationship
Raby 2008	Zambia	Low	Outpatient	HIV-	37	37 (100)	5 (14)	No
Raby 2008	Zambia	Low	Outpatient	HIV+	59	59 (100)	10 (17)	No
Chegou 2009	South Africa	Upper Middle	Outpatient	HIV-	23	23 (100)	0 (0)	No
Chen 2009	China	Lower Middle	NR	HIV-	49	41 (84)	2 (4)	Unclear
Dheda (b) 2009	South Africa	Upper Middle	Inpatient/Outpatient	HIV+	20	5 (25)	8 (40)	No
Dheda (d) 2009	South Africa	Upper Middle	Inpatient/Outpatient	HIV-	51	15 (29)	14 (27)	No
Kabeer 2009	India	Lower Middle	Inpatient/Outpatient	HIV+	64	44 (69)	12 (19)	No

Katiyar 2008	India	Lower Middle	Outpatient	HIV-	76	76 (100)	0 (0)	Unclear
Leidl (b) 2009	Uganda	Low	Outpatient	HIV+	128	19 (15)	4 (3)	Kit Donation
Markova (b) 2009	Bulgaria	Upper Middle	Outpatient	HIV+	90	13 (14)	5 (6)	No
Pai 2007	India	Lower Middle	Inpatient/Outpatient	HIV-	57	57 (100)	0 (0)	Unclear
Tahereh 2010	Iran	Lower Middle	Unclear	HIV-	81	28 (35)	6 (7)	Unclear
Tsiouris 2006	South Africa	Upper Middle	Outpatient	HIV-	13	13 (100)	0 (0)	Kit Donation
Tsiouris 2006	South Africa	Upper Middle	Outpatient	HIV+	26	26 (100)	5 (19)	Kit Donation
Veldsman 2009	South Africa	Upper Middle	Outpatient	HIV+	60	30 (50)	9 (15)	No
T-SPOT.TB								
Cattamanchi 2010	Uganda	Low	Inpatient	HIV+	212	112 (53)	54 (25)	Kit Donation
Dheda (a) 2009	South Africa	Upper Middle	Inpatient/Outpatient	HIV+	20	5 (25)	1 (5)	No

Dheda (c) 2009	South Africa	Upper Middle	Inpatient/Outpatient	HIV-	49	15 (31)	2 (4)	No
Jiang 2009	China	Lower Middle	Inpatient/Outpatient	HIV+	32	32 (100)	0 (0)	No
Leidl (a) 2009	Uganda	Low	Outpatient	HIV+	128	19 (15)	6 (5)	Kit Donation
Markova (a) 2009	Bulgaria	Upper Middle	Outpatient	HIV+	90	13 (14)	9 (10)	No
Oni 2010	South Africa	Upper Middle	Outpatient	HIV+	85	85 (100)	5 (6)	Kit Donation
Ozekinci 2007	Turkey	Upper Middle	Inpatient	HIV-	28	28 (100)	0 (0)	No
Soysal 2008	Turkey	Upper Middle	Inpatient	HIV-	102	99 (97)	4 (4)	No
Shao-ping 2009	China	Lower Middle	Inpatient	HIV-	82	22 (27)	6 (7)	No

Definition of abbreviations: IQR, interquartile range; LTBI, latent tuberculosis infection; TB, tuberculosis; NR, not reported.

* Unpublished studies.

† Indeterminate results were not excluded in calculating sensitivity estimates.

‡ Kit donation refers to donation of any test materials including kits and reagents. Work relationship refers to when one or more authors are involved in test development, consulting work, or other employment by an IGRA manufacturer.

Table 2. Head-to-Head Comparison of Sensitivity of T-SPOT.TB versus QuantiFERON-TB Gold In-Tube among Active Tuberculosis Suspects.

Author, Year	Country	HIV Status	Active TB, *n* (%)	Positive T-SPOT Result, *n* (%)	Positive QFT Result, *n* (%)	Sensitivity difference* (%)
Dheda, 2009[†]	South Africa	HIV-	15 (31), 15 (29)	14 (93)	11 (73)	20
Dheda, 2009	South Africa	HIV+	5 (25)	5 (100)	1 (20)	80
Leidl, 2009	Uganda	HIV+	19 (15)	17 (89)	14 (74)	15
Markova, 2009	Bulgaria	HIV+	13 (14)	8 (62)	12 (92)	-31

* Sensitivity difference (%) is T-SPOT sensitivity (%) – QFT-GIT sensitivity (%).

[†] Total numbers of active TB suspects evaluated by each IGRA differed within some studies; these are listed in the order *T-SPOT*, *QFT-GIT*.

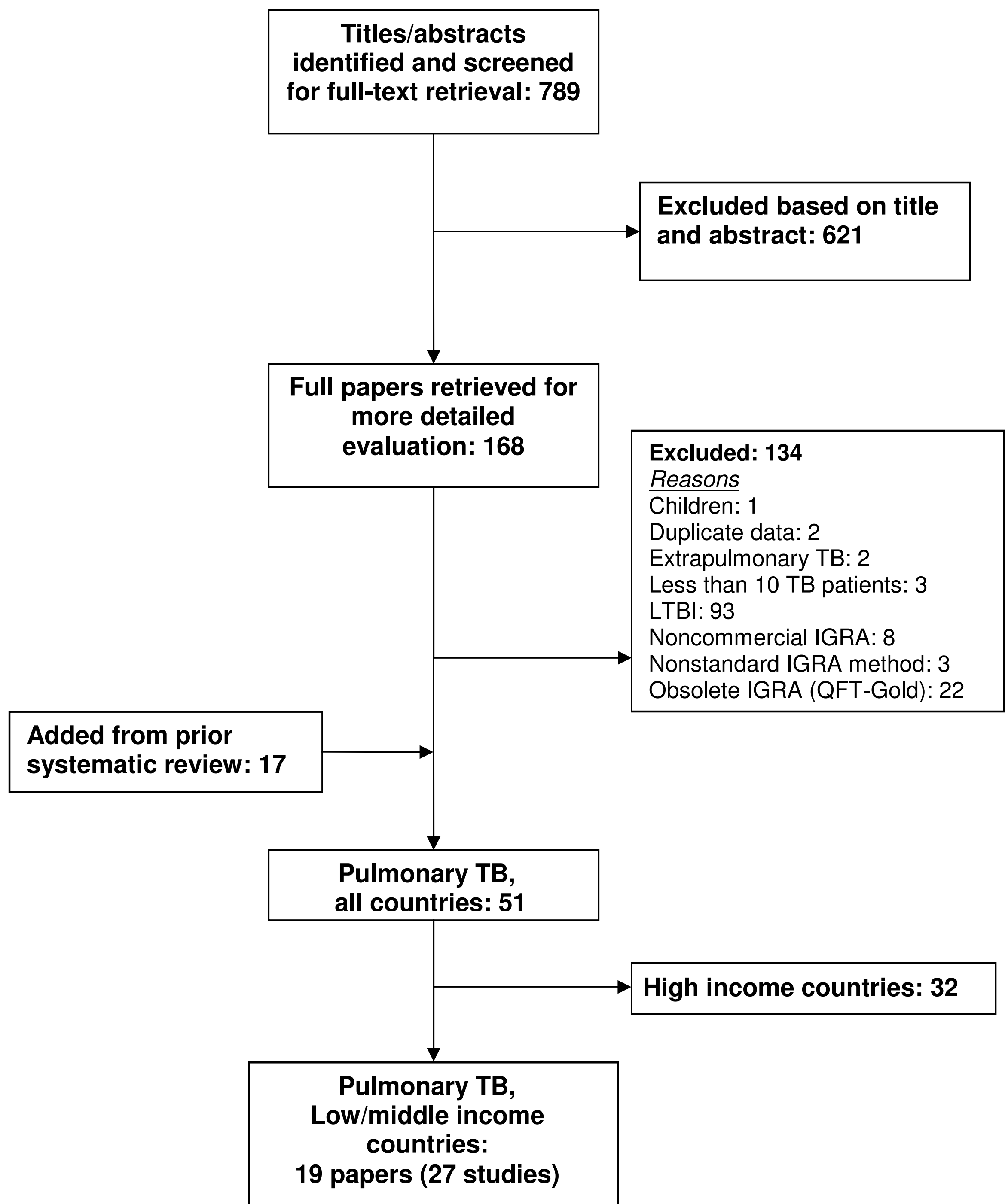


Figure 1. Study selection.

Abbreviations: IGRA, interferon-gamma release assay; LTBI, latent tuberculosis infection; TB, tuberculosis

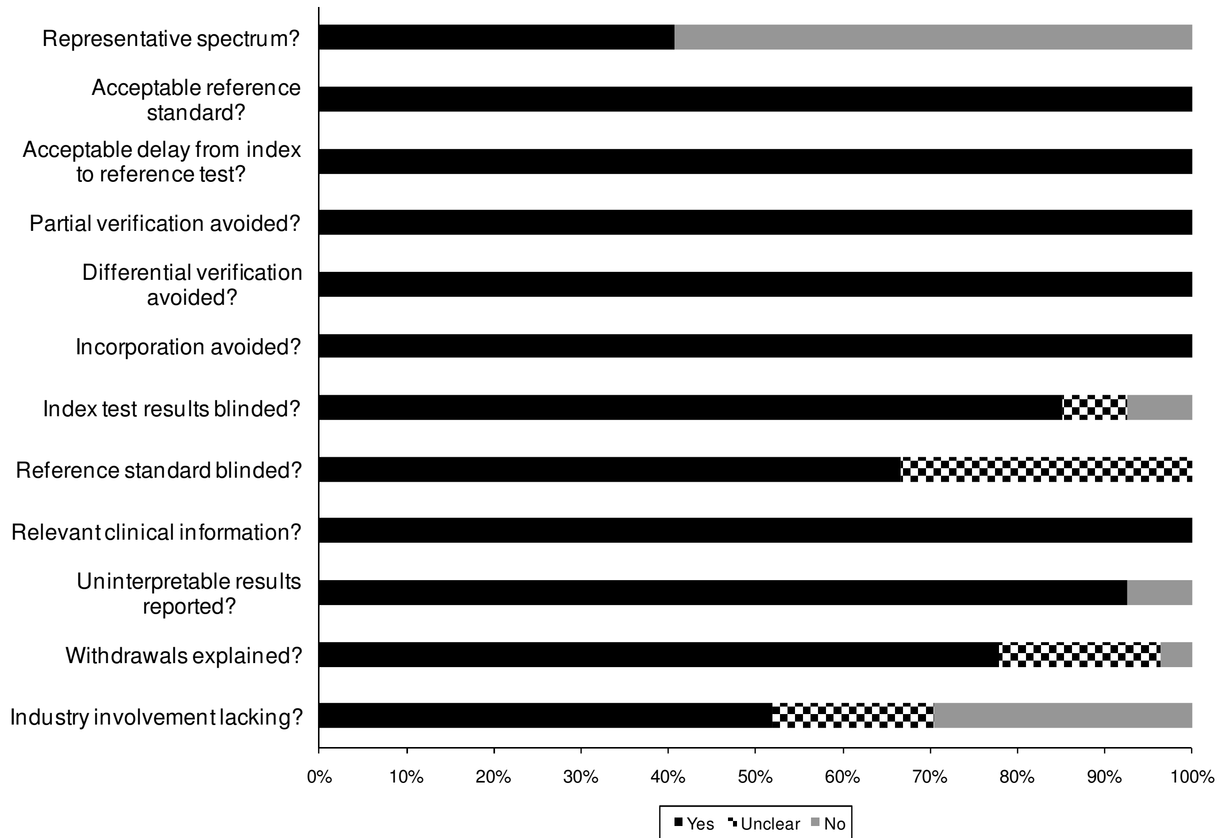


Figure 2. Assessment of study quality using the QUADAS tool.

For each QUADAS item, two reviewers independently determined whether a study did or did not meet the quality criterion, or whether it was unclear.

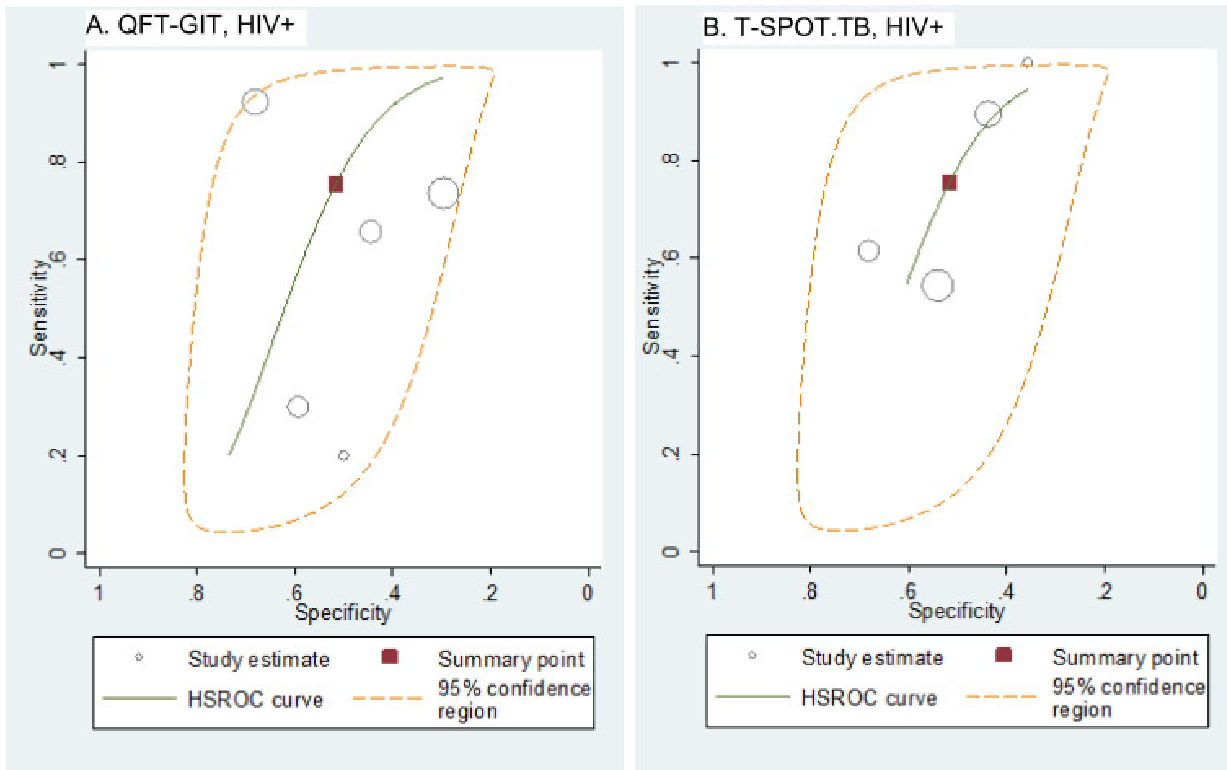


Figure 3a-b. Hierarchical Summary Receiver Operating Characteristics (HSROC) Plot of Studies that Reported both Sensitivity and Specificity in Active TB Suspects.

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

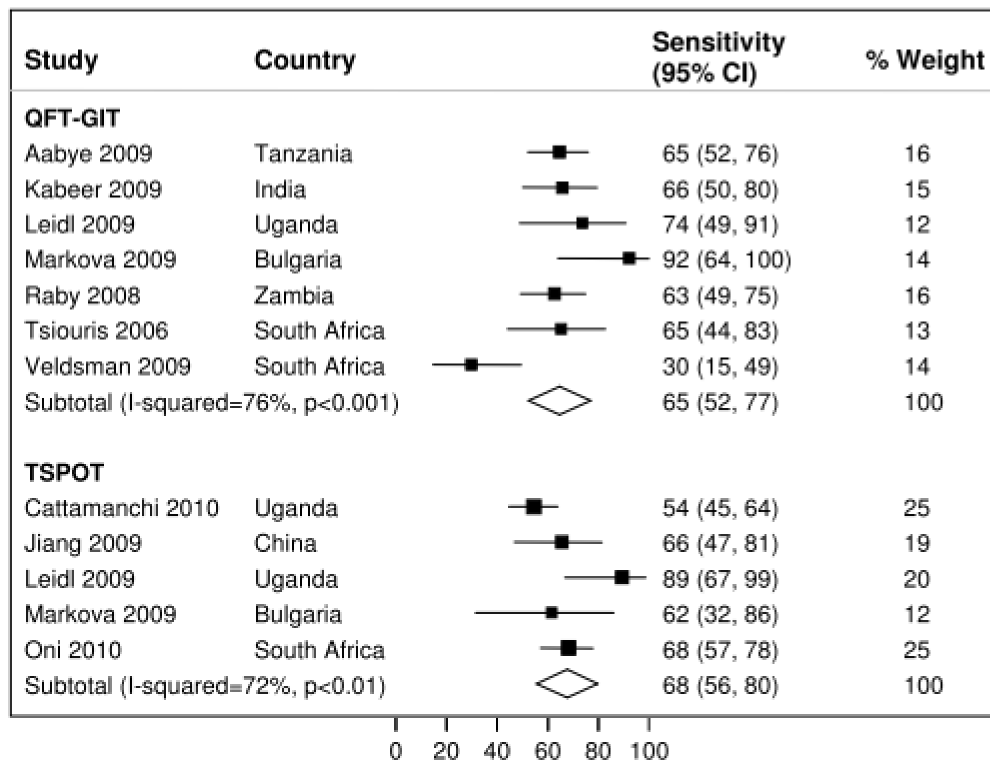


Figure 4. Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB in HIV-infected persons with confirmed active tuberculosis in low- and middle-income countries.
The forest plots display the sensitivity estimates obtained from individual studies and pooled estimates derived from random effects (DerSimonian-Laird)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

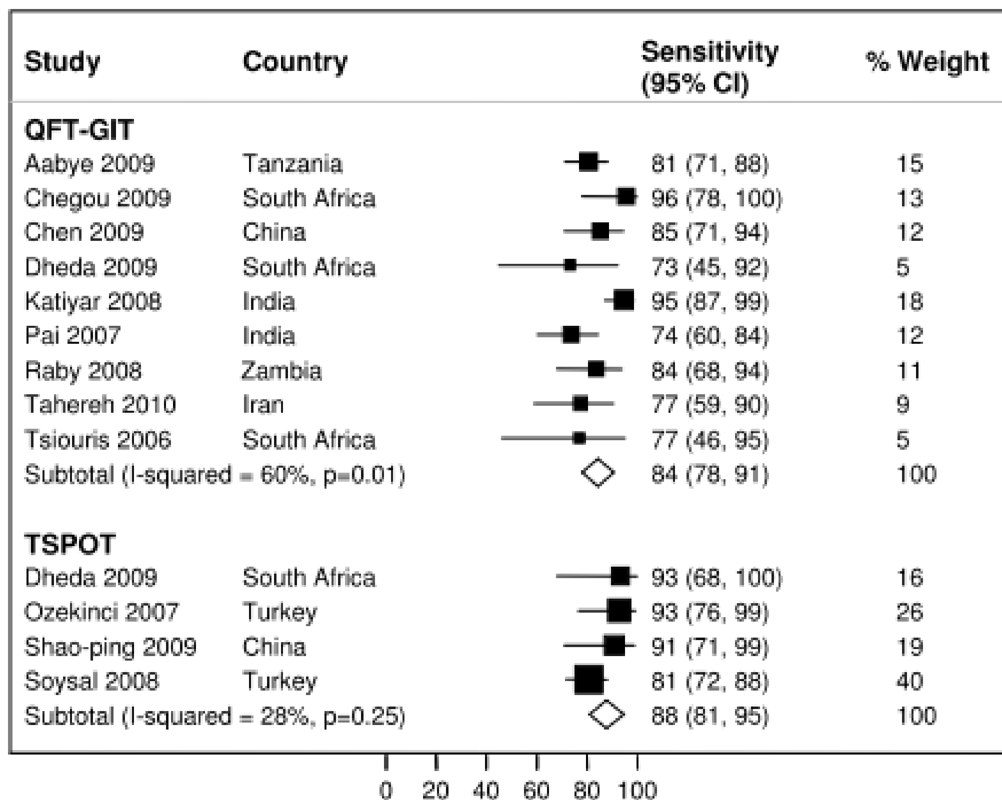


Figure 5. Sensitivity of QuantiFERON-TB Gold In-Tube and T-SPOT.TB in HIV-uninfected persons with confirmed active tuberculosis in low- and middle-income countries.

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

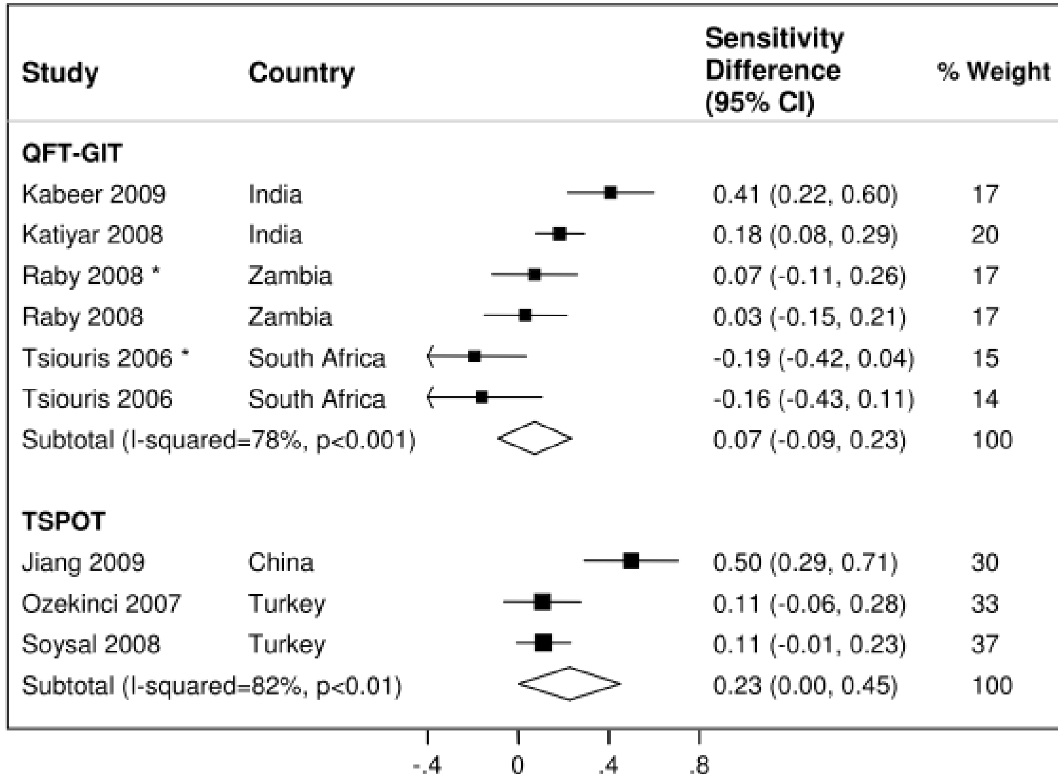
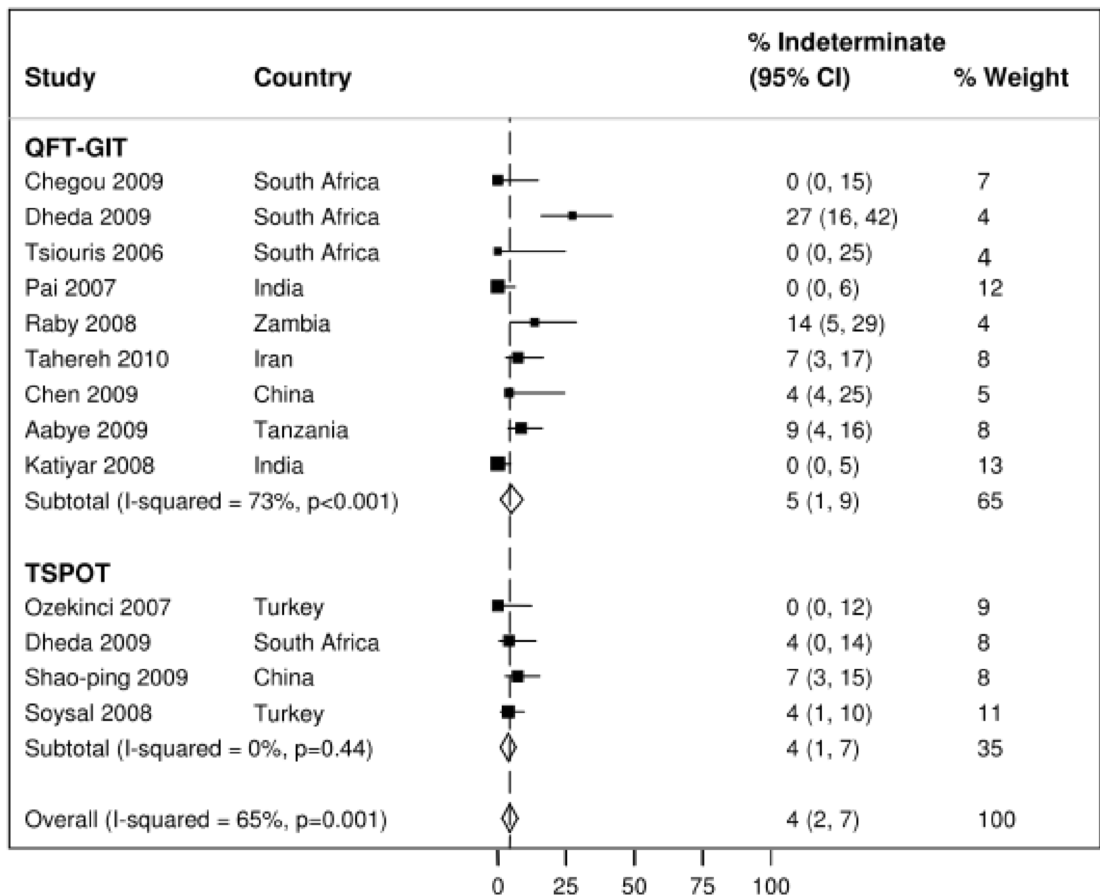


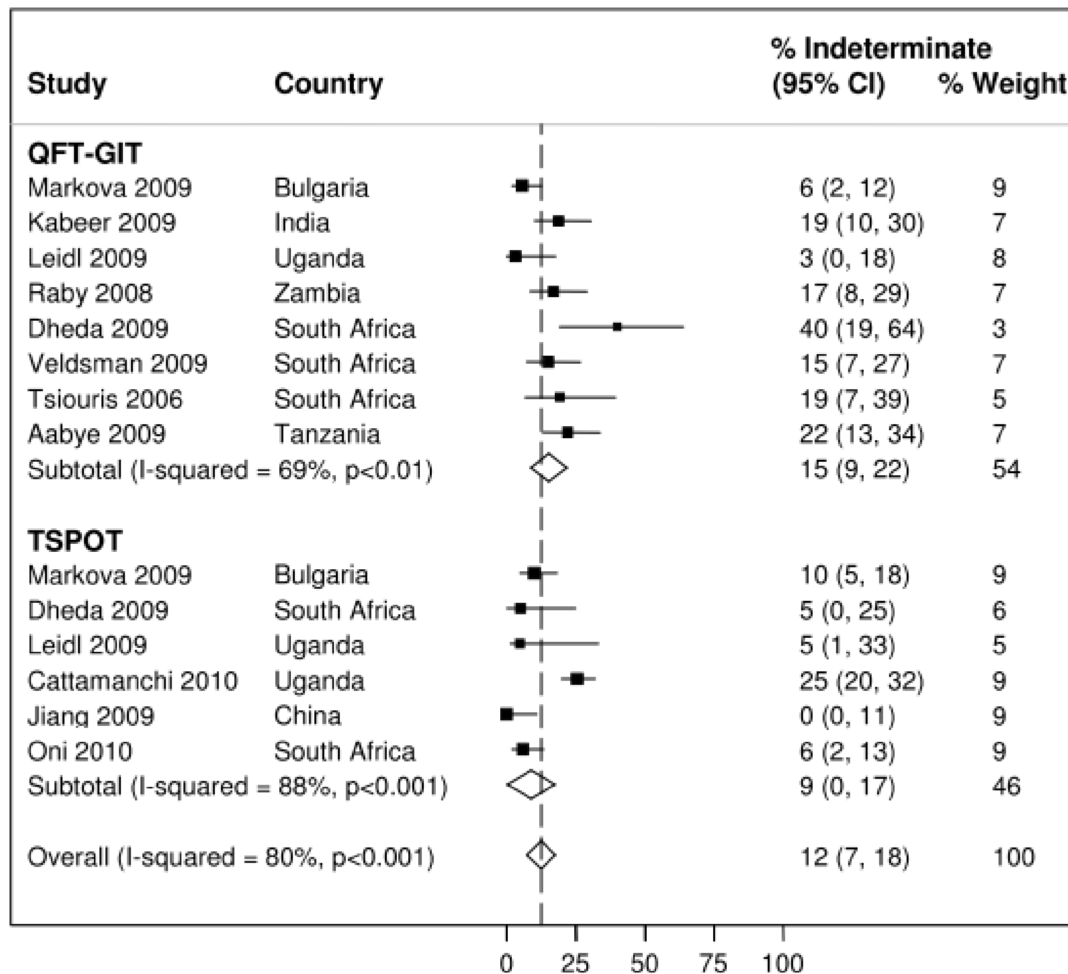
Figure 6. Percent sensitivity difference between IGRA and TST results.

The forest plots display percent differences (IGRA sensitivity - TST sensitivity) for confirmed active pulmonary TB in individual studies and pooled estimates derived from random effects (DerSimonian-Laird) modeling.
* Studies of HIV-infected patients.

John Metcalfe
Division of Epidemiology, University of California, Berkeley



Supplemental Figure 1. Proportion of indeterminate IGRA results among HIV-uninfected individuals in low- and middle-income countries.



Supplemental Figure 2. Proportion of indeterminate IGRA results among HIV-infected individuals in low- and middle-income countries.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Chapter 3. Test Variability of the Quantiferon-TB Gold In-Tube Assay in Clinical Practice

At a Glance Commentary

Scientific Knowledge on the Subject

Although interferon-gamma release assays (IGRAs) are widely used in high-income countries and numerous studies have evaluated their diagnostic performance, there are limited data on the precision of IGRA results.

What This Study Adds to the Field

In the largest precision study of an IGRA to date, we found considerable variability in TB response measured by QuantiFERON-TB Gold In-Tube (QFT-GIT, Cellestis, Australia) upon retesting of the same patient sample. Variability within individuals included differences up to 0.24 IU/ml, in either direction, when the initial TB response was between 0.25 and 0.80 IU/ml. Test results should be interpreted cautiously among individuals with positive QFT-GIT test results less than 0.59 IU/ml.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Abstract

Rationale: Although interferon-gamma release assays (IGRAs) are widely used to screen for *M. tuberculosis* infection in high-income countries, published data on repeatability are limited.

Objectives: To determine IGRA repeatability.

Methods: The study population included consecutive patients referred to The Methodist Hospital (Houston, TX, USA) between August 1, 2010 and July 31, 2011 for LTBI screening with an IGRA (QuantiFERON-TB Gold In-Tube). We performed multiple IGRA tests using leftover stimulated plasma according to a prospectively formulated quality control protocol. We analyzed agreement in interpretation of test results classified according to manufacturer-recommended criteria and repeatability of quantitative TB response.

Measurements and Main Results: During the study period, 1086 test results were obtained from 543 subjects. Per the manufacturer's cut-point, the result of the second test was discordant from that of the first in 28/366 (8%) patients with valid test results, including 13 with an initial negative result and 15 with an initial positive result. Although agreement between repeat test results was high (kappa 0.84; 95% CI 0.79-0.90), the normal expected range of within-subject variability in TB response upon retesting included differences of +/-0.60 IU/ml for all individuals (coefficient of variation (CV), 14%), and +/-0.24 IU/ml (CV 27%) for individuals whose initial TB response was between 0.25 and 0.80 IU/ml.

Conclusions: There is substantial variability in TB response when IGRAs are repeated using the same patient sample. IGRA results should be interpreted cautiously when TB response is near interpretation cut-points.

Introduction

Interferon-gamma release assays (IGRAs) are *in vitro* immunodiagnostic tests that measure effector T cell-mediated interferon-gamma (IFN-γ) response to synthetic *Mycobacterium tuberculosis*–specific polypeptides. The QuantiFERON®-TB Gold In-Tube (QFT-GIT; Cellestis, Carnegie, Australia) is a commercially available IGRA that has been recommended as an alternative to the tuberculin skin test (TST) in targeted screening for *M. tuberculosis* infection. (1)

Although IGRAs are widely used in high-income countries and numerous studies have evaluated their diagnostic performance, there are limited data on the precision of IGRA results. Clinically, such data are essential because treatment decisions could be affected by interpretation of results close to the threshold for a positive test, and for changes above or below this threshold when serial testing is performed. (2) Data on test imprecision, including repeatability (serial testing under identical conditions) and reproducibility (serial testing under changed conditions) (3) (*see* Text Box "Definitions"), are required for CE (Conformité Européene) Marking in the European Union and for premarket approval of *in vitro* diagnostic (IVD) tests by the US Food and Drug Administration (FDA). (4)

Box | Definitions

- Repeatability: The precision of a test when replicated under identical apparent conditions (e.g., same laboratory, operator, apparatus, minimal time interval); a measure of the inherent random error associated with a test. (5)
- Reproducibility: The precision of a test when replicated under different conditions (e.g., different laboratory, operator, apparatus, unspecified time interval). (5)
- TB response: The amount of IFN-γ released in response to *M. tuberculosis*-specific antigens (ESAT-6, CFP-10, or TB7.7), calculated as the difference in IFN-γ concentration in plasma from blood stimulated with antigen minus the IFN-γ concentration in plasma from blood incubated with saline (i.e., Nil). (1)
- Borderline range: TB response between 0.25 and 0.80 IU/ml; derived within our cohort from the quintiles of TB response above and below the manufacturer-recommended cut point.
- Normal expected range of within-subject variability: Mean TB response +/-1.96 repeatability standard deviations. This range about an individual's true value (the value in absence of repeatability variability) would include 95% of repeat measurements for that individual. (6)
- Low positive: A positive test result within a quantitative interval within which a repeat test may be negative based on inherent random error alone (i.e., the normal expected range includes values that cross the manufacturer-recommended cut-point (0.35 IU/ml)).

Repeatability is unaffected by intervening immune response or differences between laboratories, and thus estimates the inherent random error associated with IGRAs. Accounting for such random error when interpreting test results may contribute to improved treatment decisions. Data submitted by the manufacturer to the FDA indicate that QFT-GIT has little test-retest variability (coefficient of variation (CV), 8%). (7) Multiple independent studies have assessed longitudinal changes in TB response, (8-22) although few investigators have examined the repeatability of commercial IGRAs. (8, 10, 11, 20) These reports support a greater amount of test variability than that reported by the manufacturer, though interpretation has been limited by sample size and heterogeneous statistical methods, terminology, and epidemiologic settings.

In order to improve the clinical interpretation of QFT-GIT results, particularly results near the cut-point for a positive test, we analyzed repeatability within a large population of individuals living in a low TB incidence setting.

Some of the results of these studies have been reported previously in abstract form. (23)

Methods

Study Population

The study included consecutive employees of petrochemical companies and other individuals referred by hospitals and private clinicians for routine latent tuberculosis infection (LTBI) screening at The Methodist Hospital (Houston, Texas) between August 1, 2010 and July 31, 2011. Because this population was considered to be at low risk for *M. tuberculosis* infection, we instituted an internal quality control algorithm in which QFT-GIT testing was repeated using leftover stimulated plasma if

the initial test result was not negative with TB response <0.25 IU/ml and nil control <0.10 IU/ml. The quality control algorithm excluded individuals with both low TB response and low nil control because their test result was considered to be unambiguously negative (i.e., the test result would not have been positive even had the nil control response been lower).

Data collection

We abstracted demographic data from a de-identified, clinical database; information on nationality, TST status, and history of previous TB disease was not available.

Two technicians from a CLIA (Clinical Laboratory Improvement Amendments) -certified (24) laboratory accustomed to IGRA research (25-28) performed all QFT-GIT assays according to the manufacturer's instructions using identical instrument settings (laboratory testing details are provided in an online supplement). After the initial round of testing, we followed a prospectively formulated retesting algorithm in which QFT-GIT assays were repeated using leftover stimulated plasma from the initial venipuncture. We performed a second round of retesting in duplicate (*i.e.*, tests 3 and 4) if the first two tests were discordant based on the manufacturer's suggested cut-point (*i.e.,* negative [<0.35 IU/ml] when the initial test result was positive [≥0.35 IU/ml], or vice versa). Initial QFT-GIT assays were classified as indeterminate according to manufacturer instructions (29) and repeated once. In order to maximize clarity of presentation, and because results did not substantially differ, only the first and second tests for each subject were considered in our main analysis.

We interpreted test results according to the consensus of all available tests classified using the manufacturer's recommended cut-point. The decision to initiate LTBI treatment following test analysis was at the discretion of the referring clinician.

Statistical Analyses

For results classified according to the manufacturer-recommended cut-point, we assessed agreement between the first and second QFT-GIT assays with the kappa statistic of inter-rater agreement. (30) For quantitative results, we plotted the difference between the first and second QFT-GIT measurement against their mean, as described by Bland and Altman (see online data supplement). (31) Subjects with very low TB response (TB response < -0.35 IU/mL and < -0.5 times nil control) were excluded. (32) IFN-γ concentrations greater than 10 IU/ml may be unreliable due to the limited linear range of the ELISA reader, and were therefore truncated at 10 IU/ml.

Next, we assessed the repeatability of the ELISA portion of QFT-GIT using a linear mixed effects model fit to the numerical IFN-γ values. Such models allow global estimation of within-person standard deviation optimally weighted for the correlation structure of the repeated measurements. (33, 34) Assuming that 95% of measurements are located within +/- 1.96 SDs, the "normal expected range" of within-subject test repeatability can be calculated by expressing this SD as a percentage of the individual's mean TB response. Because test-retest differences were not normally distributed, we used a resampling procedure based on 10,000 bootstrap iterations of the dataset to verify that 95% of differences were contained within 1.96 SDs of the mean. We performed additional sensitivity analyses to examine the repeatability of QFT-GIT (1) when the initial TB response was in a borderline range (0.25-0.80 IU/ml), and (2) when subjects with TB response greater than 10 IU/ml were excluded, rather than truncated.

All *P* values were two-sided with alpha = 0.05 as the significance level. Data analysis was performed using Stata 12.1 (Stata Corporation, College Station, Texas).

Results

Between August 1, 2010 and July 31, 2011, 3234 individuals were screened for *M. tuberculosis* infection using QFT-GIT. Of these, 2819 (87.2%) were negative, 218 (6.7%) were positive, and 177 (5.5%) were indeterminate per manufacturer recommended criteria. Among those with negative test results, 2671 had a low negative TB response (<0.25 IU/ml) with low nil control (<0.1 IU/ml) and were not considered further (Figure 1). Thus, subjects with negative test results who were analyzed included (1) those with TB response <0.25 IU/ml and nil control ≥0.1 IU/ml, and (2) those with TB response 0.25–0.35 IU/ml, regardless of nil control value. Most indeterminate tests (*n*=175/177) were due to low mitogen response (mitogen minus nil control, <0.5 IU/ml). For purposes of analysis, subjects with indeterminate test results were examined separately from those with determinate test results. Subject demographic and test characteristics are summarized in Table 1.

Test Agreement

When the first two test results for all patients were analyzed according to the manufacturer-recommended cut-point, agreement was high (kappa statistic 0.84; 95% CI 0.79-0.90) (Table 2). Bland Altman plots of test-retest differences across the full and borderline range of TB response are presented as Figures E1a and E1b in the online data supplement.

Retesting Results

Upon retesting, 13/148 (9%) negative results converted to positive and 15/218 (7%) positive results reverted to negative per the manufacturer's cut-point (0.35 IU/ml). Both conversions and reversions were more likely among individuals with an initial borderline TB response (i.e., 0.25-0.349 IU/ml for conversions, and 0.35-0.80 IU/ml for reversions; p <0.001 relative to non-borderline measurements for both). Eighty-six percent (24/28) of all individuals who converted or reverted had an initial TB response within the borderline range (Table 3). Upon retesting, 11/177 (6%) indeterminate tests were re-read as negative, and two positive results (2/218, <1%) were re-read as indeterminate.

Test Variability

A total of 1086 test results from 543 subjects were available for the primary analysis. The median difference in TB response across all determinate tests within an individual was 0.06 IU/ml (IQR 0.02-0.19 IU/ml); the maximum difference in TB response between tests was >1.0 IU/ml in 14/366 (4%) individuals.

The normal expected range of variability for an individual patient included differences of +/-0.60 IU/ml (CV 14%) for all individuals, and +/-0.24 IU/ml (CV 27%) for individuals with initial TB response in the borderline range (0.25-0.80 IU/ml) (Figure 2). Results were similar when analysis included all available tests for each subject. Higher variability occurred in the sensitivity analysis excluding subjects with TB response >10 IU/ml, and among subjects with indeterminate test results (*see* Methods and online data supplement).

Discussion

We analyzed a large population of individuals from a low TB incidence setting to determine the inherent variability of QFT-GIT, with specific attention to individuals with results near the manufacturer-recommended cut-point for a positive test. We found substantial and clinically important variability in QFT-GIT results upon retesting of the same patient sample. This variability is higher than that initially reported by the manufacturer, and has important implications for interpretation of borderline test results and conversion/reversion thresholds for screening programs in low TB incidence settings.

Studies of TST variability since the 1960s have found that spontaneous changes in induration ≤6mm may occur due to host biologic variation, or differences in test administration or measurement. (35-38) These normal limits of test variability contribute to current thresholds (≥10mm) for skin test conversion. (39) Although IGRAs may be more specific than the TST and initially identify fewer persons requiring preventive therapy in some settings, (40) subsequent conversions and reversions may occur with greater frequency. (9, 13, 41, 42) The reasons for this remain uncertain, with possible explanations including a more variable immunologic reactivity of effector vs. central memory T-cells or a greater sensitivity of IGRAs to dynamic changes in the spectrum of LTBI. However, our data suggest that variability inherent to the test rather than host or pathogen factors could explain many IGRA reversions and conversions.

We found that the normal expected range of within-subject variability for the QFT-GIT assay includes a difference in the quantitative TB response up to +/-0.60 IU/ml on retesting. For test results close to the manufacturer-recommended cut-point of 0.35 IU/ml, differences of 0.24 IU/ml, in either direction, are within the normal expected range for test variability. Thus, as shown in Figure 2, positive test results less than 0.59 IU/ml could reasonably be expected to revert to negative based on the inherent variability of the test alone. Of note, in our cohort 20% (*n*=43/218) of positive results fell within this range, and 23% (*n*=10/43) of these individuals reverted to negative. Weighing four to nine months of potentially unnecessary preventive antibiotic therapy (and attendant risk of adverse drug effects) against the probability of reactivation TB, screening programs in low incidence settings may choose to limit false-positives by accepting a higher than recommended threshold for test positivity. Indeed, recognizing the limited positive predictive value of IGRAs in low-risk individuals, (43) some centers in the U.S. have already instituted such policies. (44, 45) Ultimately, as with the TST, conversion thresholds will depend on risk of future active TB disease, the epidemiologic setting, and, possibly, the magnitude of the quantitative response, (46) although establishing the evidence base for such thresholds will be challenging. (2)

Our findings build upon previous literature documenting variable amounts of QFT-GIT imprecision and calls for borderline zones of various ranges. (8, 10, 18-20, 47) Reproducibility studies of longitudinal TB response in low TB incidence settings over time periods of days to years have demonstrated changes in TB response of 16%-80% upon retesting. Repeatability data, usually from high TB incidence regions, have been less often reported. Further, interpretation of these data has been challenging due to the differing methods used to assess variability. Because this variability has traditionally been determined by dividing the pooled standard deviation by the overall mean TB response (i.e., use of coefficient of variation, or CV), persons with positive test results will have less variability, as a percentage of the mean, relative to persons with negative test results. This may account for the low variability reported by the manufacturer (CV 8.4%) during the FDA approval process, because their estimate was based on a small number of subjects who had evidence of active TB and

were likely to have had high mean TB response (7); low CV% despite relative large differences in magnitude of TB response among high positive tests is a known limitation of this measure of repeatability. (48) Independent investigators have reported results from a variety of repeatability methods. In India, Veerapathran et al. (8) used log transformation and linear mixed effects analysis to determine that "a second test performed on the same blood sample will…typically be 19% greater than the initial test." Van zyl Smit et al. (11) (South Africa) repeated readings of the same ELISA plate over a two-hour period and reported no significant variability using analysis of variance techniques. Both Detjen et al. (8) (South Africa) and Ringshausen et al. (20) (Germany) reported a high intraclass correlation coefficient (ICC; 0.991 and 0.995, respectively), suggesting excellent repeatability. Although the ICC is an important measure of repeatability, it does not provide an intuitive answer as to how much a patient's initial result might change upon retesting. Given that total imprecision is always greater than within-subject imprecision, (49) our results may be considered conservative in the context of serial testing programs where extraneous biologic variability will also be a consideration. Of note, QFT-GIT variability is poor compared with new molecular TB tests such as Xpert® MTB/RIF (Xpert; Cepheid, Inc, Sunnyvale, CA, USA) quantitative PCR (CV 6%), (50) or commonly used serologic assays. (19)

The 2010 Centers for Disease Control and Prevention (CDC) Guidelines on Interferon-γ Release Assays recommend that clinical laboratories report quantitative in addition to qualitative (i.e., positive or negative) test results. (1) Although some utility in predicting diagnosis (51) and progression to active TB (52) may exist, in low TB incidence settings reporting of quantitative test results will be most important in identifying "borderline" subjects within screening programs, for whom withholding preventive treatment pending retesting the following year may be justified.

Our study has some potential limitations. Because the majority of negative test results with TB response less than 0.25 IU/ml were not analyzed, we cannot draw conclusions as to the inherent variability of results in this group. However, we focused on test results most likely to represent clinical dilemmas within screening programs in low-incidence settings. In addition, given the operational nature of our study, we were not able to control for some potential sources of variability (e.g., operator-dependent variability, lot-to-lot variability) which may have compromised assessment of repeatability in its strictest sense. However, because our study was carried out in the course of routine clinical practice, the external generalizability of our findings is reinforced.

In conclusion, in the largest precision study of the QFT-GIT to date, we found that QFT-GIT has a normal expected range of within-subject test variability of +/-0.60 IU/ml (CV 14%) overall, and +/-0.24 IU/ml (CV 27%) among subjects with borderline TB response near the manufacturer-recommended cut-point (0.25-0.80 IU/ml). Our results suggest that in low TB incidence countries, low-risk individuals with a positive QFT-GIT result less than 0.59 IU/ml should be interpreted cautiously.

Table 1. Characteristics of the Study Population

Characteristic	Total, *n*=543
Age, years, mean (+/- SD)	43 (+/- 18)
Male, %	66
Origin of Referral,* %	
Petrochemical Corporation	74
Hospital/Clinician	26
Race/Ethnicity, %	
White	41
Black	9
Hispanic	11
Asian	10
American Native	1
Unknown	28
Number of replicates obtained per subject,† %	
1	92
2	1
3	7

Definition of abbreviations: SD, standard deviation.

Values are percentages unless otherwise stated. All categories are mutually exclusive.

*Origin of referral includes determinate tests only.

† Only the first and second tests for each subject were included in the main analysis.

Table 2. Repeatability of QFT-GIT Test Interpretation

Test Result Sequence	Total Subjects (*n*=543)
Concordant	**502 (93%)**
Positive / Positive	201
Negative / Negative	135
Indeterminate / Indeterminate	166
Discordant	**41 (7%)**
Positive / Negative	15
Negative / Positive	13
Indeterminate / Negative	11
Positive / Indeterminate	2

Definition of abbreviations: QFT-GIT, QuantiFERON® Gold-In Tube.

Note: Categorization of all available tests is provided as Table E2 in the online data supplement.

Table 3. Repeatability of QFT-GIT Test Results Stratified by Quantitative TB Response

Baseline TB Response (IU/ml)	Total Subjects	Conversion, *n* (%)*	Reversion, *n* (%)*
All subjects	366	13/148 (9)	15/218 (7)
<0.25	106	1/106 (<1)	--
0.25-0.34	42	12/42 (27)	--
0.35-0.80	66	--	12/66 (18)
0.81-3.0	76	--	2/76 (3)
3.1-9.92	76	--	1/76 (1)

Definition of abbreviations: QFT-GIT, QuantiFERON® Gold-In Tube; IFN-γ, interferon-gamma.

Note: Initial indeterminate test results were not included.

* Denominator for conversion or reversion is equal to the total and stratified population at risk (i.e., for conversion, denominator includes those initially testing negative; for reversion, denominator includes those initially testing positive)

Figure Legends

Figure 1. Study Flow Diagram

Positive, negative, and indeterminate tests were classified according to manufacturer's recommended criteria.. (29) Analyzed QFT-GIT negative test results were (1) negative <0.25 IU/ml with nil control ≥ 0.10 IU/ml, or (2) negative 0.25-0.35 IU/ml, regardless of nil control response. Only the first and second tests for each subject were included in the main analysis.

*Very low TB response was defined as TB response < -0.35 IU/mL and < -0.5 times nil control. (32)

Figure 2. Normal Expected Variability of Borderline TB Response

True mean TB response (square, defined as the average value of an unlimited number of measurements taken under the same conditions) and associated 95% variability (+/- 1.96 standard deviations (SD); the solid line to each side of the square represents 1.96 SD). Given an estimated true negative mean TB response of 0.349 IU/ml (just below test cut point for positivity) and a normal expected range of within-subject variability of +/- 0.24 IU/ml, 95% of subjects will demonstrate variability between 0.11 IU/ml and 0.59 IU/ml. Thus, the 'low positive' zone (grey shading) is defined as the interval (0.35 IU/ml–0.59 IU/ml) within which a positive result could be expected to revert to negative on retest based solely on the inherent variability of the test. Measured TB response greater than 0.59 IU/ml would be unlikely to be associated with an estimated true mean TB response less than 0.35 IU/ml, and thus should be considered to indicate a true positive value.

The repeatability of QFT-GIT assays were assessed using a linear mixed effects model fit to the numerical IFN-γ values. Borderline TB response was defined as IFN-γ concentration 0.25 IU/ml-0.80 IU/ml; the manufacturer-recommended cut point (0.35 IU/ml) is demonstrated as a dashed line.

Figure E1a. Bland-Altman Plot for Quantitative TB Response

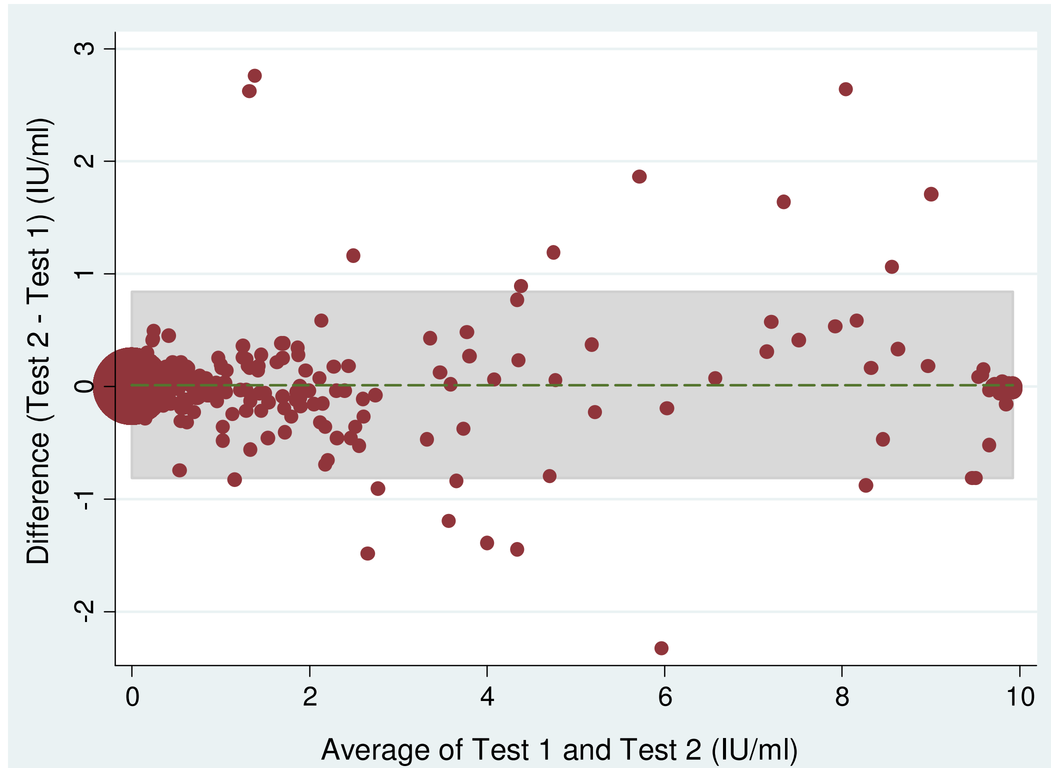
Bland-Altman analysis for first repeated measurement (Test 1 vs. Test 2) across full range of test values showed a mean difference of 0.013 IU/mL and 1.96 standard deviations (1.96 SD) of ± 0.826 IU/ml. The lower and upper limits of agreement (mean difference ± 1.96 SD) are -0.814 IU/mL and +0.839 IU/mL, respectively.

Figure E1b. Bland-Altman Plot for Borderline Quantitative TB Response

Bland-Altman analysis for first repeated measurement (Test 1 vs. Test 2) for borderline test values (0.25-0.8 IU/ml) showed a mean difference of 0.02 IU/mL and 1.96 standard deviations (1.96 SD) of ± 0.27 IU/ml. The lower and upper limits of agreement (mean difference ± 1.96 SD) are -0.24 IU/mL and +0.29 IU/mL, respectively.

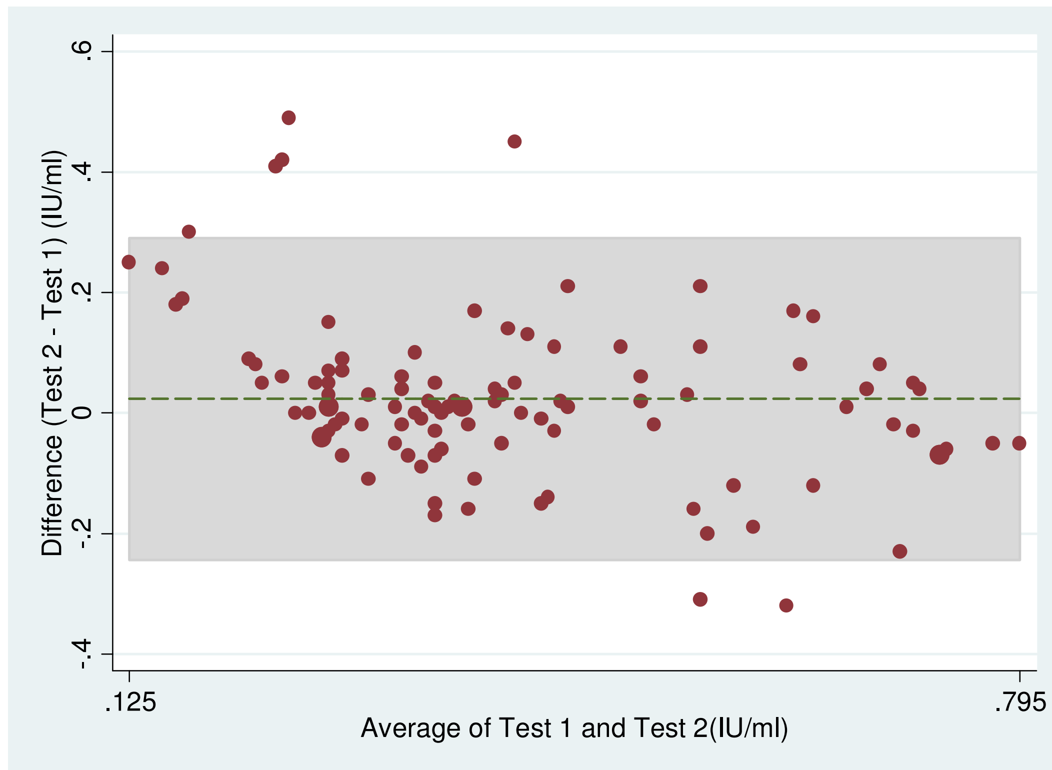
Chapter 3 Supplementary Material. Test Variability of the Quantiferon-TB Gold In-Tube Assay in Clinical Practice

Figure E1a. Bland-Altman Plot for Quantitative TB Response



Bland-Altman analysis for first repeated measurement (Test 1 vs. Test 2) across full range of test values showed a mean difference of 0.013 IU/mL **and 1.96 standard deviations (1.96 SD) of ± 0.826** IU/mL**.** The lower and upper limits of agreement (mean difference ± 1.96 SD) are -0.814 IU/mL and +0.839 IU/mL, respectively.

Figure E1b. Bland-Altman Plot for Borderline Quantitative TB Response



Bland-Altman analysis for first repeated measurement (Test 1 vs. Test 2) for borderline test values (0.25-0.80 IU/ml) showed a mean difference of 0.02 IU/mL **and 1.96 standard deviations (1.96 SD) of ± 0.27** IU/mL**.** The lower and upper limits of agreement (mean difference ± 1.96 SD) are -0.24 IU/mL and +0.29 IU/mL, respectively.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Chapter 4. Use of Microscopic-observation Drug Susceptibility Assay among Drug-Resistant Tuberculosis Suspects in Harare, Zimbabwe

Abstract

Introduction: Limited data exist on use of the microscopic-observation drug-susceptibility (MODS) assay among MDR-TB suspects in a high HIV-prevalence setting.

Methods: We retrospectively reviewed available clinical and drug susceptibility data for drug-resistant TB suspects referred for culture and drug-susceptibility testing between April 1, 2011 and March 1, 2012. The diagnostic accuracy of MODS was estimated against a reference standard including Löwenstein-Jensen (LJ) media and manual liquid (BACTEC MGIT) culture. The accuracy of MODS drug-susceptibility testing (DST) was assessed against a reference standard absolute concentration method.

Results: One hundred thirty-eight sputum samples were collected from 99 drug-resistant TB suspects; in addition, six previously cultured MDR strains were included for assessment of DST accuracy. Among persons with known HIV status, 39/59 (66%) were HIV-infected. Eighty-six percent of patients had prior TB treatment history, and 80% of individuals were on antituberculous treatment at the time of sample collection. *M. tuberculosis* was identified among 34/98 (35%) MDR-TB suspects by reference standard culture. Overall MODS sensitivity for *M. tuberculosis* detection was 85% (95% CI, 69–95%) and specificity was 93% (95% CI, 84–98%); diagnostic accuracy did not statistically differ by HIV status. The median time to positivity was significantly shorter for MODS (7 days; IQR 7–15 days) than MGIT (12 days; IQR 6–16 days) or LJ (28 days; IQR 21-35 days; $p<0.001$). Of 33 specimens with concurrent DST results, sensitivity of the MODS assay for detection of resistance to isoniazid, rifampin, and MDR-TB was 88% (95% CI, 68–97%), 96% (95% CI, 79–100%), and 91% (95% CI, 72–99%), respectively; specificity was 89% (95% CI, 52–100%), 89% (95% CI, 52–100%), and 90% (95% CI, 56–100%), respectively.

Conclusion: In a high HIV-prevalence setting, MODS diagnosed TB and drug-resistant TB with high sensitivity and shorter turnaround time compared with standard culture and DST methods.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Background

The World Health Organization (WHO) estimates a global prevalence of 650,000 cases of multidrug-resistant tuberculosis (MDR-TB), contributing to 150,000 annual deaths.[1] Despite these staggering figures, only 6% of worldwide cases are detected. [2] In particular, the emergence of MDR/XDR-TB in the southern Africa region has been associated with high mortality, [3,4] and may be seriously underestimated. [5-7]

Accurate, timely, and affordable drug susceptibility testing (DST) for surveillance and patient management in high burden countries is urgently needed. In a recent review of global MDR-TB trends, only 18 of 46 countries from the WHO African Region had nationwide drug resistance data, and only five had reported data since 2002. [6] The microscopic-observation drug-susceptibility (MODS) assay is an accurate, inexpensive, liquid culture-based diagnostic test that has been endorsed by the WHO for rapid screening of patients suspected of having MDR-TB. [8,9] Despite calls for an expanded role for use of the MODS assay in high HIV-prevalence regions, [10] data among HIV-infected TB and MDR-TB suspects [11,12] remain limited.

In order to examine the diagnostic accuracy of the MODS assay for *M. tuberculosis* detection and direct DST among MDR-TB suspects in a high HIV-prevalence region, we examined test results against a solid and liquid culture reference standard.

Methods

Study Population

We retrospectively reviewed available clinical and drug susceptibility data for drug-resistant TB suspects referred for culture and drug-susceptibility testing between April 1, 2011 and March 1, 2012. Drug-resistant tuberculosis suspects were defined by either (1) history of prior treatment (> 1 month, classified according to World Health Organization criteria [1]) or (2) contact to an individual with known or suspected drug-resistant TB. Samples were obtained from patients undergoing routine work-up and from participants of ongoing clinical studies. In addition, six previously cultured MDR isolates were included for assessment of DST accuracy. Participants of ongoing clinical studies provided written informed consent, and ethical approval was obtained from the Medical Research Council of Zimbabwe, the Institutional Review Board of the Biomedical Research and Training Institute, and the UCSF Human Research Protection Program. De-identified data from patients not participating in ongoing clinical studies and seen in the course of routine medical practice did not meet the definition for human subjects, and were exempt from ethics review.

Laboratory Methods

The Biomedical Research and Training Institute (BRTI) Tuberculosis Laboratory within the National Microbiology Reference Laboratory (NMRL) is a center for Trials of Excellence in Southern Africa (TESA). BRTI collaborates with the Ministry of Health and Child Welfare (MOHCW) in laboratory capacity building, and regularly undergoes External Quality Assurance (EQA) of DST for first-line anti-TB drugs. The most recent Centre for American Pathology (CAP) assessment in 2012 demonstrated 100% agreement for isoniazid, rifampicin, ethambutol, and streptomycin.

Sputum specimens were transported to the BRTI Tuberculosis Laboratory for standard culture, DST, and MODS testing within 48 hours of collection. Each sample was divided into two aliquots: the first aliquot underwent sputum AFB smear examination, decontamination, culture, and DST according to published guidelines, [13] and the second aliquot underwent MODS testing.

In preparation for reference standard culture and DST, the first aliquot was digested using the 4% sodium hydroxide method. The resuspended sediment was used to make a concentrated smear and inoculated onto Löwenstein-Jensen (LJ) media and in BBL™ MGIT™ Mycobacterial Growth Indicator Tubes (Becton Dickinson, Sparks, MD). MGIT broth tubes were continuously monitored for 40 days for *M. tuberculosis* growth by use of a manual MGIT reader. [14] Ziehl-Neelsen staining was used to confirm growth of Mycobacteria in all test positive tubes. MGIT cultures that had a mixture of mycobacteria and other bacterial contamination from 21 to 40 days were re-decontaminated and re-cultured. All positive cultures by MGIT were identified as *M. tuberculosis* complex by MPT64 antigen detection, [15] or by growth at different temperatures if rapid kit ID was negative. LJ media were monitored for *M. tuberculosis* growth weekly for eight weeks.

Indirect DST was performed on all positive isolates using absolute concentration measurement (MIC) on LJ media to determine susceptibility to isoniazid (0.2 and 1 ug/ml); rifampicin (32 and 64 ug/ml); ethambutol (2.8 and 4 ug/ml); and streptomycin (8 and 16 ug/ml). [16] Time to detection of growth and contamination rates were recorded for each type of culture medium.

The second sputum aliquot underwent MODS testing in accordance with published standard operating procedures. (28) All MODS test results were interpreted without knowledge of the results of the reference standard. Briefly, the sample was decontaminated using a Sodium hydroxide- Sodium citrate-NaCl solution and inoculated into Middlebrook 7H9 liquid broth containing OADC and PANTA. 900µl of this sample-broth mixture was aliquoted into each of four well columns in a 24 microtitre well plate; for each patient sample, the first two wells were drug-free, the third well contained 100µl isoniazid at 0.4 ug/ml concentration, and the fourth well contained rifampicin at 1 ug/ml. Plates were incubated at 37^0C. MODS cultures were examined using an inverted light microscope at 340 magnification every day from day 4 through Day 21 and weekly up to the end of 40 days incubation. Positive MODS cultures were identified by presence of characteristic cord formation in the drug-free control wells at time of detection of growth in drug-free wells.

During the study period, the lab transitioned from standard MODS to use of the TB MODS Kit™ (Hardy Diagnostics, Santa Maria, CA USA). Briefly, specimens were decontaminated as for standard MODS and inoculated into commercially prepared vials of Middlebrook 7H9 liquid broth containing OADC to which 100µl of PANTA was added prior to sample inoculation. Direct patient samples were inoculated into two drug-free plate wells, one plate well containing isoniazid at 0.4 ug/ml concentration, and one plate well containing rifampicin at 1 ug/ml, with examination performed as with standard MODS. Validation reports from the manufacturer document a diagnostic accuracy similar to or greater than standard MODS. [17]

Statistical Analysis

We calculated proportions with exact binomial 95% confidence intervals (CI) for the primary analyses of sensitivity, specificity, positive predictive value, and negative predictive value. A positive reference result for *M. tuberculosis* detection was defined as a positive culture on either LJ or MGIT culture. Time to positivity (TTP) was defined as the time from inoculation of the specimen in the laboratory to

report of test positivity. For calculation of sensitivity, we included indeterminate or contaminated MODS results in the denominator if they occurred in individuals with culture-positive TB; for specificity calculation, indeterminate or contaminated MODS results were excluded from analysis. For categorical variables, we compared proportions using chi-square tests. For continuous variables, we compared medians using the Wilcoxon rank-sum test. TAT was determined using survival analysis techniques, with treatment arms compared using the log-rank test. All *P* values were two-sided with alpha = 0.05 as the significance level. Data analysis was performed using Stata 12.1 (Stata Corporation, College Station, Texas).

Results

Patients and Samples

One hundred thirty-eight sputum samples were collected from 99 drug-resistant TB suspects, of whom 23 (40%) were female and the median age was 37 years (interquartile range [IQR]: 27–44). Retreatment category was available for 86 (87%) patients; 12 (14%) had no prior TB history but were contacts to known or suspected MDR-TB cases (Table 1). Among persons with known HIV status, 39/59 (66%) were HIV-infected. One patient was excluded from further analysis due to insufficient sample quantity for MODS (Figure 1).

M. tuberculosis detection

M. tuberculosis was identified in 34/98 (35%) clinical samples from either solid or liquid culture. Of these, 19/34 (56%) were MDR, 9/34 (27%) were drug-susceptible, 1/34 (3%) was rifampin-monoresistant, and 5/34 (15%) identified *M. tuberculosis* but were contaminated prior to finalization of DST. Eighteen percent (n=6/34) of cases were smear-negative. Overall MODS sensitivity for *M. tuberculosis* detection was 85% (95% CI, 69–95%) and specificity was 93% (95% CI, 84–98%) when compared with the reference standard of solid or manual liquid culture (Table 2). Negative predictive value for excluding TB among drug-resistant TB suspects was 92% (95% CI, 82–97%). In an analysis stratified by HIV status, neither sensitivity (85% for HIV-positive, 86% for HIV-negative; p=0.53) nor specificity (92% for HIV-positive, 100% for HIV-negative; p=0.31) demonstrated statistically significant differences.

Among the five TB cases not detected by the MODS assay (i.e., “false-negatives”), one was MODS indeterminate with mixed *M. tuberculosis*/nontuberculous mycobacteria (NTM) noted on manual MGIT culture, one was MODS contaminated, two were smear-negative with growth detected by MGIT following prolonged incubation (>21 days), and one was sputum smear-positive. Among four reference standard-negative, MODS-positive samples (i.e., “false-positives”), two were MODS-positive following prolonged incubation (>21 days), and one was proven to be multidrug resistant upon manual MGIT culture of a separately collected patient specimen.

Initial contamination (including specimens that were later successfully decontaminated) was similar for the MODS assay (*n*=9/138 (6.5%) specimens), manual MGIT (*n*=15/138 (10.9%), and LJ culture (*n*=8/138 (5.8%); p=0.81). Although power was limited, no difference in contamination was noted with use of the TB MODS Kit™ (6%) versus standard MODS (6%; p=0.43 for difference).

Drug Susceptibility Testing

Of 29 specimens positive by both MODS and reference standard solid or manual MGIT culture, two were absolute concentration method-indeterminate due to contamination of sub-culture. Therefore, 27 directly inoculated patient specimens and six previously cultured specimens had concurrent MODS isoniazid and rifampin wells for comparison with the absolute concentration method. Among directly inoculated samples, resistance to isoniazid was detected in 18/27 (67%), to rifampin in 18/27 (67%), and to both isoniazid and rifampin (i.e., MDR-TB) in 17/27 (63%) by the reference standard. Overall sensitivity of the MODS assay for detection of resistance to isoniazid, rifampin, and MDR-TB was 88% (95% CI, 68–97%), 96% (95% CI, 79–100%), and 91% (95% CI, 72–99%), respectively; specificity was 89% (95% CI, 52–100%), 89% (95% CI, 52–100%), and 90% (95% CI, 56–100%), respectively (Table 3).

Time to Positivity for Detection of *M. Tuberculosis* and Drug Resistance

Overall, the median time to culture positivity was significantly shorter for MODS than for the manual MGIT liquid or LJ cultures (MODS 7 days [IQR 7–15 days] vs. MGIT 12 days [IQR 6–16 days] vs. LJ 28 days [IQR 21-35 days]; $p<0.001$) (Figure 2). Median time to positivity for MODS MDR-TB diagnosis (7 days [IQR 7–15 days]) was significantly shorter than that for the absolute concentration method (71 days [IQR 51–75 days]; $p<0.001$).

Discussion

This operational study evaluated the performance of the MODS assay among drug-resistant TB suspects living in a high HIV-prevalence setting. MODS detected *M. tuberculosis* and *M. tuberculosis* drug resistance with high sensitivity and shorter time to positivity compared with reference standard culture and DST methods. Given the expanding global syndemic of HIV and MDR-TB and continued lack of an affordable, accurate, and rapid point-of-care test, these findings have implications for other limited-resource settings. [18,19]

Zimbabwe has the fourth highest TB incidence per capita (782/100,000) in the world, [2] with up to 80% of active TB cases occurring among individuals co-infected with HIV. [20] Although dramatic reductions in HIV prevalence have occurred, 16% of the adult population remains HIV-infected. [21] The World Health Organization estimates the prevalence of MDR-TB in Zimbabwe among patients with a prior history of TB treatment to be 8.3% (95% CI, 3-20%),[1] though these data originate from a national drug-resistance survey undertaken in 1995, and the true extent of drug resistant-TB in Zimbabwe is unknown. That MDR-TB incidence has increased in the country in the context of severe economic destabilization, challenges to tuberculosis control, and population displacement has been suggested, [22] though evidence in support of this are lacking.

Expanded capacity to perform DST in high burden settings is a critical need. In countries where mycobacterial culture is not routinely utilized, failure of one or more regimens of TB drugs is typically a prerequisite for referral for DST, and 12 or more months often elapse from clinical presentation to MDR-TB diagnosis. Given high early mortality [23] and the potential for ongoing transmission, [24] expedited diagnosis and early institution of effective therapy is life-saving and a critical public health mandate. Although debate continues with regards to the best scale-up option for DST in resource limited settings, [25] the high accuracy, low cost, ability to discern both isoniazid and rifampicin resistance, relative ease of operational implementation and short turnaround time should make MODS a strong consideration.

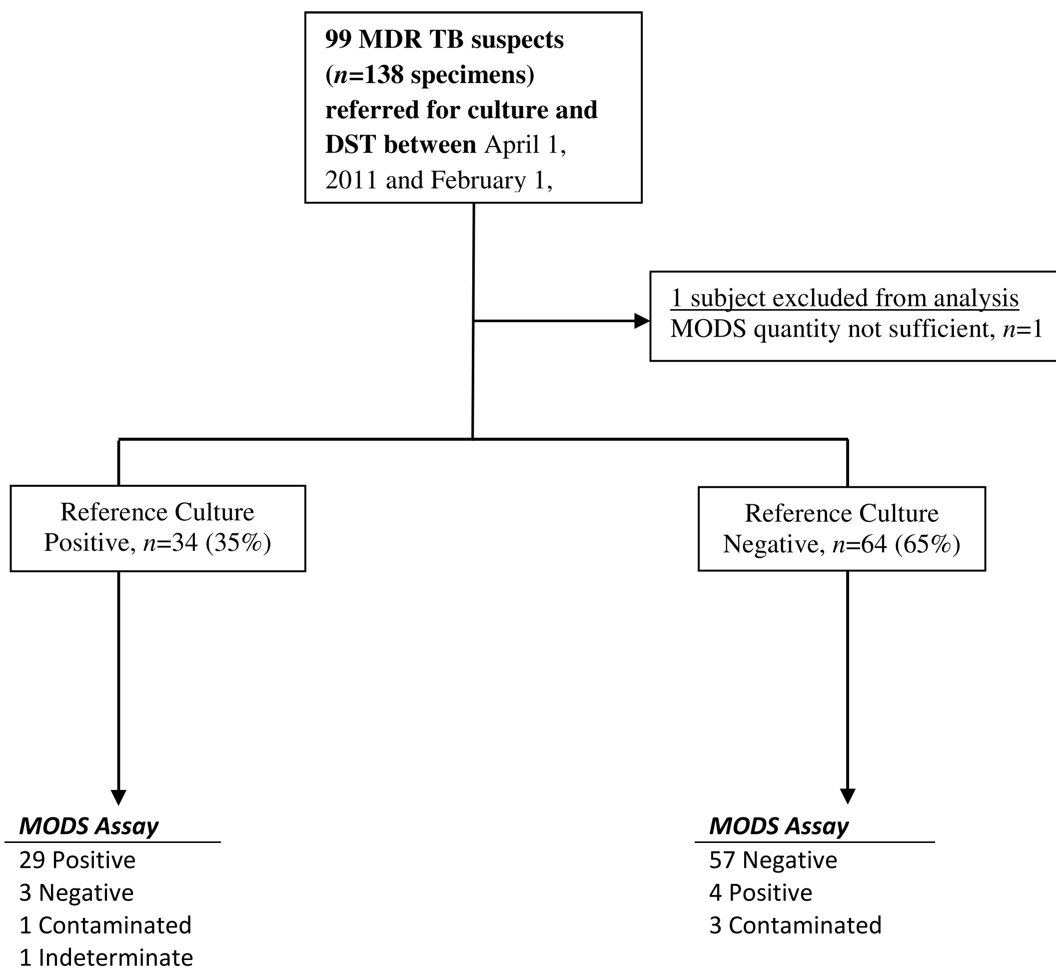
Consistent with the single other study assessing MODS diagnostic accuracy among TB suspects in a high-HIV prevalence region, [11] we found somewhat lower sensitivity for *M. tuberculosis* detection than that reported from other settings. [26] In our study, most false-negative specimens either required prolonged incubation prior to positivity, or were considered false-negative due to MODS contamination. Although culture contamination in our study was not dissimilar to that reported by other investigators, [8,26] contamination of liquid mycobacterial cultures is a known challenge for routine laboratories in Sub-Saharan Africa. [27] Sensitivity for detection of isoniazid and rifampicin drug resistance was similar to previously reported studies, [8] though negative predictive value was less optimal due to the high prevalence of drug resistance noted among MDR-TB suspects in this high burden setting. The sensitivity for detection of isoniazid resistance could be increased through use of a lower MIC (0.1ug/ml) cutpoint in MODS, including within the commercial kits. [28] Of note, in addition to the chosen reference standard, the reported accuracy of drug resistance detection is dependent upon choice of denominator for analysis. With a denominator including all patients with reference standard culture-positive disease, MODS sensitivity and negative predictive value for detection of drug resistance would be marginally lower.

Although upfront costs are higher relative to standard, noncommercial MODS ($3.50 per test at low-income country negotiated price versus $1.48 for standard MODS) [26] and FDA-approval has not yet been obtained, it has been anticipated that use of the TB MODS Kit™ (Hardy Diagnostics, Santa Maria, CA USA) will improve biologic security, attention to published standard operating procedures, and adherence to quality assurance systems. While validation data reported by the manufacturer is excellent, [17] diagnostic accuracy studies by independent investigators are underway.

A strength of our study is its operational, real-world nature. However, threats to internal or external validity include the following. First, as in many settings, routine DST of retreatment TB cases in Harare, Zimbabwe is policy but not yet standard practice, and our sample must be regarded as one of convenience. Further, due to our role as a routine clinical laboratory, we were unable to standardize specimen collection and processing, and some false-negative results may have been due to suboptimal quality in these areas. Second, power was limited to detect meaningful differences in our HIV-stratified analysis. However, that our point estimates are similar to a recent adequately powered study from a similar setting [11] lends confidence to our results. Third, we were unable to undertake a comprehensive microbiologic, molecular, and epidemiologic investigation into discordant cases. Our use of a reference standard including both solid and liquid culture methods and, often, multiple samples per patient allowed for rigorous definition of true positive results. Lastly, diagnostic accuracy is a surrogate for patient-important outcomes such as time to treatment initiation and mortality, and we were unable to assess these directly.

In conclusion, MODS detected *M. tuberculosis* and *M. tuberculosis* drug resistance with high sensitivity and more rapid time to positivity compared with standard culture and DST methods. Further, no detectable differences in diagnostic accuracy were noted for HIV-infected individuals. Prompt treatment of patients with MDR-TB and screening of their contacts will be essential to prevent further spread of drug-resistant *M. tuberculosis*; that this will occur within the context of continued socioeconomic stabilization and improved health service delivery is a strong hope. Studies focused on patient-important outcomes, along with population-based sampling methods to generate valid estimates of the prevalence and incidence of MDR-TB in modern-era Zimbabwe, are urgently needed.

Figure 1. Study Flow Diagram

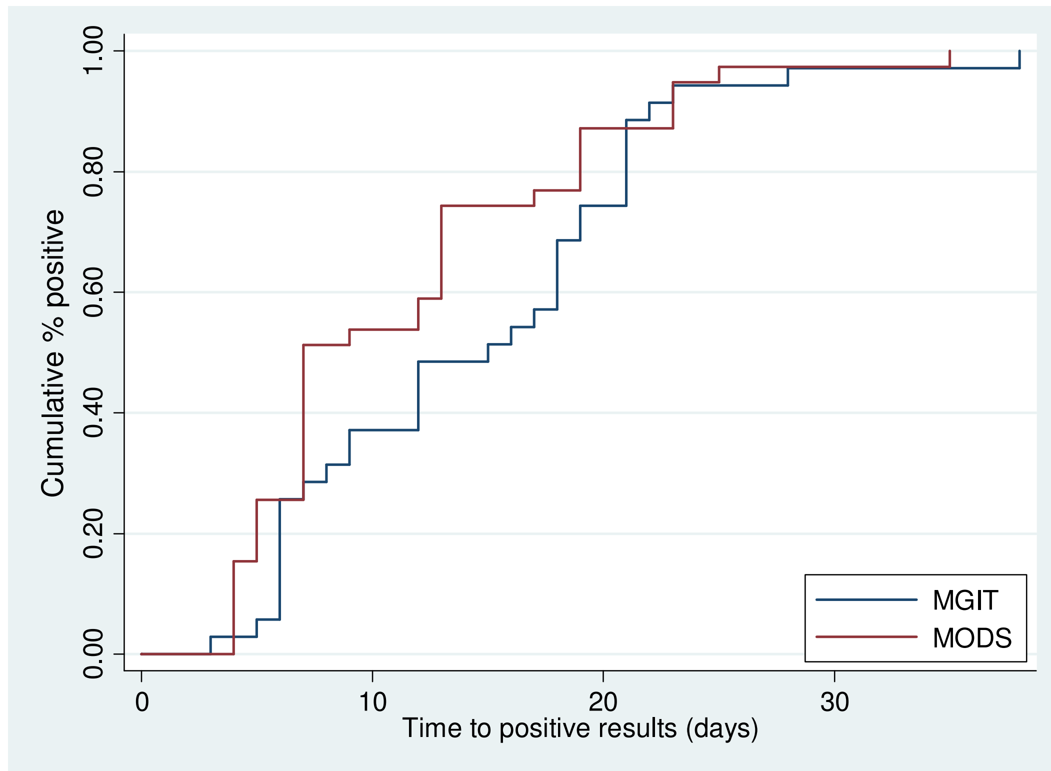


MODS=microscopic-observation drug-susceptibility.

† Six additional previously cultured isolates of known MDR status were included for analysis of drug susceptibility testing and are not included here; all were MODS-positive.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Figure 2. Kaplan-Meier Curves of Time to *M. tuberculosis* Detection



Time to positivity for *Mycobacterium tuberculosis* detection for microscopic-observation drug-susceptibility (MODS) and reference standard culture. Median time to positivity was significantly shorter for MODS than for the manual mycobacterial growth indicator tube (MGIT) MODS 7 days [IQR 7–15 days] vs. MGIT 12 days [IQR 6–16 days]; p<0.001.

Table 1. Characteristics of the Study Population

Characteristic	**Total, *n*=99**
Age, years, median (IQR)	37 (27-44)
Male, %	60
HIV-infected,* %	66
CD4+ T-cell count, median cells/uL (IQR)[†]	160 (84-285)
Site of referral, %	
Outpatient clinic	73
Inpatient ward	27
Reason for referral, %	
Default	6
Relapse	20
Treatment failure, Category I	20
Treatment failure, Category II	13
Contact with known/suspected MDR case	10
Other Retreatment	12
Unknown/Not Recorded	19
TB treatment (any) at time of sample collection, %	80
Sputum AFB smear result, %	
AFB smear-negative	70
AFB smear-positive	30
Number of samples collected, %	
One	82
≥ Two	18
MODS testing format, %	
Standard MODS	64
TB MODS Kit™ [‡]	36

Values are percentages unless otherwise stated. All categories are mutually exclusive. The denominator for each characteristic excludes missing or unknown values unless otherwise stated.
Definition of abbreviations: MODS, microscopic-observation drug-susceptibility assay.

*Among persons with known HIV status (*n*=59/99 (60%))

[†] Available for *n*=27/59 (46%) HIV-infected persons

[‡] Hardy Diagnostics, Santa Maria, CA USA

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Table 2. Comparison of the Microscopic-observation Drug-Susceptibility (MODS) Assay with Reference Standard Culture for Detection of *M. tuberculosis*

MODS Assay	Reference Standard Culture
No. of samples positive for *M. tuberculosis* by reference standard method (%)	34 (35)
All directly inoculated samples (n=98)	
Sensitivity, % (95% CI)	85 (69-95)
Specificity, % (95% CI)	93 (84-98)
Positive predictive value, % (95% CI)	88 (72-97)
Negative predictive value, % (95% CI)	92 (82-97)
HIV-positive (n=39)*	
Sensitivity, % (95% CI)	85 (55-98)
Specificity, % (95% CI)	92 (73-99)
Positive predictive value, % (95% CI)	85 (55-98)
Negative predictive value, % (95% CI)	92 (73-99)
HIV-negative (n=20)*	
Sensitivity, % (95% CI)	86 (42-100)
Specificity, % (95% CI)	100 (74-100)
Positive predictive value, % (95% CI)	100 (54-100)
Negative predictive value, % (95% CI)	93 (64-100)

*Among persons with known HIV status (*n*=59/98 (60%)).

Table 3. Drug-Susceptibility Test Results from the MODS Assay

	Isoniazid	Rifampin	Isoniazid + Rifampin (multidrug resistance)
No. of samples*	33	33	33
No. resistant (prevalence) †	18/27 (67%)	18/27 (67%)	17/27 (63%)
Sensitivity, % (95% CI)	88 (68-97)	96 (79-100)	91 (72-99)
Specificity, % (95% CI)	89 (52-100)	89 (52-100)	90 (56-100)
Positive predictive value, % (95% CI)	96 (77-100)	96 (79-100)	96 (77-100)
Negative predictive value, % (95% CI)	73 (39-94)	89 (52-100)	82 (48-98)

*Analysis limited to samples with positive microscopic-observation drug-susceptibility and reference standard culture.

† Among directly inoculated patient specimens.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Chapter 5. Evaluation of Multidrug Resistant Tuberculosis Suspects in Zimbabwe

Abstract

The extent of the MDR-TB epidemic in Sub-Saharan Africa is poorly characterized. Zimbabwe has among the highest TB incidence in the world, yet surveillance for drug resistant tuberculosis has not been undertaken since 1995. We determined that the prevalence of MDR-TB among retreatment TB cases in Zimbabwe is 19% (95% CI 13-26%), higher than the regional average. A large proportion of retreatment cases in this high HIV-burden setting have no microbiologic evidence of TB.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

The emergence of multidrug resistant (MDR) tuberculosis (TB) (TB caused by *Mycobacterium tuberculosis* resistant to at least isoniazid and rifampin) in sub-Saharan Africa has been associated with high mortality and ongoing transmission, [1] is likely underestimated, [2, 3] and poses a major risk of further destabilizing regional TB control programs. Despite these concerns, less than half of the 46 countries in the World Health Organization (WHO) African Region have provided representative drug resistance data, and only ten have reported data since 2000. [4]

Zimbabwe ranks 17th among the world's high TB burden countries, and 4th according to incidence per capita (782/100,000). [5] 15.6% of the adult population between 15-49 years is estimated to be HIV positive, and 80% of active TB cases occur in individuals co-infected with HIV. [6] Although Zimbabwe is not currently classified as a high MDR-TB burden country, the WHO annual estimated number of MDR-TB cases (970; 95% CI 406-1,980) [7] exceeds that for many of the countries in this classification. Current WHO prevalence estimates for MDR-TB among new (1.9%; 95% CI 1.0-3.3%) and retreatment (8.3%; 95% CI 3-22%) patients in Zimbabwe are based on a 1995 sub-national drug resistance survey, in which three of 36 patients tested were found to have MDR-TB. Formal national surveillance for drug resistant tuberculosis is a current leading priority for the Ministry of Health and Child Welfare (MOHCW).

It is likely that the prevalence of MDR-TB has increased in Zimbabwe. High and increasing rates of MDR-TB have been documented in neighboring southern Africa countries: South Africa, with the highest prevalence of MDR-TB in the continent and fourth highest in the world; Mozambique, one of only four African nations reporting an MDR-TB prevalence above 3% among new cases; and Botswana, where the incidence of MDR-TB appears to be rising faster than the incidence of drug-sensitive TB. [8] Zimbabweans living in these countries have been steadily returning home since the formation of the Unity Government in February 2009. Moreover, economic destabilization and population displacement have presented significant challenges to tuberculosis control, possibly increasing the prevalence of MDR-TB. [9] Yet, to date no recent data beyond anecdotal reports of MDR-TB cases among returnees from surrounding countries are available to quantify this burden. [10]

Materials and Methods. From January 1, 2012 to August 1, 2012 we prospectively recruited a population-based sample of consecutive individuals suspected of having drug-resistant pulmonary TB within Harare. Sputum smear-positive or smear-negative TB suspects with either (1) a history of prior treatment (> 1 month, classified according to World Health Organization criteria) [4] or (2) contact with an individual with known or suspected drug-resistant TB were considered drug-resistant TB suspects. Because of the historically centralized organization of TB treatment initiation in Harare (population 1.6 million), an accurate population-based sample could be obtained from a limited number of sites in the city; subjects were recruited from Harare's two central infectious diseases hospitals and a network of key polyclinics. Ethical approval was obtained from the Medical Research Council of Zimbabwe, the Institutional Review Board of the Biomedical Research and Training Institute, and the UCSF Human Research Protection Program. Cultures for mycobacteria were performed in an external quality assured laboratory using Löwenstein-Jensen (LJ) media and BBL™ MGIT™ Mycobacterial Growth Indicator Tubes (Becton Dickinson, Sparks, MD). [11] Indirect DST was performed on all *M, tuberculosis* isolates using absolute concentration measurement (MIC) on LJ media to determine susceptibility to isoniazid (0.2 and 1 ug/ml); rifampicin (32 and 64 ug/ml); ethambutol (2.8 and 4 ug/ml); and streptomycin (8 and 16 ug/ml). [12] Culture for mycobacteria and direct DST were also performed using the microscopic-observation drug-susceptibility (MODS) assay

in accordance with published standard operating procedures. [13] Drug resistance, including MDR, was classified according to results from either direct or indirect DST method. Methods published by the WHO were used to calculate the prevalence of drug-resistant tuberculosis. [4]

Results

Of 182 recruited patients, 25 (14%) were from outside Harare provincial limits and were not included in prevalence estimates. Of the remaining 157 patients (**Table 1**), 85% (n=133) were suspected retreatment cases and 15% (n=24) were new patients with a history of contact with a known or suspected MDR-TB case; of those with HIV test results available (148/157, 94%), 71% had HIV infection. Overall, 16% (n=25; 95% CI 11-23%) of individuals had organisms that were MDR, an additional 5% (n=8; 95% CI 2-10%) had rifampin or isoniazid-monoresistance, and 36% (n=56; 95% CI 28-44%) had drug-sensitive TB; 43% (n=68; 35-52%) had no evidence of TB on LJ media, manual MGIT, or MODS culture. The prevalence of MDR-TB among persons suspected of drug resistance due to a history of prior TB treatment (excluding new patients) was 19% (95% CI 13-26%). Overall, 32% (n=50) had a history of travel outside of Zimbabwe in the prior two years, and 2 of 23 (8.7%) MDR-TB patients with history of prior treatment reported TB treatment in South Africa in the prior year.

Discussion

In this study from the capital and largest city in Zimbabwe, we provide the first assessment of MDR-TB in the country since 1995. Since that time, the prevalence of MDR-TB has likely increased in Zimbabwe. Cases of MDR-TB have been contributed to by persons returning from other southern Africa countries, though the contribution to the overall prevalence of MDR-TB remains uncertain.

Despite a recent World Health Assembly resolution to achieve universal access to diagnosis and treatment of MDR-TB by 2015, the extent and course of the MDR-TB epidemic in the WHO African Region, outside of the Republic of South Africa, is poorly understood. Currently, approximately one in six persons with history of prior TB treatment will have MDR-TB in Zimbabwe, a proportion higher than the regional average.

Our study focused predominantly upon retreatment TB suspects, among whom MDR-TB cases represent (1) those with primary MDR organisms who failed their initial treatment, and (2) those who acquired MDR-TB during the course of treatment of a previous TB episode. We cannot comment on MDR-TB prevalence among new patients without risk factors for MDR-TB, a group that is thought to account for the majority of MDR-TB cases (69%) in the WHO African Region. [4] However, a recent clinical trial recruiting consecutive new patients without a history of prior TB treatment within Harare found only one case of MDR-TB among xxx recruited cases (Amina Jindani, PI RIFAQUIN, *personal communication*). Assuming the prevalence of MDR-TB among new cases has remained low, and assuming case detection was complete, in 2010, there were 1,273 (95% CI 864-1,714) incident cases of MDR-TB, somewhat more than current WHO estimates.

Valid drug resistance surveillance data must clearly distinguish new from retreatment cases, emanate from quality-assured laboratories, and be representative. That a population-based assessment of retreatment cases in Harare is generalizable to the rest of the country may be plausible because approximately one-quarter of all notified TB cases occur in Harare, though variation in the prevalence in rural areas or cities along Zimbabwe's border may exist. Lack of sampling from these areas is a limitation of our study. Although drug resistance surveys are typically performed using only sputum

smear-positive samples, we included both smear-negative and smear-positive TB suspects because (1) this group more accurately portrays the clinical scenario encountered by health care workers, and (2) it is ethically questionable to exclude smear-negative subjects in high HIV-burden regions, given the known poor sensitivity of smear microscopy.

A further unexpected finding of our study was the large proportion of individuals registered as retreatment TB cases who had no microbiologic evidence of TB (either drug-sensitive or drug-resistant), despite extensive testing. Chronic lung disease is common among HIV-infected individuals of all ages, and misclassification (and overtreatment) as TB is likely common in high HIV burden settings with limited diagnostic capability. This may have substantial implications for TB programs throughout sub-Saharan Africa.

A formal national drug resistance survey is needed to produce valid estimates of the incidence of MDR and XDR-TB among new and retreatment cases in Zimbabwe, although such activity must be considered in light of directing limited resources towards detection and management of persons with drug-resistant TB. Assistance from the WHO Green Light Committee initiative has been welcome in ensuring the availability of second-line treatment regimens, evidence-based medicine, and in-country experience. Future Zimbabwe National TB Program efforts should be informed by studies examining both traditional (e.g., previous treatment history, known case contact, institutional exposure, occupation) and nontraditional (e.g., recent migrant population) risk factors for drug resistant TB, as well as efforts to further characterize retreatment TB suspects who have no microbiologic evidence of TB. Prompt treatment and screening of contacts for patients with known MDR-TB in the context of socioeconomic stabilization and continued improvements in health service delivery will be essential to prevent further expansion of MDR-TB in Zimbabwe.

Table 1. Characteristics of Study Participants

	Drug-sensitive TB (n=56)	**MDR-TB (n=25)**	**Not TB (n=68)**
Age, median (IQR)	35 (28-43)	32 (24-40)	37 (31-45)
Male, n (%)	39 (70)	14 (56)	37 (54)
Previous TB Treatment, n (%)	48 (86)	23 (92)	56 (82)
Outpatients, n (%)	52 (93)	23 (92)	65 (96)
Sputum Smear-positive, n (%)	26 (46)	19 (76)	3 (4)
HIV-infection, n (%)	31 (57)	18 (78)	49 (77)
CD4+ T-lymphocytes, median (IQR)	194 (105-295)	146 (75-200)	257 (113-372)

Note: Eight patients with isoniazid or rifampin monoresistant TB were excluded from tabulation. Drug-sensitive TB is defined as documented sensitivity to isoniazid and rifampin; multidrug resistance is defined as resistance to at least isoniazid and rifampin; 'not TB' is defined as negative results on solid, liquid, and microscopic-observation drug-susceptibility (MODS) culture.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Concluding Remarks

There is hope on the horizon that we may witness the elimination of TB in our lifetime; the Millennium Development Goal target to slow and begin to reverse the TB epidemic by 2015 has already been achieved. The work presented here, entailing a range of analytic strategies from traditional epidemiology to loss-based prediction to repeated measures to meta-analysis, was formulated with this broader goal in mind, and has served to prepare me for a career focused on global TB elimination.

Yet, despite recent gains, substantial regional heterogeneity with respect to the success of TB control efforts continues to exist. The WHO African Region has one-quarter of the total world's TB cases, and the highest rates of cases and deaths per capita. In addition, almost 80% of TB cases among people living with HIV reside in Africa. The intersection of the MDR-TB and HIV epidemic in southern Africa continues to represent a major threat to regional TB control, signaling in essence a return to the pre-antibiotic era for many. Through this dissertation, I have come to concentrate substantial energy into coordinating research in Harare, Zimbabwe. My decision to work in this politically and economically tumultuous state unites both my personal and academic history. With the fourth highest TB incidence in the world and an 80% HIV co-infection rate, the needs of the population are palpable and efforts to address the problem are likely to have a large social benefit. Over the past year, I have enrolled over 250 persons suspected of drug-resistant TB in Zimbabwe in a prospective cohort, Trap MDR-TB (*Transmission and Pathogenesis of MDR-TB*), whose main goal is to minimize diagnostic delay on the pathway to appropriate MDR-TB treatment. In addition to qualitative work aimed at better understanding etiologies of patient, diagnostic, and referral delay, we offer comprehensive testing for MDR-TB using smear microscopy, solid culture, liquid culture (manual MGIT), microscopic-observation drug-susceptibility (MODS), and Xpert MTB/RIF testing for retreatment TB suspects; we also undertake household and congregate setting contact investigation, with symptomatic contacts offered free microbiologic testing and longitudinal follow-up. These studies aim (1) to validate or improve upon currently accepted international clinical care standards for patients at risk for drug resistant TB, and (2) to allow for determination of patients most likely to benefit from widely endorsed molecular drug susceptibility testing. We are also undertaking the first prospective study of the contribution of microbial and human genomic factors on drug resistant TB transmission in a high HIV prevalence setting; advanced genotyping methods to fully characterize serial drug resistant strains and novel statistical procedures to account for time-varying clinical covariates (e.g., treatment regimen, adherence, and CD4 count) will be used to characterize development of drug resistant TB in real-time.

Chapter 1 References

1. Menzies D, Pai M, Comstock G. Meta-analysis: New tests for the diagnosis of latent tuberculosis infection: Areas of uncertainty and recommendations for research. *Ann Intern Med* 2007;146:340-354.
2. Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: An update. *Ann Intern Med* 2008;149:177-184.
3. Ewer K, Deeks J, Alvarez L, Bryant G, Waller S, Andersen P, Monk P, Lalvani A. Comparison of t-cell-based assay with tuberculin skin test for diagnosis of mycobacterium tuberculosis infection in a school tuberculosis outbreak. *Lancet* 2003;361:1168-1173.
4. Hill PC, Fox A, Jeffries DJ, Jackson-Sillah D, Lugos MD, Owiafe PK, Donkor SA, Hammond AS, Corrah T, Adegbola RA, et al. Quantitative t cell assay reflects infectious load of mycobacterium tuberculosis in an endemic case contact model. *Clin Infect Dis* 2005;40:273-278.
5. Diel R, Nienhaus A, Lange C, Meywald-Walter K, Forssbohm M, Schaberg T. Tuberculosis contact investigation with a new, specific blood test in a low-incidence population containing a high proportion of bcg-vaccinated persons. *Respir Res* 2006;7:77.
6. Kang YA, Lee HW, Yoon HI, Cho B, Han SK, Shim YS, Yim JJ. Discrepancy between the tuberculin skin test and the whole-blood interferon gamma assay for the diagnosis of latent tuberculosis infection in an intermediate tuberculosis-burden country. *JAMA* 2005;293:2756-2761.
7. Diel R, Ernst M, Doscher G, Visuri-Karbe L, Greinert U, Niemann S, Nienhaus A, Lange C. Avoiding the effect of bcg vaccination in detecting mycobacterium tuberculosis infection with a blood test. *Eur Respir J* 2006;28:16-23.
8. Drobniewski F, Balabanova Y, Zakamova E, Nikolayevskyy V, Fedorin I. Rates of latent tuberculosis in health care staff in russia. *PLoS Med* 2007;4:e55.
9. CDC. Guidelines for using the quantiferon®-tb gold test for detecting mycobacterium tuberculosis infection, united states. December 16, 2005.
10. Dewan PK, Grinsdale J, Kawamura LM. Low sensitivity of a whole-blood interferon-gamma release assay for detection of active tuberculosis. *Clin Infect Dis* 2007;44:69-73.
11. Dosanjh DP, Hinks TS, Innes JA, Deeks JJ, Pasvol G, Hackforth S, Varia H, Millington KA, Gunatheesan R, Guyot-Revol V, et al. Improved diagnostic evaluation of suspected tuberculosis. *Ann Intern Med* 2008;148:325-336.
12. Kang YA, Lee HW, Hwang SS, Um SW, Han SK, Shim YS, Yim JJ. Usefulness of whole-blood interferon-gamma assay and interferon-gamma enzyme-linked immunospot assay in the diagnosis of active pulmonary tuberculosis. *Chest* 2007;132:959-965.
13. Kobashi Y, Mouri K, Yagi S, Obase Y, Fukuda M, Miyashita N, Oka M. Usefulness of the quantiferon tb-2g test for the differential diagnosis of pulmonary tuberculosis. *Intern Med* 2008;47:237-243.
14. Nishimura T, Hasegawa N, Mori M, Takebayashi T, Harada N, Higuchi K, Tasaka S, Ishizaka A. Accuracy of an interferon-gamma release assay to detect active pulmonary and extra-pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2008;12:269-274.
15. Ravn P, Munk ME, Andersen AB, Lundgren B, Lundgren JD, Nielsen LN, Kok-Jensen A, Andersen P, Weldingh K. Prospective evaluation of a whole-blood test using mycobacterium tuberculosis-specific antigens esat-6 and cfp-10 for diagnosis of active tuberculosis. *Clin Diagn Lab Immunol* 2005;12:491-496.
16. Winqvist N, Bjorkman P, Noren A, Miorner H. Use of a t cell interferon gamma release assay in the investigation for suspected active tuberculosis in a low prevalence area. *BMC Infect Dis* 2009;9:105.
17. Goletti D, Stefania C, Butera O, Amicosante M, Ernst M, Sauzullo I, Vullo V, Cirillo D, Borroni E, Markova R, et al. Accuracy of immunodiagnostic tests for active tuberculosis using single and combined results: A multicenter tbnet-study. *PLoS ONE* 2008;3:e3417.
18. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: From area under the roc curve to reclassification and beyond. *Stat Med* 2008;27:157-172; discussion 207-112.

19. Metcalfe JZ CA, Ho C, Grinsdale J, Kawamura M. Quantitative interferon-gamma release and classification of tuberculosis suspects with 'negative' quantiferon-gold results. *Am J Respir Crit Care Med* 2009;179:A5920.
20. Targeted tuberculin testing and treatment of latent tuberculosis infection. This official statement of the american thoracic society was adopted by the ats board of directors, july 1999. This is a joint statement of the american thoracic society (ats) and the centers for disease control and prevention (cdc). This statement was endorsed by the council of the infectious diseases society of america. (idsa), september 1999, and the sections of this statement. *Am J Respir Crit Care Med* 2000;161:S221-247.
21. Marais BJ, Gie RP, Schaaf HS, Beyers N, Donald PR, Starke JR. Childhood pulmonary tuberculosis: Old wisdom and new challenges. *Am J Respir Crit Care Med* 2006;173:1078-1090.
22. Marais BJ, Gie RP, Hesseling AH, Beyers N. Adult-type pulmonary tuberculosis in children 10-14 years of age. *Pediatr Infect Dis J* 2005;24:743-744.
23. Cellestis. Quantiferon®-tb gold package insert. .
24. Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. *Stat Appl Genet Mol Biol* 2004;3:Article18.
25. Parker CB, DeLong ER. Roc methodology within a monitoring framework. *Stat Med* 2003;22:3473-3488.
26. Greenland P, O'Malley PG. When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase a2 and c-reactive protein for stroke risk. *Arch Intern Med* 2005;165:2454-2456.
27. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837-1847.
28. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879-1886.
29. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928-935.
30. Prediction of mortality from coronary heart disease among diverse populations: Is there a common predictive function? *Heart* 2002;88:222-228.
31. Soysal A, Torun T, Efe S, Gencer H, Tahaoglu K, Bakir M. Evaluation of cut-off values of interferon-gamma-based assays in the diagnosis of m. Tuberculosis infection. *Int J Tuberc Lung Dis* 2008;12:50-56.
32. Lee JY, Choi HJ, Park IN, Hong SB, Oh YM, Lim CM, Lee SD, Koh Y, Kim WS, Kim DS, et al. Comparison of two commercial interferon-gamma assays for diagnosing mycobacterium tuberculosis infection. *Eur Respir J* 2006;28:24-30.
33. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat Med* 2006;25:127-141.
34. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med* 2008;149:751-760.

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Chapter 2 References

1. Pai M, Kalantri S and Dheda K. New tools and emerging technologies for the diagnosis of tuberculosis: part I. Latent tuberculosis. Expert Rev Mol Diagn 2006;6:413-22
2. Menzies D, Pai M and Comstock G. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. Ann Intern Med 2007;146:340-54
3. Pai M, Riley LW and Colford JM, Jr. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect Dis 2004;4:761-76
4. Pai M, Zwerling A and Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med 2008;149:177-84
5. Mazurek G, Jereb J, Vernon A, LoBue P and Goldberg S. Updated Guidelines for Using Interferon-gamma Release Assays to Detect Mycobacterium tuberculosis Infection, United States. MMWR *In press*
6. CTC. Canada Communicable Disease Report: Updated Recommendations on Interferon Gamma Release Assays for Latent Tuberculosis Infection. 2008
7. NHS. Health Protection Agency Position Statement on the use of Interferon Gamma Release Assay (IGRA) tests for tuberculosis (TB). HPA Tuberculosis Programme Board 2008
8. Lange C, Pai M, Drobniewski F and Migliori GB. Interferon-gamma release assays for the diagnosis of active tuberculosis: sensible or silly? Eur Respir J 2009;33:1250-3
9. Menzies D. Using tests for latent tuberculous infection to diagnose active tuberculosis: can we eat our cake and have it too? Ann Intern Med 2008;148:398-9
10. Pai M, Menzies D. Interferon-gamma release assays: what is their role in the diagnosis of active tuberculosis? Clin Infect Dis 2007;44:74-7
11. Sester M, Sotgiu G, Lange C, et al. Interferon-{gamma} release assays for the diagnosis of active tuberculosis: A systematic review and meta-analysis. Eur Respir J
12. Dheda K, van Zyl Smit R, Badri M and Pai M. T-cell interferon-gamma release assays for the rapid immunodiagnosis of tuberculosis: clinical utility in high-burden vs. low-burden settings. Curr Opin Pulm Med 2009;15:188-200
13. Pai M. Guidelines on IGRAs: Concordant or Discordant? 2nd Global Symposium on IGRAs. Dubrovnik, Croatia,, 2009
14. Diel R, Loddenkemper R and Nienhaus A. Evidence-based comparison of commercial interferon-gamma release assays for detecting active TB: a metaanalysis. Chest 2010;137:952-68
15. Jiang W, Shao L, Zhang Y, et al. High-sensitive and rapid detection of Mycobacterium tuberculosis infection by IFN-gamma release assay among HIV-infected individuals in BCG-vaccinated area. BMC Immunol 2009;10:31
16. Kanunfre KA, Leite OH, Lopes MI, Litvoc M and Ferreira AW. Enhancement of diagnostic efficiency by a gamma interferon release assay for pulmonary tuberculosis. Clin Vaccine Immunol 2008;15:1028-30
17. Soysal A, Torun T, Efe S, Gencer H, Tahaoglu K and Bakir M. Evaluation of cut-off values of interferon-gamma-based assays in the diagnosis of M. tuberculosis infection. Int J Tuberc Lung Dis 2008;12:50-6
18. Deville WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2002;2:9
19. Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. AJR Am J Roentgenol 2006;187:271-81
20. Leeflang MM, Deeks JJ, Gatsonis C and Bossuyt PM. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149:889-97
21. Pai M, McCulloch M, Enanoria W and Colford JM, Jr. Systematic reviews of diagnostic test evaluations: What's behind the scenes? ACP J Club 2004;141:A11-3
22. Dheda K, Smit RZ, Badri M and Pai M. T-cell interferon-gamma release assays for the rapid immunodiagnosis of tuberculosis: clinical utility in high-burden vs. low-burden settings. Curr Opin Pulm Med 2009;15:188-200

23. World Bank List of Economies.
http://siteresources.worldbank.org/DATASTATISTICS/Resources/CLASS.XLS April 2010;*accessed June 1, 2010*
24. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM and Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25
25. Fontela PS, Pant Pai N, Schiller I, Dendukuri N, Ramsay A and Pai M. Quality and reporting of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD standards. PLoS One 2009;4:e7753
26. Pai M, Minion J, Steingart K and Ramsay A. New and improved tuberculosis diagnostics: evidence, policy, practice, and impact. Curr Opin Pulm Med;16:271-84
27. Schunemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008;336:1106-10
28. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC and Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174:469-76
29. Van Deun A. What is the role of mycobacterial culture in diagnosis and case definition? In: Frieden T, ed. Toman's tuberculosis: case detection, treatment, and monitoring-questions and answers. World Health Organization 2004;2nd ed:35-43
30. Lijmer JG, Bossuyt PM and Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 2002;21:1525-37
31. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001;20:2865-84
32. Harbord R, Whiting P. Metandi: Meta-analysis of Diagnostic Accuracy using Hierarchical Logistic Regression. The Stata Journal 2009;9:211-229
33. Harbord RM, Deeks JJ, Egger M, Whiting P and Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2007;8:239-51
34. Aabye MG, Ravn P, PrayGod G, et al. The impact of HIV infection and CD4 cell count on the performance of an interferon gamma release assay in patients with pulmonary tuberculosis. PLoS One 2009;4:e4220
35. Cattamanchi A, Ssewenyana I, Davis JL, et al. Role of interferon-gamma release assays in the diagnosis of pulmonary tuberculosis in patients with advanced HIV infection. 2009
36. Chegou NN, Black GF, Kidd M, van Helden PD and Walzl G. Host markers in QuantiFERON supernatants differentiate active TB from latent TB infection: preliminary report. BMC Pulm Med 2009;9:21
37. Chen X, Yang Q, Zhang M, et al. Diagnosis of active tuberculosis in China using an in-house gamma interferon enzyme-linked immunospot assay. Clin Vaccine Immunol 2009;16:879-84
38. Dheda K, van Zyl-Smit RN, Meldau R, et al. Quantitative lung T cell responses aid the rapid diagnosis of pulmonary tuberculosis. Thorax 2009
39. Kabeer BSA, Sikhamani R, Swaminathan S, Perumal V, Paramasivam P and Raja A. Role of interferon gamma release assay in active TB diagnosis among HIV infected individuals. PLoS One 2009;4:e5718
40. Katiyar SK, Sampath A, Bihari S, Mamtani M and Kulkarni H. Use of the QuantiFERON-TB Gold In-Tube test to monitor treatment efficacy in active pulmonary tuberculosis. Int J Tuberc Lung Dis 2008;12:1146-52
41. Leidl L, Mayanja-Kizza H, Sotgiu G, et al. Relationship of immunodiagnostic assays for tuberculosis and numbers of circulating CD4+ T-cells in HIV-infection. Eur Respir J 2009
42. Markova R, Todorova Y, Drenska R, Elenkov I, Yankova M and Stefanova D. Usefulness of interferon-gamma release assays in the diagnosis of tuberculosis infection in HIV-infected patients in Bulgaria. Biotechnol. & Biotechnol 2009;23:1103-8
43. Oni T, Patel J, Gideon HP, et al. Enhanced diagnosis of HIV-1 associated tuberculosis by relating T-SPOT.TB and CD4 counts. Eur Respir J 2010
44. Ozekinci T, Ozbek E and Celik Y. Comparison of tuberculin skin test and a specific T-cell-based test, T-Spot.TB, for the diagnosis of latent tuberculosis infection. J Int Med Res 2007;35:696-703

45. Pai M, Joshi R, Bandyopadhyay M, et al. Sensitivity of a whole-blood interferon-gamma assay among patients with pulmonary tuberculosis and variations in T-cell responses during anti-tuberculosis treatment. Infection 2007;35:98-103
46. Raby E, Moyo M, Devendra A, et al. The effects of HIV on the sensitivity of a whole blood IFN-gamma release assay in Zambian adults with active tuberculosis. PLoS One 2008;3:e2489
47. Shao-ping. Enzyme-linked immunospot assay combined with serum latex agglutination test for diagnosis of pulmonary tuberculosis and concomitant pulmonary cryptococcosis. Chin J Infect Chemo 2009
48. Tahereh K, Alireza N, Massoud S and Amina K. A validity study of the QuantiFERON-TB Gold (QFT-TB) method for the diagnosis of pulmonary tuberculosis in a high risk population. Swiss Medical Weekly 2010;140:95-96
49. Tsiouris SJ, Coetzee D, Toro PL, Austin J, Stein Z and El-Sadr W. Sensitivity analysis and potential uses of a novel gamma interferon release assay for diagnosis of tuberculosis. J Clin Microbiol 2006;44:2844-50
50. Veldsman C, Kock MM, Rossouw T, et al. QuantiFERON-TB GOLD ELISA assay for the detection of Mycobacterium tuberculosis-specific antigens in blood specimens of HIV-positive patients in a high-burden country. FEMS Immunol Med Microbiol 2009
51. Oni T, Patel J, Gideon HP, et al. Enhanced diagnosis of HIV-1 associated tuberculosis by relating T-SPOT.TB and CD4 counts. Eur Respir J
52. WHO. Global Tuberculosis Control 2009. Geneva: World Health Organization 2009
53. Leeflang MM, Bossuyt PM and Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol 2009;62:5-12
54. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med 1997;16:981-91
55. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-6
56. Andersen P, Doherty TM, Pai M and Weldingh K. The prognosis of latent tuberculosis: can disease be predicted? Trends Mol Med 2007;13:175-82
57. Barry CE, 3rd, Boshoff HI, Dartois V, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol 2009;7:845-55
58. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928-35
59. Metcalfe JZ, Cattamanchi A, Vittinghoff E, et al. Evaluation of quantitative IFN-gamma response for risk stratification of active tuberculosis suspects. Am J Respir Crit Care Med 2010;181:87-93
60. Ling D, Pai M, Davids V, et al. Incremental Value Of Interferon-Gamma Release Assays For Diagnosis Of Active Tuberculosis In Smear-Negative Patients In A High-Burden Setting: A Multivariable Analysis. Abstract: Am. J. Respir. Crit. Care Med. May 2010;181:A2262
61. WHO. Strategic and Technical Advisory Group for Tuberculosis (STAG-TB): Report of the Tenth Meeting, 2010. (Accessed at http://www.who.int/tb/advisory_bodies/stag_tb_report_2010.pdf)

John Metcalfe
Division of Epidemiology, University of California, Berkeley

Chapter 3 References

1. Mazurek GH, Jereb J, Vernon A, LoBue P, Goldberg S, Castro K. Updated guidelines for using interferon gamma release assays to detect mycobacterium tuberculosis infection - united states, 2010. *MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports / Centers for Disease Control* 2010;59:1-25.
2. Pai M, O'Brien R. Serial testing for tuberculosis: Can we make sense of t cell assay conversions and reversions? *PLoS Med* 2007;4:e208.
3. Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, Poole F, Smith PG, Sriram N, Wongsrichanalai C, Linke R, O'Brien R, Perkins M, Cunningham J, Matsoso P, Nathanson CM, Olliaro P, Peeling RW, Ramsay A. Evaluation of diagnostic tests for infectious diseases: General principles. *Nature reviews Microbiology* 2006;4:S20-32.
4. Derion T. Considerations for the planning and conduct of reproducibility studies of in vitro diagnostic tests for infectious agents. *Biotechnology annual review* 2003;9:249-258.
5. American society for testing and materials (astm) international: What are repeatability and reproducibility? March/April 2009. Accessed July 25, 2012 at http://www.Astm.Org/SNEWS/MA_2009/datapoints_ma09.html
6. American society for testing and materials. E456 standard terminology relating to quality and statistics, 2012. Accessed July 25, 2012 at: http://www.Astm.Org/Standards/E456.htm
7. Cellestis. Validation report quantiferon®-tb gold in-tube: Reproducibility study. 2006 Dec. 16, 2011]. Available from: http://www.cellestis.com/IRM/Company/ShowPage.aspx/PDFs/1359-10000000/ValidationReportQFTInTubeReproducibilityStudy.
8. Detjen AK, Loebenberg L, Grewal HM, Stanley K, Gutschmidt A, Kruger C, Du Plessis N, Kidd M, Beyers N, Walzl G, Hesseling AC. Short-term reproducibility of a commercial interferon gamma release assay. *Clinical and vaccine immunology : CVI* 2009;16:1170-1175.
9. Pai M, Joshi R, Dogra S, Mendiratta DK, Narang P, Kalantri S, Reingold AL, Colford JM, Jr., Riley LW, Menzies D. Serial testing of health care workers for tuberculosis using interferon-gamma assay. *American journal of respiratory and critical care medicine* 2006;174:349-355.
10. Veerapathran A, Joshi R, Goswami K, Dogra S, Moodie EE, Reddy MV, Kalantri S, Schwartzman K, Behr MA, Menzies D, Pai M. T-cell assays for tuberculosis infection: Deriving cut-offs for conversions using reproducibility data. *PloS one* 2008;3:e1850.
11. van Zyl-Smit RN, Pai M, Peprah K, Meldau R, Kieck J, Juritz J, Badri M, Zumla A, Sechi LA, Bateman ED, Dheda K. Within-subject variability and boosting of t-cell interferon-gamma responses after tuberculin skin testing. *American journal of respiratory and critical care medicine* 2009;180:49-58.
12. Herrmann JL, Belloy M, Porcher R, Simonney N, Aboutaam R, Lebourgeois M, Gaudelus J, De Losangeles L, Chadelat K, Scheinmann P, Beydon N, Fauroux B, Bingen M, Terki M, Barraud D, Cruaud P, Offredo C, Ferroni A, Berche P, Moissenet D, Vuthien H, Doit C, Bingen E, Lagrange PH. Temporal dynamics of interferon gamma responses in children evaluated for tuberculosis. *PloS one* 2009;4:e4130.
13. Chee CB, Lim LK, Barkham TM, Koh DR, Lam SO, Shen L, Wang YT. Use of a t cell interferon-gamma release assay to evaluate tuberculosis risk in newly qualified physicians in singapore healthcare institutions. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 2009;30:870-875.
14. Lee K, Han MK, Choi HR, Choi CM, Oh YM, Lee SD, Kim WS, Kim DS, Woo JH, Shim TS. Annual incidence of latent tuberculosis infection among newly employed nurses at a tertiary care university hospital. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 2009;30:1218-1222.
15. Perry S, Sanchez L, Yang S, Agarwal Z, Hurst P, Parsonnet J. Reproducibility of quantiferon-tb gold in-tube assay. *Clinical and vaccine immunology : CVI* 2008;15:425-432.
16. Pollock NR, Campos-Neto A, Kashino S, Napolitano D, Behar SM, Shin D, Sloutsky A, Joshi S, Guillet J, Wong M, Nardell E. Discordant quantiferon-tb gold test results among us healthcare workers with increased

risk of latent tuberculosis infection: A problem or solution? *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 2008;29:878-886.
17. Yoshiyama T, Harada N, Higuchi K, Nakajima Y, Ogata H. Estimation of incidence of tuberculosis infection in health-care workers using repeated interferon-gamma assays. *Epidemiology and infection* 2009;137:1691-1698.
18. van Zyl-Smit RN, Zwerling A, Dheda K, Pai M. Within-subject variability of interferon-g assay results for tuberculosis and boosting effect of tuberculin skin testing: A systematic review. *PloS one* 2009;4:e8517.
19. Tuuminen T, Tavast E, Vaisanen R, Himberg JJ, Seppala I. Assessment of imprecision in gamma interferon release assays for the detection of exposure to mycobacterium tuberculosis. *Clinical and vaccine immunology : CVI* 2010;17:596-601.
20. Ringshausen FC, Nienhaus A, Torres Costa J, Knoop H, Schlosser S, Schultze-Werninghaus G, Rohde G. Within-subject variability of mycobacterium tuberculosis-specific gamma interferon responses in german health care workers. *Clinical and vaccine immunology : CVI* 2011;18:1176-1182.
21. Janetzki S, Schaed S, Blachere NE, Ben-Porat L, Houghton AN, Panageas KS. Evaluation of elispot assays: Influence of method and operator on variability of results. *Journal of immunological methods* 2004;291:175-183.
22. Ringshausen FC, Nienhaus A, Schablon A, Schlosser S, Schultze-Werninghaus G, Rohde G. Predictors of persistently positive mycobacterium-tuberculosis-specific interferon-gamma responses in the serial testing of health care workers. *BMC infectious diseases* 2010;10:220.
23. Metcalfe JZ, Parker M, Lew J, Graviss EA. Within-subject variability of the quantiferon-tb gold in-tube assay in a large cohort of low risk subjects. *3rd Global Symposium on IGRAs* 2012.
24. Ep21-a, estimation of total analytical error for clinical laboratory methods; approved guideline, clinical and laboratory standards institute (clsi21-a), wayne, pa, 2003.
25. Porsa E, Cheng L, Seale MM, Delclos GL, Ma X, Reich R, Musser JM, Graviss EA. Comparison of a new esat-6/cfp-10 peptide-based gamma interferon assay and a tuberculin skin test for tuberculosis screening in a moderate-risk population. *Clinical and vaccine immunology : CVI* 2006;13:53-58.
26. Porsa E, Cheng L, Graviss EA. Comparison of an esat-6/cfp-10 peptide-based enzyme-linked immunospot assay to a tuberculin skin test for screening of a population at moderate risk of contracting tuberculosis. *Clinical and vaccine immunology : CVI* 2007;14:714-719.
27. Grimes CZ, Hwang LY, Williams ML, Austin CM, Graviss EA. Tuberculosis infection in drug users: Interferon-gamma release assay performance. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2007;11:1183-1189.
28. Cruz AT, Geltemeyer AM, Starke JR, Flores JA, Graviss EA, Smith KC. Comparing the tuberculin skin test and t-spot.Tb blood test in children. *Pediatrics* 2011;127:e31-38.
29. Cellestis limited (2010). Quantiferon-tb gold in-tube: Package insert. Carnegie, victoria, australia. Available: Http://www.Cellestis.Com/irm/content/pdf/quantiferon%20us%20verg_jan2010%20no%20trims.Pdf. Accessed 2012 may 4.
30. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New york: Wiley.; 1981.
31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-310.
32. Powell RD, 3rd, Whitworth WC, Bernardo J, Moonan PK, Mazurek GH. Unusual interferon gamma measurements with quantiferon-tb gold and quantiferon-tb gold in-tube tests. *PloS one* 2011;6:e20061.
33. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963-974.
34. Diggle PJ. An approach to the analysis of repeated measurements. *Biometrics* 1988;44:959-971.
35. Furcolow ML, Watson KA, Charron T, Lowe J. A comparison of the tine and mono-vacc tests with the intradermal tuberculin test. *The American review of respiratory disease* 1967;96:1009-1027.
36. Erdtmann FJ, Dixon KE, Llewellyn CH. Skin testing for tuberculosis. Antigen and observer variability. *JAMA : the journal of the American Medical Association* 1974;228:479-481.
37. Bearman JE, Kleinman H, Glyer VV, Lacroix OM. A study of variability in tuberculin test reading. *The American review of respiratory disease* 1964;90:913-919.

38. Chaparas SD, Vandiviere HM, Melvin I, Koch G, Becker C. Tuberculin test. Variability with the mantoux procedure. *The American review of respiratory disease* 1985;132:175-177.
39. Targeted tuberculin testing and treatment of latent tuberculosis infection. American thoracic society. *MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports / Centers for Disease Control* 2000;49:1-51.
40. Zwerling A, van den Hof S, Scholten J, Cobelens F, Menzies D, Pai M. Interferon-gamma release assays for tuberculosis screening of healthcare workers: A systematic review. *Thorax* 2011.
41. Zwerling A, Cloutier-Ladurantaye J, Pietrangelo F. Conversions and reversions in health care workers in montreal, canada using quantiferon-tb-gold in-tube. *American journal of respiratory and critical care medicine* 2009:A1012.
42. Belknap R WK, Teeter L. Interferon-gamma release assays (igras) in serial testing for latent tuberculosis infection in u.S. Health care workers. *American journal of respiratory and critical care medicine* 2010:A2263.
43. Mancuso JD, Mazurek GH, Tribble D, Olsen C, Aronson NE, Geiter L, Goodwin D, Keep LW. Discordance among commercially available diagnostics for latent tuberculosis infection. *American journal of respiratory and critical care medicine* 2012;185:427-434.
44. Fong KS, Tomford JW, Teixeira L, Fraser TG, Vanduin D, Yen-Lieberman B, Gordon SM, Miranda C. Challenges of interferon-gamma release assay conversions in serial testing of health care workers in a tuberculosis control program. *Chest* 2012.
45. Gray J, Reves R, Johnson S, Belknap R. Identification of false-positive quantiferon-tb gold in-tube assays by repeat testing in hiv-infected patients at low risk for tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012;54:e20-23.
46. Menzies D. Interpretation of repeated tuberculin tests. Boosting, conversion, and reversion. *American journal of respiratory and critical care medicine* 1999;159:15-21.
47. Schablon A, Harling M, Diel R, Ringshausen FC, Torres Costa J, Nienhaus A. Serial testing with an interferon-gamma release assay in german healthcare workers. *GMS Krankenhaushygiene interdisziplinar* 2010;5.
48. Quan H, Shih WJ. Assessing reproducibility by the within-subject coefficient of variation with random effects models. *Biometrics* 1996;52:1195-1203.
49. Krouwer JS, Rabinowitz R. How to improve estimates of imprecision. *Clinical chemistry* 1984;30:290-292.
50. van Zyl-Smit RN, Binder A, Meldau R, Mishra H, Semple PL, Theron G, Peter J, Whitelaw A, Sharma SK, Warren R, Bateman ED, Dheda K. Comparison of quantitative techniques including xpert mtb/rif to evaluate mycobacterial burden. *PloS one* 2011;6:e28815.
51. Metcalfe JZ, Cattamanchi A, Vittinghoff E, Ho C, Grinsdale J, Hopewell PC, Kawamura LM, Nahid P. Evaluation of quantitative ifn-gamma response for risk stratification of active tuberculosis suspects. *American journal of respiratory and critical care medicine* 2010;181:87-93.
52. Diel R, Loddenkemper R, Niemann S, Meywald-Walter K, Nienhaus A. Negative and positive predictive value of a whole-blood interferon-{gamma} release assay for developing active tuberculosis: An update. *American journal of respiratory and critical care medicine* 2011;183:88-95.

Chapter 4 References

1. WHO (2010) Multidrug and extensively drug-resistant tuberculosis: 2010 global report on surveillance and response. Geneva, Switzerland. World Health Organization.
2. World_Health_Organization (2009) Global tuberculosis control - epidemiology, strategy, financing.
3. Gandhi NR, Moll AP, Lalloo U, Pawinski R, Zeller K, et al. (2009) Successful integration of tuberculosis and HIV treatment in rural South Africa: the Sizonq'oba study. J Acquir Immune Defic Syndr 50: 37-43.
4. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, et al. (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet 368: 1575-1580.
5. Ben Amor Y, Nemser B, Singh A, Sankin A, Schluger N (2008) Underreported threat of multidrug-resistant tuberculosis in Africa. Emerg Infect Dis 14: 1345-1352.
6. Wright A, Zignol M, Van Deun A, Falzon D, Gerdes SR, et al. (2009) Epidemiology of antituberculosis drug resistance 2002-07: an updated analysis of the Global Project on Anti-Tuberculosis Drug Resistance Surveillance. Lancet 373: 1861-1873.
7. IOM. Addressing the Threat of Drug-Resistant Tuberculosis: A Realistic Assessment of the Challenge. Workshop Summary. ; 2009; Washington, DC. The National Academies Press.
8. Minion J, Leung E, Menzies D, Pai M (2010) Microscopic-observation drug susceptibility and thin layer agar assays for the detection of drug resistant tuberculosis: a systematic review and meta-analysis. Lancet Infect Dis 10: 688-698.
9. WHO. Non-commercial culture and drug-susceptibility testing methods for screening of patients at risk of multi-drug resistant tuberculosis: policy statement. http://whqlibdoc.who.int/publications/2011/9789241501620_eng.pdf (accessed June 25, 2012).
10. Minion J, Pai M (2010) Expanding the role of the microscopic observation drug susceptibility assay in tuberculosis and HIV management. Clin Infect Dis 50: 997-999.
11. Shah NS, Moodley P, Babaria P, Moodley S, Ramtahal M, et al. (2011) Rapid diagnosis of tuberculosis and multidrug resistance by the microscopic-observation drug-susceptibility assay. Am J Respir Crit Care Med 183: 1427-1433.
12. Reddy KP, Brady MF, Gilman RH, Coronel J, Navincopa M, et al. (2010) Microscopic observation drug susceptibility assay for tuberculosis screening before isoniazid preventive therapy in HIV-infected persons. Clin Infect Dis 50: 988-996.
13. World Health Organization. Laboratory services in tuberculosis control. Part II. Microscopy. Geneva, Switzerland: WHO, 1998. .
14. Becton Dickinson Manual Mycobacterial Growth Systems. Accessed on July 29, 2012 at: http://www.bd.com/ds/productCenter/MT-Manual.asp.
15. SD Standard Diagnostics. Accessed on July 29, 2012 at: http://www.alere.es/UserFiles/TBAgMPT64.pdf.
16. Canetti G, Froman S, Grosset J, Hauduroy P, Langerova M, et al. (1963) Mycobacteria: Laboratory Methods for Testing Drug Sensitivity and Resistance. Bull World Health Organ 29: 565-578.
17. Hardy Diagnostics TB MODS Kit Package Insert. Accessed on August 13, 2012 at http://tbevidence.org/documents/rescentre/sop/MODS_Flyer.pdf.
18. Wells CD, Cegielski JP, Nelson LJ, Laserson KF, Holtz TH, et al. (2007) HIV infection and multidrug-resistant tuberculosis: the perfect storm. J Infect Dis 196 Suppl 1: S86-107.
19. Hesseling AC, Kim S, Madhi S, Nachman S, Schaaf HS, et al. (2012) High prevalence of drug resistance amongst HIV-exposed and -infected children in a tuberculosis prevention trial. Int J Tuberc Lung Dis 16: 192-195.
20. Easterbrook PJ, Gibson A, Murad S, Lamprecht D, Ives N, et al. (2004) High rates of clustering of strains causing tuberculosis in Harare, Zimbabwe: a molecular epidemiological study. J Clin Microbiol 42: 4536-4544.
21. Halperin DT, Mugurungi O, Hallett TB, Muchini B, Campbell B, et al. (2011) A surprising prevention success: why did the HIV epidemic decline in Zimbabwe? PLoS Med 8: e1000414.
22. Bateman C (2008) Zimbabwe meltdown fuelling MDRTB? S Afr Med J 98: 15-16.

23. Abdool Karim SS, Naidoo K, Grobler A, Padayatchi N, Baxter C, et al. (2010) Timing of initiation of antiretroviral drugs during tuberculosis therapy. N Engl J Med 362: 697-706.
24. Vella V, Racalbuto V, Guerra R, Marra C, Moll A, et al. (2011) Household contact investigation of multidrug-resistant and extensively drug-resistant tuberculosis in a high HIV prevalence setting. Int J Tuberc Lung Dis 15: 1170-1175, i.
25. Moore DA, Shah NS (2011) Alternative methods of diagnosing drug resistance--what can they do for me? J Infect Dis 204 Suppl 4: S1110-1119.
26. Leung E, Minion J, Benedetti A, Pai M, Menzies D (2012) Microcolony culture techniques for tuberculosis diagnosis: a systematic review. Int J Tuberc Lung Dis 16: 16-23, i-iii.
27. Chihota VN, Grant AD, Fielding K, Ndibongo B, van Zyl A, et al. (2010) Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting. Int J Tuberc Lung Dis 14: 1024-1031.
28. Gumbo T (2010) New susceptibility breakpoints for first-line antituberculosis drugs based on antimicrobial pharmacokinetic/pharmacodynamic science and population pharmacokinetic variability. Antimicrob Agents Chemother 54: 1484-1491.

Chapter 5 References

1. Gandhi NR, Moll A, Sturm AW, et al. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet **2006** Nov 4;368(9547):1575-80.
2. Ben Amor Y, Nemser B, Singh A, Sankin A, Schluger N. Underreported threat of multidrug-resistant tuberculosis in Africa. Emerging infectious diseases **2008** Sep;14(9):1345-52.
3. Shah NS, Richardson J, Moodley P, et al. Increasing drug resistance in extensively drug-resistant tuberculosis, South Africa. Emerging infectious diseases **2011** Mar;17(3):510-3.
4. World Health Organization. Multidrug and extensively drug-resistant tuberculosis: 2010 global report on surveillance and response. Geneva, Switzerland.
5. World Health Organization. Global tuberculosis control: a short update to the 2009 report. Geneva, Switzerland.
6. USAID. Report on the global AIDS epidemic. **2008**.
7. World Health Organization. Zimbabwe Tuberculosis Profile, 2011. Accessed on October 13, 2012 at https://extranet.who.int/sree/Reports?op=Replet&name=/WHO_HQ_Reports/G2/PROD/EXT/TBCountryProfile&ISO2=ZW&outtype=PDF.
8. Dye C. Doomsday postponed? Preventing and reversing epidemics of drug-resistant tuberculosis. Nature reviews Microbiology **2009** Jan;7(1):81-7.
9. Bateman C. Zimbabwe meltdown fuelling MDRTB? S Afr Med J **2008** Jan;98(1):15-6.
10. IOM Regional Office for Southern Africa. IOM detects possible TB problem amongst returning migrants at Plumtree Centre - Zimbabwe. A Bulletin of News, Information, and Analysis on Migration Health in Southern Africa. Issue Two. Nov 2009. Accessed on Oct 13, 2012 at: http://iomzimbabwe.org/index.php?option=com_content&view=article&id=51&catid=1.
11. Becton Dickinson Manual Mycobacterial Growth Systems. Accessed on July 29, 2012 at: http://www.bd.com/ds/productCenter/MT-Manual.asp.
12. Canetti G, Froman S, Grosset J, et al. Mycobacteria: Laboratory Methods for Testing Drug Sensitivity and Resistance. Bull World Health Organ **1963**;29:565-78.
13. Hardy Diagnostics TB MODS Kit Package Insert. Accessed on August 13, 2012 at http://tbevidence.org/documents/rescentre/sop/MODS_Flyer.pdf.