



# Interferon release does not add discriminatory value to smear-negative HIV–tuberculosis algorithms

M.X. Rangaka<sup>\*,#,†</sup>, H.P. Gideon<sup>#</sup>, K.A. Wilkinson<sup>#,+,§</sup>, M. Pai<sup>‡</sup>,  
J. Mwansa-Kambafwile<sup>#</sup>, G. Maartens<sup>§,\*\*,††</sup>, J.R. Glynn<sup>†</sup>, A. Boulle<sup>\*</sup>, K. Fielding<sup>†</sup>,  
R. Goliath<sup>#</sup>, R. Titus<sup>#</sup>, S. Mathee<sup>##</sup> and R.J. Wilkinson<sup>#,+,§,††</sup>

**ABSTRACT:** Clinical algorithms for evaluating HIV-infected individuals for tuberculosis (TB) prior to isoniazid preventive therapy (IPT) perform poorly, and interferon- $\gamma$  release assays (IGRAs) have moderate accuracy for active TB. It is unclear whether, when used as adjunct tests, IGRAs add any clinical discriminatory value for active TB diagnosis in the pre-IPT assessment.

779 sputum smear-negative HIV-infected persons, established on or about to commence combined antiretroviral therapy (ART), were screened for TB prior to IPT. Stepwise multivariable logistic regression was used to develop clinical prediction models. The discriminatory ability was assessed by receiver operator characteristic area under the curve (AUC). QuantiFERON<sup>®</sup>-TB Gold in-tube (QFT-GIT) was evaluated.

The prevalence of smear-negative TB by culture was 6.4% (95% CI 4.9–8.4%). Used alone, QFT-GIT and the tuberculin skin test (TST) had comparable performance; the post-test probability of disease based on single negative tests was 3–4%. In a multivariable model, the QFT-GIT test did not improve the ability of a clinical algorithm, which included not taking ART, weight <60 kg, no prior history of TB, any one positive TB symptom/sign (cough  $\geq$  2 weeks) and CD4+ count <250 cells per mm<sup>3</sup>, to discriminate smear-negative culture-positive and -negative TB (72% to 74%; AUC comparison  $p=0.33$ ). The TST marginally improved the discriminatory ability of the clinical model (to 77%, AUC comparison  $p=0.04$ ).

QFT-GIT does not improve the discriminatory ability of current TB screening clinical algorithms used to evaluate HIV-infected individuals for TB ahead of preventive therapy. Evaluation of new TB diagnostics for clinical relevance should follow a multivariable process that goes beyond test accuracy.

**KEYWORDS:** Diagnostic research, HIV, incremental value, interferon- $\gamma$  release assay, *Mycobacterium tuberculosis*, QuantiFERON<sup>®</sup>-TB Gold in-tube

The HIV-1 epidemic has derailed tuberculosis (TB) control in sub-Saharan Africa. It is recognised that additional interventions, apart from the otherwise successful directly observed treatment short course (DOTS) strategy, are required to control TB [1]. The World Health Organization (WHO) therefore promotes isoniazid preventive therapy (IPT), intensified TB case finding and TB infection control to curb TB in HIV-infected individuals as part of the “Three Is” initiative [2]. Active TB should be excluded prior to starting IPT to avoid drug resistance from isoniazid monotherapy [3]. Where sputum culture for *Mycobacterium tuberculosis* has been

employed as a gold standard method in various settings, the median prevalence of positivity has ranged from 0.7% in population-based surveys to 8.5% in HIV-1 voluntary counselling and testing services [4]. Asymptomatic and smear-negative forms of disease are also common and so the operational need to screen for prevalent TB has become very important [5, 6].

The gold standard for intensified TB case finding remains sputum culture. It is more sensitive than smear, but takes several weeks and is expensive. Thus, screening algorithms based on TB symptoms have been suggested but sensitivity and specificity

## AFFILIATIONS

\*Centre for Infectious Disease and Epidemiology Research (CIDER),

#Clinical Infectious Diseases Research Initiative, Institute of Infectious Diseases and Molecular Medicine,

§Dept of Medicine,

\*\*Division of Clinical Pharmacology, University of Cape Town, and

##Provincial Administration of the Western Cape, Cape Town, South Africa.

†London School of Hygiene and Tropical Medicine,

††MRC National Institute for Medical Research, and

‡Division of Medicine, Imperial College London, London, UK.

‡McGill University and Montreal Chest Institute, Montreal, QC, Canada.

## CORRESPONDENCE

M.X. Rangaka

Centre for Infectious Disease and Epidemiology Research (CIDER)

School of Health Sciences

University of Cape Town

Observatory

7925

Cape Town

South Africa

E-mail: mxrangaka@yahoo.co.uk

Received:

April 05 2011

Accepted after revision:

June 04 2011

First published online:

June 30 2011

European Respiratory Journal

Print ISSN 0903-1936

Online ISSN 1399-3003

This article has supplementary material available from [www.erj.ersjournals.com](http://www.erj.ersjournals.com)

vary between strategies, studies and location [5, 7–11]. A recently proposed algorithm to screen HIV-infected persons living in resource-poor areas for TB reported that any one symptom/sign out of cough of any duration, fever, night sweats or weight loss were 79% sensitive with a negative predictive value of 98% (in a 5% TB prevalence setting), but only 50% specific. The authors concluded that reliably ruling out active TB might have priority over generating a consistent group of subjects with positive screening and no TB. However, with such poor specificity, an implicit consequence of the proposed algorithm is that the majority of those with false positive results would still require further investigation to rule out active TB disease [12]. This highlights the need for alternative or additional techniques.

Infection by *M. tuberculosis* can be inferred from evidence of immune sensitisation either *in vivo* by the Mantoux tuberculin skin test (TST) or *in vitro* by antigen-specific release of interferon (IFN)- $\gamma$  by T-cells (IFN- $\gamma$  release assay; IGRA). These present additional TB screening strategies. However, the utility of IGRAs in high-TB incidence settings where HIV-1 co-infection is also common remains unclear because, like the TST, IGRAs cannot distinguish between latent infection and active disease. Used as single stand-alone tests, both IGRA and the TST are imperfect for diagnosing those who have prevalent active disease as they have moderate positive predictive value and can be falsely negative [13–17]. Similarly, a single positive result may not necessarily mean the presence of active disease. Based on existing data reporting poor performance of current commercial IGRAs in low- and middle-income countries, a recent WHO Strategic and Technical Advisory Group meeting on the use of commercial IGRAs for the diagnosis of latent and active TB planned to support an approach to develop a “negative” WHO policy recommendation to discourage the use of commercial IGRAs in those countries [18]. However, the utility of IGRAs may be better appreciated in multivariable clinical algorithms. It is thus important to understand the incremental or added value of IGRAs in active TB diagnosis, over and beyond conventional tests for active TB [19, 20]. Considering IGRAs with other predictors might modify both its accuracy estimates of sensitivity, specificity and likelihood ratios, and the overall posterior probability of active TB disease [21, 22]. This is the crux of a multivariable evaluation. Such an evaluation of IGRAs has not previously been explored and in the context of the pre-IPT evaluation. The WHO has identified these as priority research questions in the IGRA and TB prevention areas [23].

Our hypothesis was that when used as adjunct tests, even poorly to moderately performing TB biomarkers will improve the ability of current multivariable clinical algorithms to discriminate those with and without TB. Therefore, we assessed the incremental or additive discriminatory value of QuantiFERON<sup>®</sup>-TB Gold in-tube (QFT-GIT; Cellestis, Carnegie, Australia) or Mantoux TST when added as adjunct tools to detect active TB among sputum smear-negative persons who were being prescribed or about to commence antiretroviral therapy (ART) and being assessed for IPT.

## MATERIAL AND METHODS

### Participant recruitment and study setting

The University of Cape Town (Cape Town, South Africa) research ethics committee approved this study (REC 013/2007) and written informed consent was provided. The study setting was the

Ubuntu Clinic in Khayelitsha Site B, an HIV-TB outpatient facility 30 km east of Cape Town, South Africa. Overall TB incidence during the study was ~1,500 per 100,000 inhabitants. Consecutive participants undergoing screening for an ongoing pragmatic randomised controlled trial (RCT) were invited. The trial was established to determine the effectiveness of the combination of IPT and ART provided under routine clinical practice to reduce the risk of active TB in HIV-infected persons (clinicaltrials.gov identifier NCT00463086 and Lancet D-09-02885 approved protocol). Participants were eligible if they were willing to provide additional consent for further screening by TST and QFT-GIT. They were screened regardless of initial TB symptoms or signs and regardless of eligibility for the ART-IPT study. The selection criteria for the RCT were not applied when assessing patients' eligibility for evaluation by QFT-GIT. Screening occurred between November 2007 and September 2009. Adults aged  $\geq 18$  yrs, who were already prescribed ART or about to commence ART and whose TST response could be assessed by the study team in 48–72 h, were eligible to participate. There were no other specific exclusion criteria for the QFT-GIT substudy. To meet our study objectives, all analyses were undertaken on a restricted dataset of sputum smear-negative participants. This evaluation of standard QFT-GIT occurred under routine clinical practice and was not intended as an optimisation study for QFT-GIT.

### Study procedures

All TB screening tools were administered on the day of screening. Participants were evaluated for the following TB symptoms or signs at the baseline screening visit: cough  $\geq 2$  weeks; night sweats in the last 2 weeks; self-reported weight loss for those newly prescribed ART or weight loss  $\geq 1.5$  kg in 1 month for patients already taking ART at screening; fever  $\geq 2$  weeks; and significant lymphadenopathy ( $>2$  cm in diameter) on examination. The algorithm reflects South African guidelines including the TB Programme and the National IPT guidelines for HIV-infected persons. Chest radiography (CXR) did not form part of the screening strategy.

A single sputum sample for TB microscopy and culture was obtained. Patients who failed to produce sputum either spontaneously or following ultrasonic nebulisation with hypertonic saline were either asked to return the following day (with a sample) or nebulisation was repeated if necessary at their next clinic visit. A window period of up to 1 month for sputum collection was permitted. Specimens were sent to an academic reference laboratory at Groote Schuur Hospital (Cape Town) for processing. The laboratory is part of the South African National Health Laboratory Service. A single positive culture (BACTEC mycobacterial growth indicator (Becton Dickinson Microbiology Systems, Cockeysville, MD, USA) and Lowenstein-Jensen media) for *M. tuberculosis* was regarded as definite active disease and formed the reference standard criterion for diagnostic accuracy in this study. Given the high rate of active TB in the recruited population, all cases with positive cultures were considered as definite TB and thus referred for prompt DOTS. Upon referral to the TB programme, further assessments initiated included physical examination and/or CXR. During this study period, the reference laboratory reported low cross-contamination; of 500 dummy sputum samples, none were false positive for *M. tuberculosis* (M. Nicol, Division of Medical Microbiology, Faculty of Health Sciences, University of Cape Town and

National Health Laboratory Service, Cape Town, South Africa; personal communication).

The TST (2 TU RT23 purified protein derivative (PPD); Statens Serum Institut, Copenhagen, Denmark) was administered on the volar aspect of the left forearm by personnel trained in its administration. The TST induration was recorded after 48–72 h by the ballpoint pen and ruler method. Phlebotomy for QFT-GIT was performed on the same day that the TSTs were administered and preceded placement of PPD. QFT-GIT assays were performed in a Cellestis-accredited laboratory and interpreted according to the manufacturer’s guidelines. Results were also recorded quantitatively (antigen-stimulated IFN- $\gamma$  corrected for background, expressed in IU·mL<sup>-1</sup>). Laboratory technicians were blinded to TST results, TB symptoms and signs and culture results, while clinicians were blinded to QFT-GIT results.

**Diagnostic accuracy of single tests for prevalent sputum smear-negative, culture-positive TB**

Using culture positivity as the gold standard, we computed standard test accuracy measures to detect prevalent disease. Post-test probabilities of disease based on negative test results were calculated to assess the clinical utility of a negative single stand-alone test result for ruling out active TB. Discriminatory ability was evaluated by receiver operating characteristic area or area under the curve of sensitivity *versus* 1 minus specificity (AUC). That is, based on the results of a stand-alone test, do patients with culture-positive TB have higher risk predictions than those without [24]?

**Selection of potential predictors of TB**

All analyses were performed with STATA 10 MP (StataCorp, College Station, TX, USA). Candidate clinical predictors were determined *a priori* based on clinical judgment and the published literature [10, 25] and they included: ART status (taking ART or not), sex, age, weight at screening, most recent CD4 count (no older than 6 months), prior TB and TB symptoms and clinical signs. TST at the clinically relevant cut-off of 5 mm and QFT-GIT at standard manufacturer’s cut-offs were pre-specified for multivariable analysis. Further details are provided in the online supplementary material, S-1.

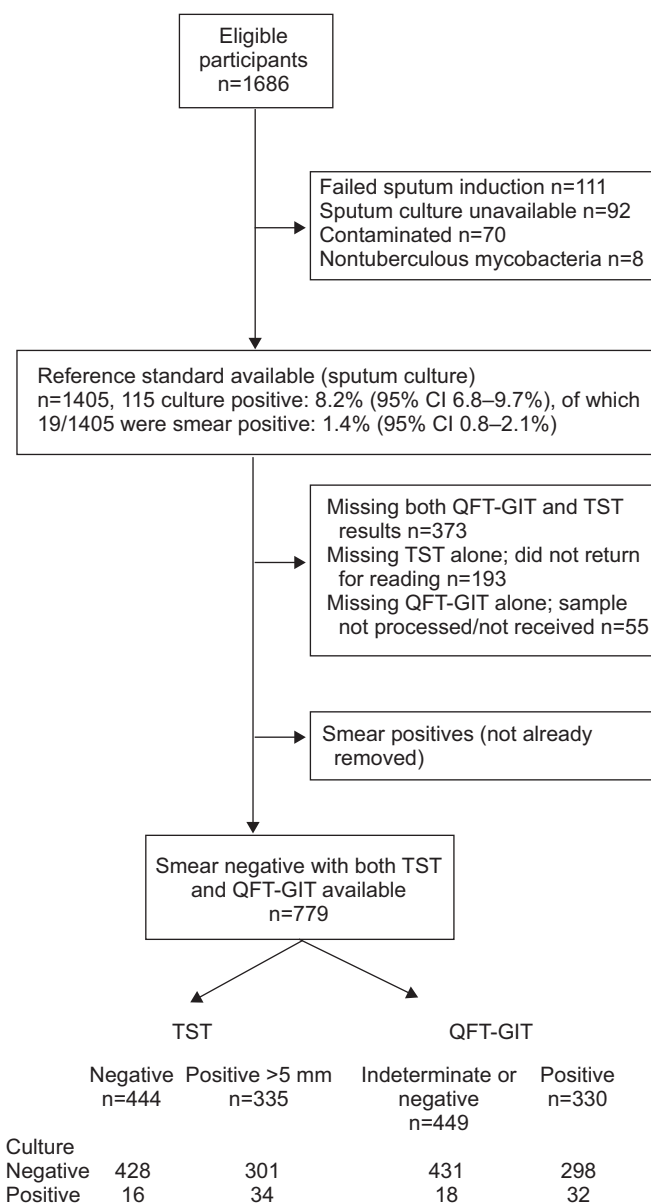
**Prediction model development and assessment of added discriminatory value**

Multivariable logistic regression analysis was performed to develop diagnostic models for culture-positive TB. First, a reduced, most parsimonious, clinical model without additional tests of TB infection (TST and QFT-GIT) was derived. TB tests of infection were added singly to the reduced clinical model and then simultaneously to explore the added predictive value of a single test and of combined tests, respectively. The ability of a multi-variable model to discriminate *M. tuberculosis* culture-positive and -negative patients was assessed by AUC. That is, based on the results of the test when added to multiple clinical predictors, do patients with the outcome have higher risk predictions than those without? The AUC was chosen as a measure of model discrimination because it is an objective estimate where individual predicted probabilities, and hence all probability cut-off points, are used in the assessment of model performance. Full details on model selection, calibration and discrimination are provided in the online supplementary material, S-2.

**RESULTS**

**Characteristics of the reference groups**

1,686 eligible participants were consecutively screened between November 2007 and September 2009 (fig. 1). Of these, 1,405 had culture results available, with an overall prevalence of culture-positive TB of 8.2% (95% CI 6.8–9.7%); after exclusions (fig. 1), 779 smear-negative participants were available for analysis. The prevalence of smear-negative, culture-positive TB was 6.4% (95% CI 4.9–8.4%). 56% of TB cases had at least one TB symptom or sign upon referral for DOTS. 88% of cases had CXR performed and 57% of those had signs suggestive of TB. The median (interquartile range) duration between the baseline clinical assessment and referral was 6 (4–12) weeks. The baseline characteristics of participants with both TST and QFT-GIT



**FIGURE 1.** Flow-chart detailing recruitment into the study. This shows participant flow and the numbers used in the final multivariable discriminatory value analysis. QFT-GIT: QuantIFERON®-TB Gold in-tube; TST: tuberculin skin test.

**TABLE 1** Characteristics of sputum smear-negative patients stratified by *Mycobacterium tuberculosis* culture status

Clinical and laboratory features	TB culture positive	TB culture negative	p-value
<b>Patients n</b>	50	729	
<b>Age yrs</b>	35 (31–40)	36 (31–42)	0.71
<b>Age ≥35 yrs</b>	46%	45%	0.92
<b>Male</b>	68%	75%	0.25
<b>No prior TB</b>	82%	62%	0.004
<b>CD4+ count cells per mm<sup>3</sup></b>	169 (98–239)	198 (136–315)	0.03
<b>CD4+ &lt;200 cells per mm<sup>3</sup></b>	62%	50% (n=721)	0.12
<b>CD4+ &lt;250 cells per mm<sup>3</sup></b>	80%	66% (n=721)	0.05
<b>Weight kg</b>	60 (54–65)	66 (58–76)	<0.001
<b>Weight &lt;60 kg</b>	52%	33% (n=722)	0.01
<b>Not on ART at screening</b>	54%	34%	0.004
<b>Symptoms and signs of TB</b>			
Cough ≥2 weeks	10%	4% (n=728)	0.05
Night sweats	10%	2% (n=728)	0.002
Self-reported fever n/N	1/49	3/727	0.230 (exact)
Nodes on examination n/N	1/49	1/728	0.122 (exact)
<b>Weight loss</b>	18%	5% (n=728)	<0.0001
<b>Any one positive TB symptom or sign</b>	26%	8% (n=728)	<0.0001
<b>Tests of TB infection</b>			
TST positive at 5 mm cut-off	68%	41%	<0.0001
TST positive at 10 mm cut-off	66%	37%	<0.0001
TST positive at 15 mm cut-off	54%	26%	<0.0001
TST mm	15 (0–20)	0 (0–15)	<0.0001
Manufacturer's cut-offs QFT-GIT positive	64%	41%	0.004 (exact)
QFT-GIT negative	30%	53%	
QFT-GIT indeterminate	6%	7%	
QFT-GIT quantitative	0.5 (0.1–2.6)	0.12 (0–0.85)	0.003
QFT-GIT tertiles			
<0.03	16% (n=45)	33% (n=695)	0.005 (score test for trend)
0.03 to <0.42	31% (n=45)	32% (n=695)	
≥0.42	53% (n=45)	35% (n=695)	
Either TST 5 mm/QFT-GIT positive (indeterminate included with negatives)	80%	56%	0.001
Either TST 5 mm/QFT-GIT positive (indeterminate results excluded)	83% (n=48)	59% (n=692)	0.001

Data are presented as median (interquartile range), unless otherwise stated. TB: tuberculosis; ART: antiretroviral therapy; TST: tuberculin skin test; QFT-GIT: QuantiFERON®-TB Gold in-tube.

results stratified by culture status are shown in table 1. Patients with prevalent TB had a lower CD4 count (median 169 *versus* 198 cells per mm<sup>3</sup>), were less likely to be taking ART at screening (54% *versus* 34%) and were more likely to report any of a range of symptoms and signs of TB, but the overall frequency of any one TB symptom or sign was low. At least 80% of those with culture-positive disease had either a positive TST (at 5 mm) or QFT-GIT result. The highest proportion of those with TB was represented in the highest QFT-GIT tertile compared with the lower strata (p=0.005, score test for trend).

#### Diagnostic accuracy of single tests for prevalent sputum smear-negative, culture-positive TB

The diagnostic accuracy measures of the TST and QFT-GIT as single stand-alone tests are moderate and comparable regardless of whether QFT-GIT results included indeterminate results or not (table 2). The combined variable of either TST or QFT-GIT positive was the most sensitive test, 80% (95% CI 66–90%) and

83% (95% CI 70–93%) when QFT results included indeterminate results as negatives or when indeterminate results are excluded, respectively. The clinical utility of single negative tests to rule out smear-negative, culture-positive TB was poor and the post-test probability of disease following application of negative tests remained close to the pre-test probability of 6%. The discriminatory ability of the individual predictors to identify those with disease and those without disease, based on AUC, was limited. Both the TST and QFT-GIT were comparable, and also comparable to the variable of any one positive TB symptom/sign (table 2). Details of other accuracy measures and those for TB symptoms are presented in table 2.

#### Predictors of prevalent active TB

The following clinical covariates emerged as significant independent predictors of TB during univariable logistic regression analyses (table 3): not being on ART at screening, weight <60 kg and CD4+ count <250 cells per mm<sup>3</sup>. No history of TB

**TABLE 2** Diagnostic accuracy of single tests: tuberculosis (TB) symptoms and signs, tuberculin skin test (TST) and QuantiferON®-TB Gold in-tube (QFT-GIT)

Clinical and laboratory tests	Sensitivity %	Specificity %	Pr D+ T+ <sup>‡</sup> %	Pr D- T- <sup>‡</sup> %	LR <sup>+</sup> <sup>§</sup>	LR <sup>-</sup> <sup>§</sup>	% Post-test Pr <sup>f</sup>	AUC <sup>##</sup>
<b>TB symptoms and signs</b>								
Cough for ≥2 weeks	10 (3–22)	96 (94–97)	14 (5–30)	94 (92–96)	2.4 (1.0–6)	0.9 (0.9–1)	5.7 (4.8–6.5)	53 (49–57)
Night sweats	10 (3–22)	98 (96–99)	23 (8–45)	94 (92–96)	4.3 (1.7–11)	0.9 (0.8–1)	5.7 (4.8–5.7)	54 (50–58)
Self-reported fever	2 (0.1–11)	99.6 (98.8–99.9)	25 (0.6–81)	94 (92–95)	5.0 (0.5–47)	1.0 (0.9–1.0)	5.7 (5.7–6.5)	51 (49–53)
Weight loss	18 (9–31)	95 (93–96)	20 (9–34)	94 (93–96)	3.5 (1.8–6.9)	0.9 (0.8–1.0)	5.7 (4.8–6.5)	57 (51–62)
Nodes on examination (n=777, prevalence 6% (95% CI 5–8%))	2 (0.1–11)	99.9 (99–100)	50 (1–99)	94 (92–95)	15 (0.9–234)	1.0 (0.9–1.0)	5.7 (5.7–6.5)	51 (49–53)
Any one positive TB symptom/sign	26 (15–40)	92 (90–94)	18 (10–29)	95 (93–96)	3.2 (1.9–5.4)	0.8 (0.7–1.0)	4.8 (3.8–5.7)	59 (53–65)
<b>Tests for TB infection/disease</b>								
TST (5 mm cut-off)	68 (53–81)	59 (55–62)	10 (7–14)	96 (94–98)	1.7 (1.3–2.0)	0.5 (0.4–0.8)	2.9 (2.0–4.8)	63 (57–70)
TST (10 mm cut-off)	66 (51–79)	63 (59–66)	11 (8–15)	96 (94–98)	1.8 (1.4–2.2)	0.5 (0.4–0.8)	2.9 (2.0–4.8)	64 (58–71)
TST (15 mm cut-off)	54 (39–68)	74 (71–77)	12 (8–18)	96 (94–97)	2.1 (1.6–2.8)	0.6 (0.5–0.8)	3.8 (2.9–4.8)	64 (57–71)
QFT-GIT, indeterminates included with negatives	64 (49–77)	59 (56–63)	10 (7–13)	96 (94–98)	1.6 (1.3–2.0)	0.6 (0.4–0.9)	3.8 (2.9–5.7)	62 (55–69)
QFT-GIT, without indeterminates (n=728, prevalence 7% (5–8%))	68 (53–81)	56 (52–60)	10 (8–13)	96 (94–98)	1.6 (1.3–1.9)	0.6 (0.4–0.9)	3.8 (2.9–5.7)	62 (55–69)
Either TST 5 mm/QFT-GIT positive (indeterminate included as negative)	80 (66–90)	44 (40–48)	9 (6–12)	97 (95–99)	1.4 (1.2–1.7)	0.5 (0.3–0.8)	2.9 (2.0–4.8)	62 (56–68)
Either TST 5 mm/QFT-GIT positive (indeterminate results excluded)	83 (70–93)	41 (37–45)	9 (6–12)	97 (95–99)	1.4 (1.2–1.6)	0.4 (0.2–0.8)	2.9 (1.0–4.8)	62 (57–68)

Data are presented with 95% confidence intervals. Pr: probability; D: disease (culture-positive TB); T: test; LR: likelihood ratio; AUC: area under the curve. Restricted to patients with smear negative results for *Mycobacterium tuberculosis* with all three TST, QFT-GIT and culture results available (n=779, prevalence 6% (5–8%)), except where indicated. <sup>‡</sup>: probability of a culture-positive result given a positive test. <sup>§</sup>: probability of a culture negative result given a negative test. <sup>¶</sup>: likelihood ratios for a positive (LR+) and negative (LR-) test. <sup>f</sup>: negative test; post-test odds/(1+post-test odds) where post-test odds=(Prevalence/(1-Prevalence)) x LR-. <sup>##</sup>: AUC of sensitivity and 1 minus specificity. Disease prevalence in the study (proportion culture positive) was used as the best estimate of pre-test probability.

in the past was a strong predictor of prevalent TB; that is, those who had been treated for active TB were less likely to have prevalent TB. Patients with any one positive TB symptom/sign were 4.0 times as likely to have TB compared with those with no symptoms (OR 4.0 (95% CI 2.0–7.9)). The TST at the 5, 10 and 15-mm threshold (OR 3.0 (95% CI 1.6–5.6), OR 3.3 (95% CI 1.8–6.0) and OR 3.3 (95% CI, 1.9–5.9), respectively) and TST quantitative results also emerged as significant predictors. Patients with a positive QFT-GIT response (quantitative and at the manufacturer's thresholds) were 2.7 times more likely to have culture positive TB (OR 2.7 (95% CI 1.5–5.2)). Either TST or QFT-GIT positivity was also a significant predictor of TB. There was a tendency towards increasing positivity for those with indeterminate and positive results compared with those with negative responses; however, the score test for trend was not significant (p=0.88). When QFT-GIT responses were grouped into tertiles, it was seen that individuals in the highest then the middle tertiles were three- and two-fold as likely, respectively, to have culture-confirmed active TB compared with those with negative responses (score test for trend p=0.005).

#### Multivariable logistic regression models and the discriminatory value of TST or QFT-GIT when added to the final multivariable clinical model

Our final clinical prediction model included the following predictors: not being on ART at screening; weight <60 kg; no prior history of TB; any one positive TB symptom or sign; and CD4+ count <250 cells per mm<sup>3</sup> (table 4). Compared with individual predictors of TB in table 2, this multivariate model had significantly improved discrimination for active TB (AUC of 0.72 for the model versus AUC range of 0.51–0.64 for predictors; individual AUC comparison p-values are not displayed). The clinical model was then extended by adding the TST (at 5 mm) and QFT-GIT (manufacturer's cut-off) as separate tests and with both included in the extended model. Models showed good calibration using the Hosmer–Lemeshow goodness of fit test (table 4). The odds ratios for the extended models with TST and QFT-GIT, added singly and together, were similar and the confidence intervals overlapped. However, there was no significant difference in the AUC between the QFT-GIT extended model and the clinical model (p=0.33 for AUC comparisons). By contrast, the discriminatory ability of the clinical model was statistically significantly improved by addition of the TST (p=0.04 for AUC comparisons; table 4). However, the confidence levels of the TST and QFT-GIT extended clinical models overlapped. Simultaneously adding both TST and QFT-GIT to the best clinical model also resulted in an AUC that was significantly different from the clinical model, suggesting further overall improved discriminatory ability for culture-positive and -negative TB.

In secondary analyses, addition of either a positive TST (5 mm) or QFT-GIT result, and of quantitative QFT-GIT divided into tertiles of increasing reactivity did not improve the final clinical model (for details see the online supplementary material, S-3).

#### DISCUSSION

Undiagnosed TB is a major concern in HIV-infected persons. Existing clinical algorithms are not highly accurate in smear-negative patients and available biomarkers, such as IGRAs, perform moderately. Our main finding was that even when considered as an adjunct test, QFT-GIT did not improve the

**TABLE 3** Univariable predictors of active tuberculosis (TB)

Clinical and laboratory tests	OR (95% CI)
<b>Clinical observations</b>	
Not on ART at screening	2.3 (1.3–4.1)
Male	1.4 (0.8–2.7)
Age yrs	0.99 (0.96–1.02)
Age ≥35 yrs	1.02 (0.6–1.8)
Weight kg	0.95 (0.92–0.97)
Weight <60 kg	2.2 (1.3–4.0)
CD4+ <200 cells per mm <sup>3</sup>	1.6 (0.9–2.9)
CD4+ <250 cells per mm <sup>3</sup>	2.0 (1.0–4.1)
CD4+ count cells per mm <sup>3</sup>	1.0 (0.99–1.0)
No prior TB	2.8 (1.3–5.9)
<b>TB symptoms and signs</b>	
Cough for ≥2 weeks	2.6 (1.0–7.0)
Night sweats	4.6 (1.6–13.2)
Self-reported fever	5.0 (0.5–49.2)
Weight loss	4.1 (1.9–9.1)
Nodes on examination	15.1 (0.9–246)
Any one positive TB symptom/sign	4.0 (2–7.9)
<b>Tests for TB infection/disease</b>	
TST (5 mm cut-off)	3.0 (1.6–5.6)
TST (10 mm cut-off)	3.3 (1.8–6.0)
TST (15 mm cut-off)	3.3 (1.9–5.9)
TST mm (quantitative)	1.1 (1.02–1.1)
QFT-GIT (manufacturer's cut-offs)	
Negative	1
Indeterminate	1.6 (0.4–5.7)
Positive	2.7 (1.5–5.2)
QFT-GIT (tertiles)	
<0.03	1
0.03 to <0.42	2.1 (0.8–5.3)
≥0.42	3.2 (1.4–7.6)
QFT-GIT (quantitative)	1.1 (1.0–1.2)
Either TST 5 mm/QFT-GIT positive (indeterminate included as negative)	3.1 (1.6–6.4)
Either TST 5 mm/QFT-GIT positive (indeterminate results excluded)	3.5 (1.6–7.5)

Following basic explorations of data, age and weight were dichotomised at medians of 35 yrs and 60 kg, respectively. ART: anti-retroviral therapy; TST: tuberculin skin test; QFT-GIT: QuantIFERON<sup>®</sup>-TB Gold in-tube.

ability of our best multivariable clinical model to discriminate between culture positive and culture negative. This incremental discriminatory value of QFT-GIT when added to current clinical algorithms used in assessment of HIV-infected persons in whom TB must be ruled out prior to TB prevention had not previously been explored. The WHO has identified this as a gap in the IGRA body of knowledge [18, 23].

In a conventional analysis, QFT-GIT interpreted according to the manufacturer's guidelines, did not have high sensitivity (64–68%) and had poor specificity (56–59%) for active disease in this setting. Furthermore, single negative test results did not significantly reduce the post-test probability of disease. This is not new. Moderate sensitivity for active TB was also shown in a recently published systematic review and meta-analysis for

the diagnosis of latent TB infection in HIV-infected individuals (61% (95% CI 41–75%) for QFT-GIT, using culture as a surrogate reference standard) [26]. Thus, neither QFT-GIT nor the TST should be used in isolation as a rule out or rule in test for active TB prior to IPT provision. This conclusion is in line with a similar body of evidence that suggested that stand-alone IGRA tests have limited application in excluding active TB in challenging clinical scenarios [14, 26–29]. Based on such data, WHO plans to issue a statement to discourage the use of commercial IGRAs in low- and middle-income countries [18]. However, simple test accuracy based on single stand-alone tests does not in itself demonstrate clinical relevance [20, 21, 30, 31]. It is important to demonstrate the extent to which IGRAs have true added value in the clinical context, when all identified predictors of TB are considered.

In our multivariable analyses, not being on ART at screening, weight <60 kg, no history of prior TB, any one positive TB symptom or clinical sign (includes cough ≥2 weeks) and CD4+ count <250 cells per mm<sup>3</sup> all emerged as strong independent predictors of smear-negative, culture-positive TB (AUC 72%). Not having had TB in the past is strongly associated with prevalent active TB in HIV-infected individuals and has previously been cited as a predictor of risk in ART cohorts [32–34]. The value of this predictor is interesting and was highlighted in our recent immunological evaluation of T-cell responses of HIV-infected patients without active TB who were newly starting ART [35]. Patients with a prior history of TB who had previously completed TB treatment had a reduced response in an in house IGRA than those who had not previously received TB treatment, possibly suggesting the likelihood of having a lower bacterial load [35]. Clinical prediction models derived for similar settings should therefore continue to evaluate this predictor.

QFT-GIT did not improve the ability of the clinical model to discriminate between culture positive and culture negative. By contrast, addition of the TST (5-mm cut-off) and the combination of both the TST (5-mm cut-off) and QFT-GIT at the standard manufacturer's cut-off statistically significantly improved the discriminatory ability of the clinical tools (AUC ~80% for both models). However, the AUC confidence intervals overlapped. The conclusion that, in its current format, QFT-GIT may add very little to current clinical tools is consistent with recent evidence. A paper by METCALFE *et al.* [19] explored whether QFT-GIT improves the classification of HIV-uninfected, smear-negative pulmonary and extrapulmonary TB suspects into researcher-selected risk categories. They concluded that although QFT-GIT improved risk stratification of patients when added to clinical and demographic risk factors, there was no additional benefit when clinical judgment of experienced clinicians was also considered.

Despite the current poor performance of commercial IGRAs, the technology remains a significant operational and conceptual advance. It is logistically simpler than the TST, especially in the provider/patient environment and in scoring. The TST could not be determined for 193 patients as they failed to return for their readings. By contrast, fewer (55 subjects) QFT-GIT results were unobtainable, mainly as a result of blood samples not being processed. A further improved IGRA might therefore provide greater clinical utility. Two approaches being pursued include the addition of new antigens [36] or the evaluation of different cytokine biomarkers [37]. It is also

**TABLE 4** Multivariable logistic regression estimates for culture-positive disease in clinical model with and without tests of tuberculosis (TB) infection

Multivariable predictors	A. Clinical model	B. With TST (5 mm)	C. With QFT-GIT	D. With TST (5 mm) and QFT-GIT
<b>Clinical</b>				
Not on ART at screening	1.2 (0.60–2.4)	1.2 (0.60–2.5)	1.3 (0.63–2.6)	1.3 (0.62–2.6)
Weight <60 kg	2.3 (1.3–4.1)	2.6 (1.4–4.7)	2.6 (1.4–4.8)	2.8 (1.5–5.1)
No prior TB	2.7 (1.2–6.0)	2.5 (1.2–5.6)	2.4 (1.1–5.3)	2.4 (1.1–5.2)
Any one positive TB symptom/sign	3.1 (1.5–6.3)	2.9 (1.4–6.1)	3.0 (1.4–6.3)	2.9 (1.4–6.2)
CD4+ count <250 cells per mm <sup>3</sup>	1.5 (0.66–3.4)	1.8 (0.8–4.2)	1.6 (0.7–3.8)	1.8 (0.8–4.2)
<b>Tests of TB infection</b>				
TST positive at 5 mm		3.5 (1.9–6.7)		2.7 (1.4–5.4)
QFT-GIT (manufacturer's cut-offs)				
Positive			3.1 (1.6–5.9)	2.1 (1.1–4.3)
Indeterminate			1.5 (0.4–5.4)	1.5 (0.4–5.5)
Negative			1	1
<b>Incremental value performance measures</b>				
Hosmer–Lemeshow goodness of fit p-value <sup>#</sup>	0.791	0.993	0.684	0.895
Akaike information criterion <sup>†</sup>	351	337	343	336
AUC (95% CI) <sup>§</sup>	0.72 (0.65–0.80)	0.77 (0.70–0.84)	0.74 (0.67–0.82)	0.78 (0.71–0.85)
AUC comparison p-value <sup>§</sup>		0.04	0.33	0.02
LRT p-value <sup>§</sup>		<0.001	0.002	<0.001

Data are presented as OR (95% CI), unless otherwise stated. ART: antiretroviral therapy; TST: tuberculin skin test; QFT-GIT: QuantiFERON<sup>®</sup>-TB Gold in-tube; AUC: area under the curve; LRT: likelihood ratio test. <sup>#</sup>: H<sub>0</sub> (null): there is no difference between observed and model-predicted probabilities; a p-value closer to 1 indicates good calibration. <sup>†</sup>: a small Akaike information criterion infers minimum prediction error. <sup>§</sup>: models B–D compared with A. The LRT is for the reduced model nested in the full model. p<0.05 indicate that the added predictor is statistically important to the model. AUC comparison between the clinical model extended with TST (5 mm) and the model extended with both TST (5 mm) and QFT-GIT, p=0.51.

evident from our finding of a prevalence of culture positivity of 6.4%, in the absence of symptoms in many cases, that the distinction between latent and active TB may be arbitrary. TB infection appears to exist as a spectrum and the immune responses that characterise active and latent disease therefore inevitably overlap [38, 39]. Insight into the distinction between active and latent TB, and the necessity to potentially factor multiple biomarkers to achieve consistent risk stratification comes from transcriptional profiling of the blood of TB patients [40]. 393 differentially elevated or reduced transcripts characterised active TB but ~10% latent patients also had a transcriptional pattern that resembled active disease.

Our finding that TST may offer some utility in pre-IPT assessment was unexpected. The TST is often falsely negative in moderate-to-severely immunocompromised HIV-1 infected persons, which often limits its utility [41]. However, evaluation hitherto has been as a stand-alone test and not as part of a multivariable clinical process. A major pitfall of adopting conventional test accuracy measures as the source of information for considering a test in clinical practice is that accuracy estimates are assumed to be constant [22, 42]. These can differ across populations, within groups of patients with different risk profiles (e.g. age, symptoms and signs, etc.) and can vary according to other test results [42]. Therefore, when those additional predictors of disease are considered in clinical diagnosis, as simulated in our multivariable models, the accuracy of the test is modified [42]. This possibly explains the improved discriminatory ability of TB symptoms when considered in multivariable models, despite poor individual

sensitivity estimates. Similarly, the marginal difference in the added discriminatory ability between TST and QFT-GIT, despite comparable sensitivity estimates, is explained. As a tool for assessing the models' ability to discriminate *M. tuberculosis* culture-positive and -negative patients, the AUC may be insensitive to model changes [43, 44] and does not give an absolute indication of risk prediction. This work may be improved by an assessment in the independent absolute risk of culture-positive TB disease, of the model without TB tests of infection from addition of TST and QFT-GIT, by the use of new risk stratification methods [43, 45, 46]. A clinical scoring rule based on the best-extended clinical model would further provide an application of the model in practice; this would aid our appreciation of the models' absolute clinical relevance. The TST identifies HIV-infected persons who will benefit from short-term IPT [47] as well as those that may benefit from continuous therapy [48]. However, there is little enthusiasm for performing TST in programmes. If addition of TST to existing clinical screening algorithms could additionally help stratify those who should further be evaluated by culture, this might prove useful. However, the possibility that this utility may be location specific means the TST should be validated prospectively in multivariable clinical models in other similar settings.

Our study had limitations. CXR was not included in our multivariable assessment as radiography was not part of our pre-IPT screening protocol. The decision not to include radiography in our 2007 protocol was based on national IPT guidelines and supported by data from two IPT trials, one local [8] and the other a study that evaluated the value of CXR for Botswana's

national IPT programme [7]. The local study found that CXR added very little to a clinical algorithm that included two of more standard TB symptoms and signs. The Botswana study similarly suggested that screening of TB symptoms alone was adequate to rule out TB and that CXR added very little (0.2% of 560 HIV asymptomatic patients screened in the Botswana programme had TB diagnosed solely on the basis of radiography). However, a number of studies have since been published on the utility of CXR in the pre-IPT evaluation of HIV-infected patients but, overall, with inconclusive results. Based on four such studies (n=2,805), a WHO-led patient meta-analysis suggested that the addition of CXR results to their best clinical rule increased sensitivity from about 79% to 90% but decreased specificity further from 50% to 40% [12]. This will unlikely be a useful rule-in strategy for a clinical setting in a high-burden country. This lack of consensus on the utility of the CXR in the pre-IPT evaluation of HIV infected patients is further reflected in the recently published WHO guidelines for intensified TB case-finding and IPT for people living with HIV in resource-constrained settings, a symptoms screen and not CXR is mandated [49].

A single sputum specimen, ultrasonically induced in some cases, was sent for microscopy and culture. It is possible that more culture-positive disease may have been diagnosed if more than one sample had been sent. All samples were processed in an accredited laboratory according to strict protocols to prevent cross-contamination. Our clinical model included absolute weight rather than body mass index. Although this potentially simplifies clinical algorithms, it may not be very useful at the level of the individual.

Our study also had strengths. First, this was a large, single-centre study undertaken in the context of high TB incidence and among HIV-infected patients, either already on ART or about to commence ART, evaluated for TB prevention. It focused on a difficult-to-diagnose group of smear-negatives. Secondly, our multivariable models were developed on a large dataset that included a small number of *a priori* determined predictors, and the ratio of TB cases to the number of parameters in our models was around the cited rule of thumb of 1:10, therefore limiting model over-fitting [50]. Furthermore, selection of predictors was based on clinical judgment and published literature and not just on statistical significance (p-values/Akaike information criterion). Thirdly, objective measures of incremental value and the tests' ability to discriminate TB patients from those without TB were used, thus going beyond conventional standard test accuracy.

In summary, QFT-GIT, in its current formulation, does not improve the discriminatory ability of current TB screening clinical algorithms and should not be used alone to rule out or rule in active TB prior to IPT provision. However, an improved IGRA might provide greater clinical utility. Further research is required to improve IGRAs either through the addition of new antigens or the evaluation of different cytokine biomarkers. Evaluation of new TB diagnostics for clinical relevance should follow a multivariable process that goes beyond test accuracy.

## SUPPORT STATEMENT

This work was supported by the Foundation for Innovative New Diagnostics (FIND), the Wellcome Trust (grants 084323, 084670 and 088316), the Dept of Health of South Africa and the Medical Research Council of the UK and the European Union (SANTE/2005/105-061-102). M. Pai was supported by a New Investigator Award from

the Canadian Institutes of Health Research. When this work was conducted, M. Pai was a consultant for FIND. M. Pai is a consultant for the Bill and Melinda Gates Foundation.

## STATEMENT OF INTEREST

None declared.

## ACKNOWLEDGEMENTS

We thank patients and staff at Ubuntu Clinic in Khayelitsha, South Africa for inclusion and clinical evaluation, respectively. The ART plus IPT Study Team is thanked for its assistance and V. Azevedo (Khayelitsha Subdistrict, City of Cape Town Health Dept, Cape Town, South Africa) for valuable input.

## REFERENCES

- 1 Maartens G, Wilkinson RJ. Tuberculosis. *Lancet* 2007; 370: 2030–2043.
- 2 World Health Organization. WHO Three I's Meeting. Intensified Case Finding (ICF), Isoniazid Preventive Therapy (IPT), and TB Infection Control (IC) with people living with HIV. Report of a joint World Health Organization HIV/AIDS and TB department meeting. Geneva, World health Organization, 2008.
- 3 Balcells ME, Thomas SL, Godfrey-Faussett P, *et al.* Isoniazid preventive therapy and risk for resistant tuberculosis. *Emerg Infect Dis* 2006; 12: 744–751.
- 4 Kranzer K, Houben RM, Glynn JR, *et al.* Yield of HIV-associated tuberculosis during intensified case finding in resource-limited settings: a systematic review and meta-analysis. *Lancet Infect Dis* 2010; 10: 93–102.
- 5 Bassett IV, Wang B, Chetty S, *et al.* Intensive tuberculosis screening for HIV-infected patients starting antiretroviral therapy in Durban, South Africa. *Clin Infect Dis* 2010; 51: 823–829.
- 6 Meintjes G, Wilkinson RJ. Undiagnosed active tuberculosis in HIV-infected patients commencing antiretroviral therapy. *Clin Infect Dis* 2010; 51: 830–832.
- 7 Mosimaneotsile B, Talbot EA, Moeti TL, *et al.* Value of chest radiography in a tuberculosis prevention programme for HIV-infected people, Botswana. *Lancet* 2003; 362: 1551–1552.
- 8 Mohammed A, Ehrlich R, Wood R, *et al.* Screening for tuberculosis in adults with advanced HIV infection prior to preventive therapy. *Int J Tuberc Lung Dis* 2004; 8: 792–795.
- 9 Day JH, Charalambous S, Fielding KL, *et al.* Screening for tuberculosis prior to isoniazid preventive therapy among HIV-infected gold miners in South Africa. *Int J Tuberc Lung Dis* 2006; 10: 523–529.
- 10 Cain KP, McCarthy KD, Heilig CM, *et al.* An algorithm for tuberculosis screening and diagnosis in people with HIV. *N Engl J Med* 2010; 362: 707–716.
- 11 Corbett EL, Bandason T, Cheung YB, *et al.* Prevalent infectious tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control. *Int J Tuberc Lung Dis* 2009; 13: 1231–1237.
- 12 Getahun H, Kittikraisak W, Heilig CM, *et al.* Development of a standardized screening rule of tuberculosis in people living with HIV in resource-constrained settings: individual participant data meta-analysis of observational studies. *PLoS Med* 2011; 8: e1000391.
- 13 Seshadri C, Uiso LO, Ostermann J, *et al.* Low sensitivity of T-cell based detection of tuberculosis among HIV co-infected Tanzanian in-patients. *East Afr Med J* 2008; 85: 442–449.
- 14 Nicol MP, Davies MA, Wood K, *et al.* Comparison of T-SPOT.TB assay and tuberculin skin test for the evaluation of young children at high risk for tuberculosis in a community setting. *Paediatrics* 2009; 123: 38–43.
- 15 Liao CH, Lai CC, Tan CK, *et al.* False-negative results by enzyme-linked immunospot assay for interferon- $\gamma$  among patients with culture-confirmed tuberculosis. *J Infect* 2009; 59: 421–423.



- 16 Bamford AR, Crook AM, Clark JE, *et al.* Comparison of interferon- $\gamma$  release assays and tuberculin skin test in predicting active tuberculosis (TB) in children in the UK: a paediatric TB network study. *Arch Dis Child* 2010; 95: 180–186.
- 17 Ling DI, Pai M, Davids V, *et al.* Are interferon- $\gamma$  release assays useful for diagnosing active tuberculosis in a high-burden setting? *Eur Respir J* 2011; 38: 649–656.
- 18 World Health Organization. Strategic and Technical Advising Group for Tuberculosis (STAG-TB). Report of the Tenth Meeting. Geneva, World Health Organization, 2010.
- 19 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.
- 20 Moons KG, Biesheuvel CJ, Grobbee DE. Test research *versus* diagnostic research. *Clin Chem* 2004; 50: 473–476.
- 21 Moons KG, Grobbee DE. Diagnostic studies as multivariable, prediction research. *J Epidemiol Community Health* 2002; 56: 337–338.
- 22 Moons KG, van ES GA, Michel BC, *et al.* Redundancy of single diagnostic test evaluation. *Epidemiology* 1999; 10: 276–281.
- 23 World Health Organization TB/HIV Working Group Stop TB Partnership. Priority research questions for TB/HIV in HIV-prevalent and resource-limited settings. Report No.: WHO/HTM/TB/2010. Geneva, World Health Organization, 2010.
- 24 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36.
- 25 Corbett EL, Zezai A, Cheung YB, *et al.* Provider-initiated symptom screening for tuberculosis in Zimbabwe: diagnostic value and the effect of HIV status. *Bull World Health Organ* 2010; 88: 13–21.
- 26 Cattamanchi A, Smith R, Steingart KR, *et al.* Interferon- $\gamma$  release assays for the diagnosis of latent tuberculosis infection in HIV-infected individuals: a systematic review and meta-analysis. *J Acquir Immune Defic Syndr* 2011; 56: 230–238.
- 27 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. *BMC Infect Dis* 2010; 10: 75.
- 28 Davies MA, Connell T, Johannisen C, *et al.* Detection of tuberculosis in HIV-infected children using an enzyme-linked immunospot assay. *AIDS* 2009; 23: 961–969.
- 29 Mazurek M, Jereb J, Vernon A, *et al.* Updated guidelines for using interferon- $\gamma$  release assays to detect *Mycobacterium tuberculosis* infection – United States, 2010. *MMWR Recomm Rep* 2010; 59: 1–25.
- 30 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.
- 31 Kattan MW. Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst* 2003; 95: 634–635.
- 32 Lawn SD, Badri M, Wood R. Tuberculosis among HIV-infected patients receiving HAART: long term incidence and risk factors in a South African cohort. *AIDS* 2005; 19: 2109–2116.
- 33 Lawn SD, Badri M, Wood R. Risk factors for tuberculosis among HIV-infected patients receiving antiretroviral treatment. *Am J Respir Crit Care Med* 2005; 172: 1348.
- 34 Lawn SD, Myer L, Bekker LG, *et al.* Burden of tuberculosis in an antiretroviral treatment programme in sub-Saharan Africa: impact on treatment outcomes and implications for tuberculosis control. *AIDS* 2006; 20: 1605–1612.
- 35 Wilkinson KA, Seldon R, Meintjes G, *et al.* Dissection of regenerating T-cell responses against tuberculosis in HIV-infected adults sensitized by *Mycobacterium tuberculosis*. *Am J Respir Crit Care Med* 2009; 180: 674–683.
- 36 Dosanjh DP, Hinks TS, Innes JA, *et al.* Improved diagnostic evaluation of suspected tuberculosis. *Ann Intern Med* 2008; 148: 325–336.
- 37 Chegou NN, Black GF, Kidd M, *et al.* Host markers in QuantiFERON supernatants differentiate active TB from latent TB infection: preliminary report. *BMC Pulm Med* 2009; 9: 21.
- 38 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: 845–855.
- 39 Young DB, Gideon HP, Wilkinson RJ. Eliminating latent tuberculosis. *Trends Microbiol* 2009; 17: 183–188.
- 40 Berry MP, Graham CM, McNab FW, *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010; 466: 973–977.
- 41 Rangaka MX, Wilkinson KA, Seldon R, *et al.* Effect of HIV-1 infection on T-cell-based and skin test detection of tuberculosis infection. *Am J Respir Crit Care Med* 2007; 175: 514–520.
- 42 Moons KG, van Es GA, Deckers JW, *et al.* Limitations of sensitivity, specificity, likelihood ratio, and Bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 1997; 8: 12–17.
- 43 Cook NR. Statistical evaluation of prognostic *versus* diagnostic models: beyond the ROC curve. *Clin Chem* 2008; 54: 17–23.
- 44 Greenland P. Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond" by M.J. Pencina, R.B. D'Agostino Sr, R.B. D'Agostino Jr, R.S. Vasan, *Statistics in Medicine* (DOI: 10.1002/sim.2929). *Stat Med* 2008; 27: 188–190.
- 45 Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; 21: 128–138.
- 46 Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, *et al.* 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.
- 47 Akolo C, Adetifa I, Shepperd S, *et al.* Treatment of latent tuberculosis infection in HIV infected persons. *Cochrane Database Syst Rev* 2010; CD000171.
- 48 Samandari T, Agizew TB, Nyirenda S, *et al.* 6-month *versus* 36-month isoniazid preventive treatment for tuberculosis in adults with HIV infection in Botswana: a randomised, double-blind, placebo-controlled trial. *Lancet* 2011; 377: 1588–1598.
- 49 StopTB. Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings. Geneva, World Health Organization, 2011.
- 50 Peduzzi P, Concato J, Kemper E, *et al.* A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49: 1373–1379.