

# Evaluation of Quantitative IFN- $\gamma$ Response for Risk Stratification of Active Tuberculosis Suspects

John Z. Metcalfe<sup>1</sup>, Adithya Cattamanchi<sup>1</sup>, Eric Vittinghoff<sup>2</sup>, Christine Ho<sup>3,4</sup>, Jennifer Grinsdale<sup>3</sup>, Philip C. Hopewell<sup>1,3</sup>, L. Masae Kawamura<sup>3</sup>, and Payam Nahid<sup>1,3</sup>

<sup>1</sup>Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, and <sup>2</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco; <sup>3</sup>Tuberculosis Control Section, Department of Public Health; and <sup>4</sup>Centers for Disease Control and Prevention, San Francisco, California

**Rationale:** The contribution of interferon- $\gamma$  release assays (IGRAs) to appropriate risk stratification of active tuberculosis suspects has not been studied.

**Objectives:** To determine whether the 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 (IFN)- $\gamma$  results measured by QuantiFERON-TB Gold (Cellestis, Carnegie, Australia).

**Measurements and Main Results:** Of 660 patients meeting eligibility criteria, 65 (10%) had culture-proven tuberculosis. The odds of active tuberculosis increased by 7% (95% confidence interval [CI], 3–11%) for each doubling of IFN- $\gamma$  level. The addition of quantitative IFN- $\gamma$  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-risk or lower-risk categories. However, quantitative IFN- $\gamma$  results did not significantly improve appropriate risk reclassification beyond that provided by clinician assessment of risk (4%; 95% CI, –7 to +22%;  $P = 0.14$ ).

**Conclusions:** Higher quantitative IFN- $\gamma$  results were associated with active tuberculosis, and added clinical value to a prediction model incorporating conventional risk factors. Although this benefit may be attenuated within highly experienced centers, the predictive accuracy of quantitative IFN- $\gamma$  levels should be evaluated in other settings.

**Keywords:** interferon-gamma release assay; Quantiferon; risk prediction; risk reclassification

Interferon- $\gamma$  release assays (IGRAs) are *in vitro* immunodiagnostic tests that measure effector T cell mediated interferon (IFN)- $\gamma$  response to *Mycobacterium tuberculosis*-specific antigens. IGRAs are as sensitive and more specific than the tuberculin skin test for detecting latent tuberculosis infection (LTBI) (1,

(Received in original form June 30, 2009; accepted in final form October 1, 2009)

Supported by grants T32 HL007185 (J.Z.M.), K23 HL094141 (A.C.), R01 AI034238 (P.C.H.), and K23 HL092629 (P.N.) from the National Institutes of Health.

Correspondence and requests for reprints should be addressed to John Z. Metcalfe, M.D., M.P.H., University of California, San Francisco Division of Pulmonary and Critical Care Medicine, 1001 Potrero Avenue, Rm 5K1, San Francisco, CA 94110-0111. E-mail: john.metcalfe@ucsf.edu

This article has an online supplement, which is accessible in this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org)

Am J Respir Crit Care Med Vol 181, pp 87–93, 2010

Originally Published in Press as DOI: 10.1164/rccm.200906-0981OC on October 1, 2009  
Internet address: [www.atsjournals.org](http://www.atsjournals.org)

## AT A GLANCE COMMENTARY

### Scientific Knowledge on the Subject

The role of interferon- $\gamma$  release assays (IGRAs) in the evaluation of active tuberculosis suspects is controversial. To date, whether IGRAs improve classification of smear-negative tuberculosis suspects into clinically relevant risk categories has not been examined.

### What This Study Adds to the Field

Quantitative interferon- $\gamma$  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 considered.

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 to experience widespread use, could be used for targeted screening of LTBI in all circumstances in which the tuberculin skin test (TST) is used (9).

Although the advantages of IGRAs in diagnosing LTBI are well established, their role in evaluating active tuberculosis suspects remains unclear. IGRAs have variable, although 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 individual patients' risk for 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 high risk are typically initiated on antituberculosis 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- $\gamma$  response measured by QFT-G (Cellestis) to routine clinical evaluation improves risk stratification of individuals suspected of having smear-negative pulmonary and extrapulmonary tuberculosis.

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, Centers for Disease Control and Prevention, and Infectious Diseases Society of America 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 more than 14 days before or 14 days after their initial clinic visit, or (3) performed more than 7 days into a course of tuberculosis treatment, were excluded. In addition, patients younger than 10 years of age (where adult-type, nonpaucibacillary disease is uncommon) (21, 22), with a positive AFB smear examination, a 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

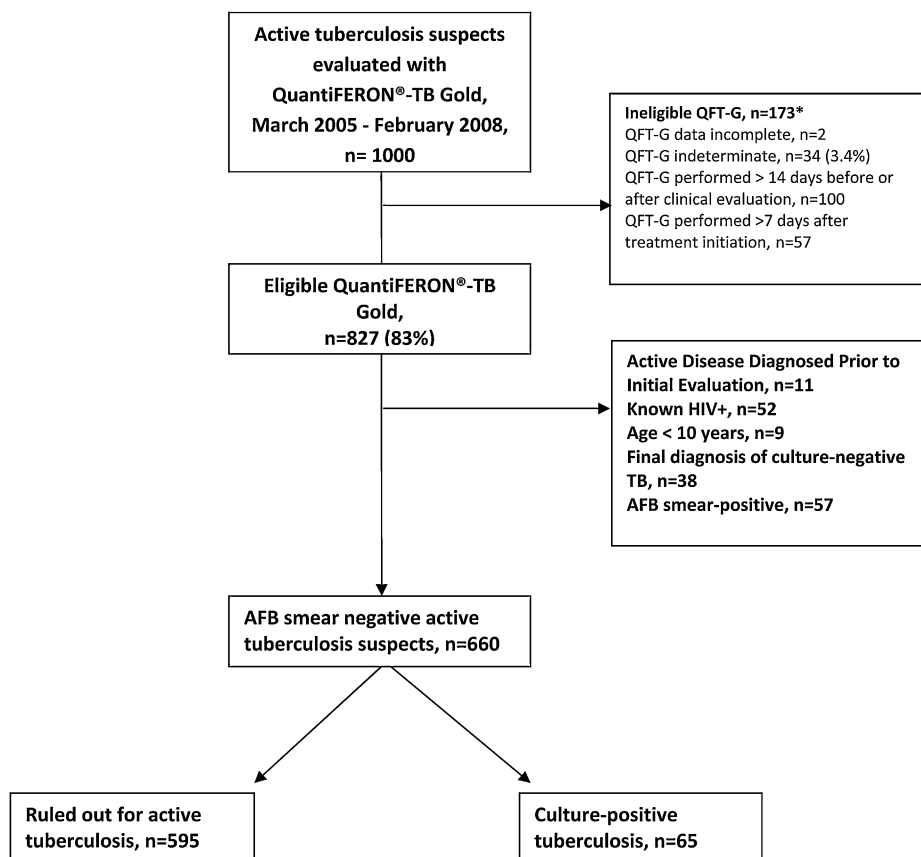
The analysis included the following steps. First, a novel model selection procedure, the deletion/substitution/addition (DSA) algorithm (24), was used to select the optimal prediction model for culture-confirmed tuberculosis using standard clinical and demographic variables. Cova-

riates were considered for inclusion in the model based on previous studies of risk factors for active tuberculosis and included the following: age, sex, foreign birth, homelessness, contact with an active tuberculosis case, previous history of active tuberculosis, predisposing medical condition (diabetes mellitus, silicosis, cancer, or condition requiring use of immunosuppressive medications), clinical symptoms of active tuberculosis (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 on the basis of 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- $\gamma$  results were added to the clinical prediction model. Performance of the prediction models with and without quantitative IFN- $\gamma$  results were then compared using receiver-operator characteristic analysis (25) and net reclassification index (NRI) (18). Based on prespecified 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 were obtained using 10-fold cross-validation. Bootstrap confidence intervals for the NRI estimate are reported based on 1,000 resampling iterations.

All *P* values were two-sided with  $\alpha = 0.05$  as the significance level. All analyses were performed using Stata 10 (Stata Corporation,



**Figure 1.** Study Flow Diagram. \*Some patients were excluded for more than one reason.

College Station, TX) and R, version 2.8.1 (R Project for Statistical Computing, <http://cran.r-project.org>).

## RESULTS

Of 1,000 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- $\gamma$  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 noncases 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- $\gamma$  level was significantly higher in patients with tuberculosis compared with those without (1.1 IU/ml vs. 0.37 IU/ml;  $P < 0.001$ ), and higher IFN- $\gamma$  levels were associated with increased odds of active tuberculosis (odds ratio [OR], 1.07; 95% CI, 1.03–1.11) for each doubling of IFN- $\gamma$  level. For example, a patient with a quantitative IFN- $\gamma$  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. Eighty-five percent of cases had quantitative

IFN- $\gamma$  results in the upper three quintiles of IFN- $\gamma$  concentration ( $\geq 0.23$  IU/ml), whereas 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 before 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 before 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- $\gamma$ Results and Risk Reclassification

The addition of quantitative IFN- $\gamma$  results to the baseline prediction model, including demographic and clinical predictors, significantly improved model accuracy (area under the curve [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 appropriately reclassified into higher- or lower-

**TABLE 1. DESCRIPTION OF ACID FAST BACILLI SMEAR-NEGATIVE ACTIVE TUBERCULOSIS SUSPECTS**

Characteristic	Total	Noncases (n = 595)	Culture-Positive Cases (n = 65)	P Value
Age, years, 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- $\gamma$ concentration, IU/ml, median (IQR)	0.46 (0.06–2.42)	0.37 (0.05–2.38)	1.1 (0.36–3.77)	<0.01
Quintiles of IFN- $\gamma$ 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 to active TB case, %	2	2	6	0.03
Predisposing medical condition <sup>†</sup> , %	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	

*Definition of abbreviations:* BCG = bacillus Calmette-Guérin vaccine; CXR = chest radiograph; IQR = Interquartile range; QFT-G = QuantiFERON-TB Gold; TB = tuberculosis.

Values are percentages unless otherwise stated.

\* Positive at the manufacturer-recommended cut point of 0.35 IU/ml.

<sup>†</sup> Predisposing medical condition: diabetes mellitus, silicosis, cancer, or condition requiring use of immunosuppressive medications.

TABLE 2. COEFFICIENTS AND SUMMARY STATISTICS FOR PREDICTION MODELS

	Baseline Clinical Prediction Model	Baseline Prediction Model with IFN- $\gamma$ Results	Baseline Prediction Model with Clinician Suspicion	Baseline Prediction Model with Clinician Suspicion and IFN- $\gamma$ 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	0.0 23
US birth <sup>†</sup>	1.80	2.85	2.01	2.95
Foreign birth, $\leq 2$ yr in US <sup>†</sup>	1.41	1.58	2.33	2.45
Foreign born, 3–12 yr in US <sup>†</sup>	2.71	3.37	2.09	2.65
Contact to active case	2.43	2.11	3.69	3.09
High clinical suspicion <sup>‡</sup>			19.43	19.31
Intermediate clinical suspicion <sup>‡</sup>			5.53	4.83
Quantitative IFN- $\gamma$ result (effect size per each doubling, IU/ml)		1.07		1.07
AIC	400	374	346	323
AUC	0.71 (0.64–0.77)	0.78 <sup>§</sup> (0.73–0.84)	0.82 (0.77–0.88)	0.86 <sup>  </sup> (0.81–0.91)

*Definition of abbreviations:* AIC = Akaike information criterion, a measure of the goodness of fit of a statistical model with lower values indicating better fit; 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; CXR = chest radiograph.

\* Reference category: inactive disease or normal CXR.

<sup>†</sup> Reference category: foreign born, >12 years in US.

<sup>‡</sup> Reference category: low clinical suspicion.

<sup>§</sup> Significant difference ( $P < 0.001$ ) between this model and previous model without quantitative IFN- $\gamma$  results.

<sup>||</sup> Significant difference ( $P = 0.02$ ) between this model and previous model without quantitative IFN- $\gamma$  results.

risk categories (Table 3). In comparison to the clinical model alone, both case reclassification (11 more cases classified as high risk and 1 fewer as low risk) and noncase reclassification (88 more noncases designated as low risk and no more noncases classified as high risk) were improved. Results were similar when alternate thresholds were used to define risk categories (see Table E1 in the online supplement). Findings on chest radiograph remained the strongest predictor of active tuberculosis.

### Secondary Analysis

We performed a secondary analysis to determine whether quantitative IFN- $\gamma$  levels improved risk reclassification beyond a prediction model that includes clinician suspicion. First, we evaluated whether a similar benefit in risk reclassification occurred when clinician suspicion, rather than quantitative IFN- $\gamma$  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, the addition of quantitative IFN- $\gamma$  results to this expanded model, including clinician suspicion, significantly increased accuracy (AUC, 0.82; 95% CI, 0.77–0.88 vs. AUC, 0.86; 95% CI, 0.81–0.91;  $P = 0.02$ ), but not net reclassification index (NRI, 4%; 95% CI, –0.07 to 0.22;  $P = 0.14$ ). Improved prediction among tuberculosis cases was outweighed by worse performance among noncases (Table 4). The addition of QFT-G results at the manufacturer-recommended cut point of 0.35 IU/ml in place of quantitative IFN- $\gamma$  levels did not materially affect results obtained in either the primary or secondary analysis. To further explore performance in cases and noncases, we examined individual patients' risk before and after quantitative IFN- $\gamma$  level was added to the model. The majority of culture-proven cases showed an appropriate increase in predicted risk with addition of quantitative IFN- $\gamma$  results (Figure 2A). However, both decreased (appropriate) and increased (inappropriate) risk prediction was common among noncases (Figure 2B).

TABLE 3. RISK RECLASSIFICATION FOLLOWING INCORPORATION OF IFN- $\gamma$  RESULTS. COMPARISON TO BASELINE CLINICAL PREDICTION MODEL

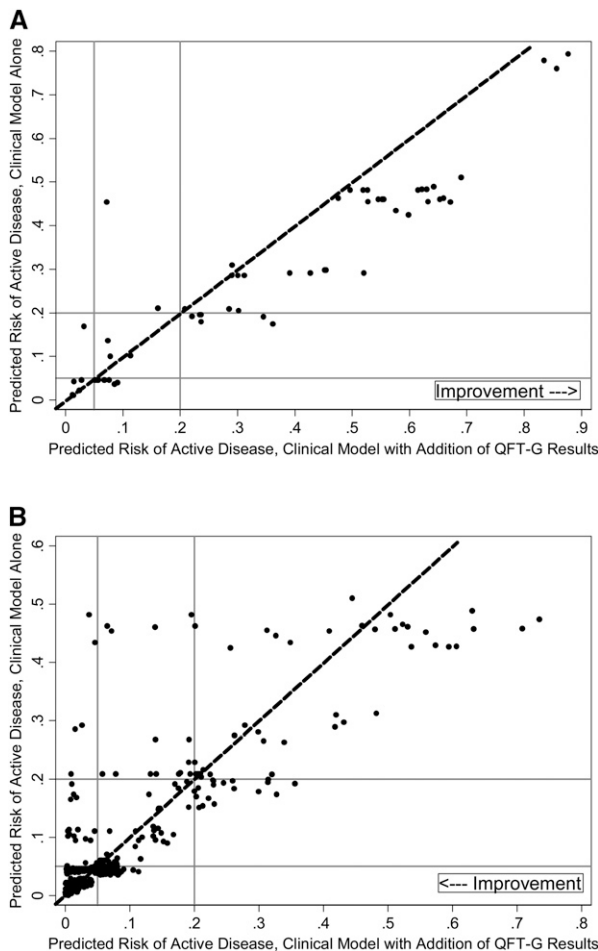
Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN- $\gamma$ Results				Percent Appropriately Reclassified
	$\leq 5\%$ risk	5–20% risk	>20% risk	Total No.	
In 65 patients who developed culture-positive disease					
$\leq 5\%$ risk	7	3	0	10	30
5–20% risk	1	27	12	40	28
>20% risk	1	0	14	15	7
Total No.	9	30	26	65	
In 595 patients who ruled out for active tuberculosis					
$\leq 5\%$ 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% ( $P < 0.001$ ). Reclassification among patients who developed culture-positive disease = 20% ( $P < 0.01$ ); reclassification among patients who ruled out for active tuberculosis = 11.9% ( $P < 0.001$ ).

**TABLE 4. RISK RECLASSIFICATION FOLLOWING INCORPORATION OF IFN- $\gamma$  RESULTS. COMPARISON TO EXPANDED CLINICAL PREDICTION MODEL**

Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN- $\gamma$ Results			Total No.	Percent Appropriately Reclassified
	$\leq 5\%$ risk	5–20% risk	$>20\%$ risk		
In 65 patients who developed culture-positive disease					
$\leq 5\%$ 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					
$\leq 5\%$ 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% ( $P = 0.31$ ). Reclassification among patients who developed culture-positive disease = 18.5% ( $P < 0.01$ ); reclassification among patients who ruled out for active tuberculosis = –14.8% ( $P = 1$ ).



**Figure 2.** (A) Changes in predicted risk of active disease following incorporation of quantitative IFN- $\gamma$  results, cases. (B) Changes in predicted risk of active disease following incorporation of quantitative IFN- $\gamma$  results, noncases. Dashed diagonal lines represent no change in risk prediction with addition of quantitative IFN- $\gamma$  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 noncases. Horizontal and vertical solid lines indicate lower (5%) and upper (20%) risk cut points.

## DISCUSSION

In this study, we found that quantitative IFN- $\gamma$  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- $\gamma$  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 highly experienced tuberculosis centers in low-incidence settings.

Risk prediction has long been used in the cardiovascular (26, 27) and cancer (28) literature 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, although these measures alone do not describe the predictive accuracy of these assays or the extent to which they improve on readily available clinical information (29). In the absence of an established risk prediction model for AFB smear-negative tuberculosis, we used the DSA routine (24) to identify the optimal prediction model. This state-of-the-art procedure considers nonlinear 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 those 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- $\gamma$  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 weigh the net reclassification results more heavily than improvements in discrimination represented by increases in AUC. Although 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 use for evaluating risk prediction models (29, 34).

The changes in predicted risk of active tuberculosis following consideration of quantitative IFN- $\gamma$  results were not uniform. Among intermediate and high-risk patients who eventually ruled out for active tuberculosis, the addition of quantitative IFN- $\gamma$  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 who were eventually ruled out for tuberculosis were inappropriately reclassified as intermediate risk after consideration of quantitative IFN- $\gamma$  results. The possibility that quantitative IFN- $\gamma$  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 on both the base prediction model and choice of risk categories. We recognize that addition of IFN- $\gamma$  to suboptimal base models could produce large improvements in both discrimination and risk reclassification. We used 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 prespecified, 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 IFN- $\gamma$  results are not reported by the SFDPH laboratory. The dramatic improvement in model performance with the addition of clinician suspicion, however, indicates that crucial information is obtained in the workup process beyond our measured covariates. Prospective studies should attempt to better define these factors. Third, the test characteristics of QFT-G In-Tube (QFT-G-IT), the most recent generation of this assay, may differ from QFT-G as used in this study. Lastly, our analysis is most relevant to tuberculosis referral centers with experienced clinicians operating in low-incidence settings.

In conclusion, quantitative IFN- $\gamma$  results obtained from QFT-G improved clinical evaluation of tuberculosis suspects compared with 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.

**Conflict of Interest Statement:** J.Z.M. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. A.C. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. E.V. has received consultancy fees from Kronos Longevity Research Institute, NPS Pharmaceuticals, Tethys Biosciences (\$1001–\$5000), and Zelos Therapeutics (\$5001–\$10,000). He has also received royalties from Springer-Verlag (\$5001–\$10,000). C.H. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.G. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. P.C.H. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. L.M.K. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. P.N. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

**Acknowledgment:** The authors thank the staff at the San Francisco Department of Public Health, Tuberculosis Control Section.

## References

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Mazurek GH, Jereb J, Lobue J, Pademarco MF, Metchock B, Vernon A; Division of Tuberculosis Elimination, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention (CDC). Guidelines for using the QuantiFERON-TB Gold test for detecting *Mycobacterium tuberculosis* infection, United States. *MMWR Recomm Rep* 2005 Dec 16;54(RR-15):49–55.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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 TB net-study. *PLoS One* 2008;3:e3417.
- 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 207–212.
- Metcalfe JZ, Cattamanchi A, 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.
- American Thoracic Association and Centers for Disease Control. Targeted tuberculin testing and treatment of latent tuberculosis infection. Official Statement of the ATS and CDC, July 1999. *Am J Respir Crit Care Med* 2000;161:S221–S247.

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, Hesselning AH, Beyers N. Adult-type pulmonary tuberculosis in children 10–14 years of age. *Pediatr Infect Dis J* 2005; 24:743–744.
23. QuantiFERON-TB Gold: the whole blood IFN-gamma test measuring responses to ESAT-6 & CFP-10 peptide antigens [package insert]. Carnegie, Australia: Cellestis Ltd.; 2007. Doc No. 05980000D. Accessed 2009 Oct 28. Available from: <http://www.cellestis.com/IRM/Company/ShowPage.aspx?CPID=1247>
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. Diverse Populations Collaborative Group. 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.