

# UCSF

## UC San Francisco Previously Published Works

### Title

Test Variability of the QuantiFERON-TB Gold In-Tube Assay in Clinical Practice

### Permalink

<https://escholarship.org/uc/item/6vt758b8>

### Journal

American Journal of Respiratory and Critical Care Medicine, 187(2)

### ISSN

1073-449X

### Authors

Metcalfe, John Z  
Cattamanchi, Adithya  
McCulloch, Charles E  
et al.

### Publication Date

2013-01-15

### DOI

10.1164/rccm.201203-0430oc

Peer reviewed

# Test Variability of the QuantiFERON-TB Gold In-Tube Assay in Clinical Practice

John Z. Metcalfe<sup>1</sup>, Adithya Cattamanchi<sup>1</sup>, Charles E. McCulloch<sup>2</sup>, Justin D. Lew<sup>3</sup>, Ngan P. Ha<sup>3</sup>, and Edward A. Graviss<sup>3</sup>

<sup>1</sup>Division of Pulmonary and Critical Care Medicine, Francis J. Curry International Tuberculosis Center, San Francisco General Hospital, San Francisco, California; <sup>2</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California; and <sup>3</sup>Department of Pathology and Genomic Medicine, The Methodist Hospital Research Institute, Houston, Texas

**Rationale:** Although IFN- $\gamma$  release assays (IGRAs) are widely used to screen for *Mycobacterium 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) between August 1, 2010 and July 31, 2011 for latent tuberculosis (TB) infection screening with an IGRA (QuantiFERON-TB Gold In-Tube; Cellestis, Carnegie, Australia). 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, 1,086 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 (8%) of 366 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% confidence interval, 0.79–0.90), the normal expected range of within-subject variability in TB response on retesting included differences of  $\pm 0.60$  IU/ml for all individuals (coefficient of variation, 14%), and  $\pm 0.24$  IU/ml (coefficient of variation, 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.

**Keywords:** interferon- $\gamma$  release assay; QuantiFERON; repeatability; imprecision

IFN- $\gamma$  release assays (IGRAs) are *in vitro* immunodiagnostic tests that measure effector T cell-mediated IFN- $\gamma$  response to synthetic *Mycobacterium tuberculosis*-specific polypeptides. The QuantiFERON-TB Gold In-Tube (QFT-GIT; Cellestis, Carnegie,

(Received in original form March 8, 2012; accepted in final form October 9, 2012)

Supported in part by the National Institutes of Health (K23 AI094251 to J.Z.M. and K23 HL094141 to A.C.) and the Robert Wood Johnson Foundation (AMFDP Medical Faculty Development Award to J.Z.M.).

**Author Contributions:** Study conception, J.Z.M. and E.A.G. Initial manuscript draft, J.Z.M., E.A.G., and C.E.M. Data analysis, J.Z.M. and C.E.M. Editing and revisions of manuscript, J.Z.M., E.A.G., A.C., and C.E.M. Laboratory procedures, J.D.L., N.P.H., and E.A.G.

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, San Francisco General Hospital, 1001 Potrero Avenue, Room 5K1, San Francisco, CA 94110-0111. E-mail: john.metcalfe@ucsf.edu

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

Am J Respir Crit Care Med Vol 187, Iss. 2, pp 206–211, Jan 15, 2013

Copyright © 2013 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201203-0430OC on October 26, 2012

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

## AT A GLANCE COMMENTARY

### Scientific Knowledge on the Subject

Although IFN- $\gamma$  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 tuberculosis response measured by QuantiFERON-TB Gold In-Tube (Cellestis, Carnegie, Australia) on retesting of the same patient sample. Variability within individuals included differences up to 0.24 IU/ml, in either direction, when the initial tuberculosis response was between 0.25 and 0.80 IU/ml. Positive QuantiFERON-TB Gold In-Tube test results less than 0.59 IU/ml should be interpreted cautiously.

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 impacted 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 Table 1), are required for CE (Conformité Européenne) Marking in the European Union and for premarket approval of *in vitro* diagnostic tests by the US Food and Drug Administration (FDA) (4).

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%) (5). Multiple independent studies have assessed longitudinal changes in tuberculosis (TB) response (6–20), although few investigators have examined the repeatability of commercial IGRAs (7, 9, 10, 18). These reports support a greater amount of test variability than that reported by the manufacturer, although interpretation has been limited by sample size and heterogeneous statistical methods, terminology, and epidemiologic settings.

TABLE 1. DEFINITIONS

Term	Definition
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 (51).
Reproducibility	The precision of a test when replicated under different conditions (e.g., different laboratory, operator, apparatus, unspecified time interval) (51).
TB response	The amount of IFN- $\gamma$ released in response to <i>M. tuberculosis</i> -specific antigens (ESAT-6, CFP-10, TB7.7), calculated as the difference in IFN- $\gamma$ concentration in plasma from blood stimulated with antigen minus the IFN- $\gamma$ 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 $\pm$ 1.96 repeatability SD. This range about an individual's true value (the value in absence of repeatability variability) would include 95% of repeat measurements for that individual (52).
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]).

Definition of abbreviation: TB = tuberculosis.

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 (21).

## 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, TX) 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. The quality control algorithm excluded individuals with both low TB response ( $< 0.25$  IU/ml) and low nil control ( $< 0.10$  IU/ml) 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 deidentified, clinical database; information on nationality, TST status, or history of previous TB disease was not available.

Two technicians from a Clinical Laboratory Improvement Amendments certified (22) laboratory accustomed to IGRA research (23–26) performed all QFT-GIT assays according to the manufacturer's instructions using identical instrument settings (laboratory testing details are provided in the 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 [ $\geq 0.35$  IU/ml], or vice versa). Initial QFT-GIT assays were classified as indeterminate according to manufacturer instructions (27) and repeated once. 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 after 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 interrater agreement (28). 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 supplement) (29). Subjects with very low TB response (TB response less than  $-0.35$  IU/ml and less than  $-0.5$  times nil control) were excluded (30). IFN- $\gamma$  concentrations greater than 10 IU/ml may be unreliable because of 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- $\gamma$  values. Such models allow global estimation of within-person SD optimally weighted for the correlation structure of the repeated measurements (31, 32). Assuming that 95% of measurements are located within  $\pm 1.96$  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 when the initial TB response was in a borderline range (0.25–0.80 IU/ml), and when subjects with TB response greater than 10 IU/ml were excluded, rather than truncated.

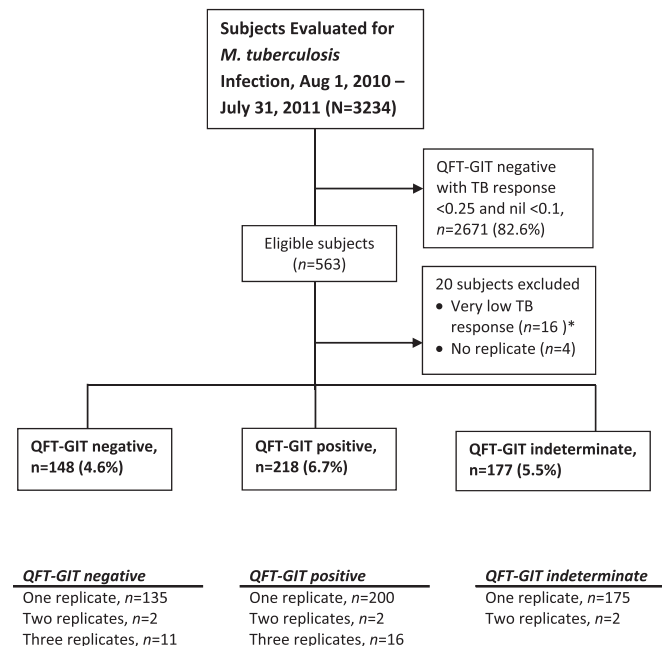
All *P* values were two-sided with  $\alpha = 0.05$  as the significance level. Data analysis was performed using Stata 12.1 (Stata Corporation, College Station, TX).

## RESULTS

Between August 1, 2010 and July 31, 2011, 3,234 individuals were screened for *M. tuberculosis* infection using QFT-GIT. Of these, 2,819 (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, 2,671 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 those with TB response less than 0.25 IU/ml and nil control greater than or equal to 0.1 IU/ml, and 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 2.

### 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% confidence interval, 0.79–0.90) (Table 3). Bland-Altman plots of test-retest differences



**Figure 1.** Study flow diagram. Positive, negative, and indeterminate tests were classified according to manufacturer's recommended criteria (27). Analyzed QuantiFERON Gold-In Tube (QFT-GIT) negative test results were negative less than 0.25 IU/ml with nil control greater than or equal to 0.10 IU/ml, or 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 less than  $-0.35$  IU/ml and less than  $-0.5$  times nil control (30). TB = tuberculosis.

across the full and borderline range of TB response are presented as Figures E1A and E1B in the online supplement.

### Retesting Results

On retesting, 13 (9%) of 148 negative results converted to positive and 15 (7%) of 218 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 nonborderline measurements for both). Eighty-six percent (24 of 28) of all individuals who converted or reverted had an initial TB response within the borderline range (Table 4). On retesting, 11 (6%) of 177 indeterminate tests were reread as negative, and two positive results (2 of 218;  $<1\%$ ) were reread as indeterminate.

### Test Variability

A total of 1,086 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 greater than 1 IU/ml in 14 (4%) of 366 individuals.

The normal expected range of variability for an individual patient included differences of  $\pm 0.60$  IU/ml (CV 14%) for all individuals, and  $\pm 0.24$  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

**TABLE 2. CHARACTERISTICS OF THE STUDY POPULATION**

Characteristic	Total (n = 543)
Age, yr, mean ( $\pm$ SD)	43 ( $\pm$ 18)
Male, %	66
Origin of referral, %*	
Petrochemical corporation	74
Hospital/clinician	26
Race or ethnicity, %	
White	41
Black	9
Hispanic	11
Asian	10
American Native	1
Unknown	28
Number of replicates obtained per subject, % <sup>†</sup>	
1	92
2	1
3	7

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

\* Origin of referral includes determinate tests only.

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

sensitivity analysis excluding subjects with TB response greater than 10 IU/ml, and among subjects with indeterminate test results (see METHODS and online 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 on 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 less than or equal to 6 mm may occur because of host biologic variation, or differences in test administration or measurement (33–36). These normal limits of test variability contribute to current thresholds ( $\geq 10$  mm) for skin test conversion (37). Although IGRAs may be more specific than the TST and initially identify fewer persons requiring preventive therapy in some settings (38), subsequent conversions and reversions may occur with greater frequency (7, 11, 39, 40). The reasons for this remain speculative, with possible

**TABLE 3. 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 abbreviation: QFT-GIT = QuantiFERON Gold-In Tube.

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

**TABLE 4. REPEATABILITY OF QFT-GIT TEST RESULTS STRATIFIED BY QUANTITATIVE TB RESPONSE**

Baseline TB Response (IU/ml)	Total Subjects	Conversion, n (%) <sup>*</sup>	Reversion, n (%) <sup>*</sup>
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	76	—	2/76 (3)
3.1–9.92	76	—	1/76 (1)

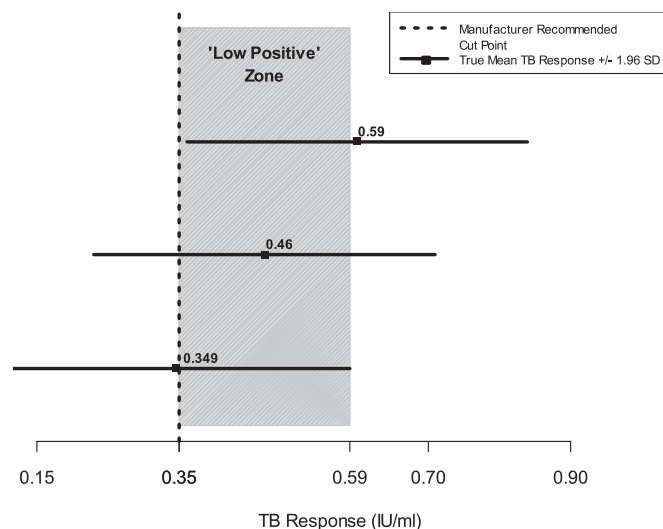
*Definition of abbreviations:* QFT-GIT = QuantiFERON Gold-In Tube; TB = tuberculosis.

Initial indeterminate test results were not included.

<sup>\*</sup>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).

explanations including greater variability in immunologic reactivity of effector versus central memory T cells, or a superior 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  $\pm 0.60$  IU/ml on retesting. For



**Figure 2.** Normal expected variability of borderline tuberculosis (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 ( $\pm 1.96$  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  $\pm 0.24$  IU/ml, 95% of subjects demonstrate variability between 0.11 and 0.59 IU/ml. Thus, the low positive zone (gray shading) is defined as the interval (0.35–0.59 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 is 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 QuantiFERON Gold-In Tube (QFT-GIT) assays was assessed using a linear mixed effects model fit to the numerical IFN- $\gamma$  values. Borderline TB response was defined as IFN- $\gamma$  concentration 0.25–0.80 IU/ml; the manufacturer-recommended cut point (0.35 IU/ml) is demonstrated as a dashed line.

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% ( $n = 43$  of 218) of positive results fell within this range, and 23% ( $n = 10$  of 43) of these individuals reverted to negative. Weighing 4 to 9 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 (41), some centers in the United States have already instituted such policies (42, 43). 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 (44), although establishing the evidence base for such thresholds will be challenging (2).

Our findings build on previous literature documenting variable amounts of QFT-GIT imprecision and calls for borderline zones of various ranges (6, 8, 16–18, 45). 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% on retesting. Repeatability data, usually from high TB incidence regions, have been less often reported. Further, interpretation of these data has been challenging because of the differing methods used to assess variability. Because this variability has traditionally been determined by dividing the pooled SD by the overall mean TB response (i.e., use of CV), persons with positive test results 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 (5); low CV percent despite relative large differences in magnitude of TB response among high positive tests is a known limitation of this measure of repeatability (46). Independent investigators have reported results from a variety of repeatability methods. In India, Veerapathran and coworkers (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 and coworkers (9) (South Africa) repeated readings of the same ELISA plate over a 2-hour period and reported no significant variability using analysis of variance techniques. Both Detjen and coworkers (10) (South Africa) and Ringshausen and coworkers (18) (Germany) reported a high intraclass correlation coefficient (0.991 and 0.995, respectively) suggesting excellent repeatability. Although the intraclass correlation coefficient is an important measure of repeatability, it does not provide an intuitive answer as to how much a patient’s initial result might change on retesting. Given that total imprecision is always greater than within-subject imprecision (47), our results may be considered conservative in the context of serial testing programs where extraneous biologic variability is also a consideration. Of note, QFT-GIT variability is high compared with new molecular TB tests, such as Xpert MTB/RIF (Cepheid, Inc., Sunnyvale, CA) quantitative polymerase chain reaction (CV 6%) (48), or commonly used serologic assays (17).

The 2010 Centers for Disease Control and Prevention Guidelines on IGRAs recommend that clinical laboratories report quantitative in addition to qualitative (i.e., positive or negative) test

results (1). Although some utility in predicting diagnosis (49) and progression to active TB (50) may exist, in low TB incidence settings reporting of quantitative test results is 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 most 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 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 performed 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  $\pm 0.60$  IU/ml (CV 14%) overall, and  $\pm 0.24$  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.

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

## References

- 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 Morb Mortal Wkly Rep Recomm Rep* 2010;59:1–25.
- 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.
- Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, Poole F, Smith PG, Sriram N, Wongsrichanalai C, Linke R, et al. Evaluation of diagnostic tests for infectious diseases: general principles. *Nat Rev Microbiol* 2006;4:S20–S32.
- Derion T. Considerations for the planning and conduct of reproducibility studies of in vitro diagnostic tests for infectious agents. *Biotechnol Annu Rev* 2003;9:249–258.
- Cellestis. Validation report QuantiFERON-TB Gold In-Tube: reproducibility study. 2006 [accessed 2011 Dec 16]. Available from: <http://www.cellestis.com/IRM/Company/ShowPage.aspx/PDFs/1359-10000000/ValidationReportQFTInTubeReproducibilityStudy>
- Detjen AK, Loebenberg L, Grewal HM, Stanley K, Gutschmidt A, Kruger C, Du Plessis N, Kidd M, Beyers N, Walzl G, et al. Short-term reproducibility of a commercial interferon gamma release assay. *Clin Vaccine Immunol* 2009;16:1170–1175.
- 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. *Am J Respir Crit Care Med* 2006;174:349–355.
- Veerapathran A, Joshi R, Goswami K, Dogra S, Moodie EE, Reddy MV, Kalantri S, Schwartzman K, Behr MA, Menzies D, et al. T-cell assays for tuberculosis infection: deriving cut-offs for conversions using reproducibility data. *PloS One* 2008;3:e1850.
- van Zyl-Smit RN, Pai M, Praph K, Meldau R, Kieck J, Juritz J, Badri M, Zumla A, Sechi LA, Bateman ED, et al. Within-subject variability and boosting of T-cell interferon-gamma responses after tuberculin skin testing. *Am J Respir Crit Care Med* 2009;180:49–58.
- Herrmann JL, Belloy M, Porcher R, Simonney N, Aboutaam R, Lebourgeois M, Gaudelus J, De Losangeles L, Chadelat K, Scheinmann P, et al. Temporal dynamics of interferon gamma responses in children evaluated for tuberculosis. *PLoS ONE* 2009;4:e4130.
- 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. *Infect Control Hosp Epidemiol* 2009;30:870–875.
- 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. *Infect Control Hosp Epidemiol* 2009;30:1218–1222.
- Perry S, Sanchez L, Yang S, Agarwal Z, Hurst P, Parsonnet J. Reproducibility of QuantiFERON-TB Gold In-Tube assay. *Clin Vaccine Immunol* 2008;15:425–432.
- Pollock NR, Campos-Neto A, Kashino S, Napolitano D, Behar SM, Shin D, Sloutsky A, Joshi S, Guillet J, Wong M, et al. Discordant QuantiFERON-TB Gold Test results among US healthcare workers with increased risk of latent tuberculosis infection: a problem or solution? *Infect Control Hosp Epidemiol* 2008;29:878–886.
- 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. *Epidemiol Infect* 2009;137:1691–1698.
- 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.
- 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*. *Clin Vaccine Immunol* 2010;17:596–601.
- 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. *Clin Vaccine Immunol* 2011;18:1176–1182.
- 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. *J Immunol Methods* 2004;291:175–183.
- 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 Infect Dis* 2010;10:220.
- 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. Presented at the 3rd Global Symposium on IGRAs 2012, January 15, 2012, Waikoloa, Hawaii.
- Clinical and Laboratory Standards Institute. Ep21-a, estimation of total analytical error for clinical laboratory methods; approved guideline. Wayne, PA: Clinical and Laboratory Standards Institute; 2003.
- 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. *Clin Vaccine Immunol* 2006;13:53–58.
- 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. *Clin Vaccine Immunol* 2007;14:714–719.
- Grimes CZ, Hwang LY, Williams ML, Austin CM, Graviss EA. Tuberculosis infection in drug users: interferon-gamma release assay performance. *Int J Tuberc Lung Dis* 2007;11:1183–1189.
- Cruz AT, Geltmeyer 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–e38.
- Cellestis Limited. (2010). QuantiFERON-TB Gold In-Tube: package insert. Carnegie, Victoria, Australia [accessed 2012 May 4]. Available from: [Http://www.Cellestis.Com/irm/content/pdf/quantiferon%20us%20verg\\_jan2010%20no%20trims.Pdf](http://www.Cellestis.Com/irm/content/pdf/quantiferon%20us%20verg_jan2010%20no%20trims.Pdf)
- Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: Wiley; 1981.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
- Powell RD III, 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.
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–974.

32. Diggle PJ. An approach to the analysis of repeated measurements. *Biometrics* 1988;44:959–971.
33. Furcolow ML, Watson KA, Charron T, Lowe J. A comparison of the tine and mono-vacc tests with the intradermal tuberculin test. *Am Rev Respir Dis* 1967;96:1009–1027.
34. Erdtmann FJ, Dixon KE, Llewellyn CH. Skin testing for tuberculosis: antigen and observer variability. *JAMA* 1974;228:479–481.
35. Bearman JE, Kleinman H, Glycer VV, Lacroix OM. A study of variability in tuberculin test reading. *Am Rev Respir Dis* 1964;90:913–919.
36. Chaparas SD, Vandiviere HM, Melvin I, Koch G, Becker C. Tuberculin test: variability with the Mantoux procedure. *Am Rev Respir Dis* 1985;132:175–177.
37. American Thoracic Society. Targeted tuberculin testing and treatment of latent tuberculosis infection. *MMWR Morb Mortal Wkly Rep Recomm Rep* 2000;49:1–51.
38. Zwerling A, van den Hof S, Scholten J, Cobelens F, Menzies D, Pai M. Interferon-gamma release assays for tuberculosis screening of health-care workers: a systematic review. *Thorax* 2012;67:62–70.
39. Zwerling A, Cloutier-Ladurantaye J, Pietrangelo F. Conversions and reversions in health care workers in Montreal, Canada using QuantiFERON-TB-Gold In-Tube [abstract]. *Am J Respir Crit Care Med* 2009;179:A1012.
40. Belknap RWK, Teeter L. Interferon-gamma release assays (IGRAS) in serial testing for latent tuberculosis infection in U.S. health care workers [abstract]. *Am J Respir Crit Care Med* 2010;181:A2263.
41. 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. *Am J Respir Crit Care Med* 2012;185:427–434.
42. 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;142:55–62.
43. Gray J, Reeves 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. *Clin Infect Dis* 2012;54:e20–23.
44. Menzies D. Interpretation of repeated tuberculin tests: boosting, conversion, and reversion. *Am J Respir Crit Care Med* 1999;159:15–21.
45. Pai M, Joshi R, Dogra S, Zwerling AA, Gajalakshmi D, Goswami K, Reddy MV, Kalantri A, Hill PC, Menzies D, *et al.* T-cell assay conversions and reversions among household contacts of tuberculosis patients in rural India. *Int J Tuberc Lung Dis* 2009;13:84–92.
46. Quan H, Shih WJ. Assessing reproducibility by the within-subject coefficient of variation with random effects models. *Biometrics* 1996;52:1195–1203.
47. Krouwer JS, Rabinowitz R. How to improve estimates of imprecision. *Clin Chem* 1984;30:290–292.
48. van Zyl-Smit RN, Binder A, Meldau R, Mishra H, Semple PL, Theron G, Peter J, Whitelaw A, Sharma SK, Warren R, *et al.* Comparison of quantitative techniques including Xpert MTB/RIF to evaluate mycobacterial burden. *PLoS ONE* 2011;6:e28815.
49. 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. *Am J Respir Crit Care Med* 2010;181:87–93.
50. 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. *Am J Respir Crit Care Med* 2011;183:88–95.
51. American Society for Testing and Materials (ASTM) International. What are repeatability and reproducibility? March/April 2009 [accessed 2012 July 25]. Available from: [http://www.Astm.Org/SNEWS/MA\\_2009/datapoints\\_ma09.html](http://www.Astm.Org/SNEWS/MA_2009/datapoints_ma09.html)
52. American Society for Testing and Materials. E456 standard terminology relating to quality and statistics, 2012 [accessed 2012 July 25]. Available from: <http://www.Astm.Org/Standards/E456.htm>