Reproducibility of QuantiFERON-TB Gold In-Tube Assay

Sharon Perry, Luz Sanchez, Shufang Yang, Zubin Agarwal, Philip Hurst, and Julie Parsonnet

Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford, California, and Santa Clara Valley Health and Hospitals System, San Jose, California

Received 3 September 2007/Returned for modification 7 November 2007/Accepted 6 January 2008

Studies are needed to characterize the reproducibility of QuantiFERON-TB Gold (QFT-G) for targeted U.S. screening populations. Members of northern California households were tested with the QFT-G in-tube assay (QFT-G-IT) at two home visits 3 months apart. Reproducibility and agreement with the tuberculin skin test (TST) were assessed. Monte Carlo simulation was used to evaluate the role of test-related error. Of 63 individuals (49 adults and 14 children) completing QFT-G-IT at both time points, 79% were foreign-born (98% from Latin America) and 68% reported Mycobacterium bovis BCG vaccination. At the baseline visit, 23 (37%) were TST positive and 15 (24%) were QFT-G-IT positive ($\chi^2 = 0.48 \pm 0.11$). At 3 months, 3/48 (6.3%; 95% confidence interval [95CI], 2 to 17) of those initially QFT-G-IT negative converted, and 5/15 (33%; 95CI, 15 to 58) of those initially QFT-G-IT positive reverted. Among the 8 individuals with inconsistent QFT-G-IT results, the maximum gamma interferon response at either visit was 0.68 IU/ml versus means of 4.99 (± 3.74) and 6.95 (± 5.6) for 10 persistent positives at the first and second visits, respectively. Expected false-reversion and -conversion rates were 32% (90CI, 25 to 39%) and 6.95% (90CI, 4.6 to 9.8%) when the sensitivity and specificity were assumed to average 70% and 98%, respectively. Transient responses to QFT-G-IT are common, and low positive results need to be interpreted with caution. Further studies are needed to characterize the predictive value of the test for U.S. foreign-born and other targeted screening populations.

Foreign-born individuals constitute the fastest growing segment of tuberculosis (TB) cases in the United States. In 2006, TB incidence among the foreign-born in the United States was 21.9/100,000, or 9.5 times that among those born in the United States (8). Community control efforts depend on reliable detection of those who would benefit from treatment of latent disease. Long the bulwark of U.S. surveillance, the tuberculin skin test (TST) has several well-known limitations for screening immigrant populations, including cross-reactivity with nontuberculous mycobacterium (NTM) and Mycobacterium tuberculosis bacillus Calmette-Guérin (BCG) vaccination as well as reduced sensitivity in immunocompromised hosts, active TB patients, children, and the elderly (28).

After more than a century, the development of T-cell-based gamma interferon (IFN-γ) release assays (IGRAs) for the detection of latent TB infection represents a significant advancement in TB diagnostics (36). Advantages of these in vitro blood tests include direct and objective measurement of cell-mediated immune response and the need for only one patient visit. Newer generations of the assays incorporate highly specific M. tuberculosis targets, such as the ESAT-6 (early secretory antigen 6) and CFP-10 (culture filtrate protein 10) complex from region of difference 1 (RD1) of the M. tuberculosis genome (3). Because the RD1 region is absent from most NTM as well as from strains used in BCG vaccines, the tests are considered highly specific for infection with M. tuberculosis (1, 42). One IGRA is now commercially available in the United States: QuantiFERON-TB Gold (QFT-G; Cellestis Ltd., Carnegie, Victoria, Australia), approved by the U.S. Food and Drug Administration in 2005 (26). A recent meta-analysis estimated the pooled average specificity and sensitivity of this assay to be 97.7% and 67%, respectively (29). A simplified QFT-G blood collection system (the QFT-G in-tube assay [QFT-G-IT]) was approved in October 2007 by the U.S. FDA; another IGRA format, the enzyme-linked immunospot (ELISPOT) assay (T-SPOT.TB; Oxford Immunotec, Oxon, United Kingdom) available in Europe, is currently undergoing U.S. FDA review. The Centers for Disease Control and Prevention has recommended use of the QFT-G in “all circumstances” where the TST is used (26).

Despite broad enthusiasm for the tests, their role in screening programs remains to be defined (29). The biological basis for discordance with TST in the assessment of high-risk populations is not understood (21, 34). Studies with children (9) and immune-compromised hosts (5) are also still needed. Although adoption of the QFT-G is being actively investigated by U.S. public health departments (12), only one cross-sectional study has been reported from the United States (39). There are no U.S. studies, and only a handful internationally (16, 33), examining reproducibility for the intended screening populations. In the United States, serial testing studies are needed to characterize utility in screening foreign-born and other targeted high-risk groups, for whom repeat testing is often conducted either intentionally (to assess boosting) or incidentally (immigrants may see a variety of health care providers who each perform a TB test). In the absence of a true “gold standard” for latent TB infection, serial testing studies can also be useful for evaluating discordance with TST (31).

As part of a prospective study of infectious disease transmission in immigrant households in northern California, we tested household members on two occasions 3 months apart in order to evaluate the reproducibility of the QFT-G-IT.
MATERIALS AND METHODS

Design. The study was nested within the Stanford Infection and Family Transmission (SIFT) study, a prospective study of infectious disease transmission in households of Santa Clara County, CA (37). Households with an index case of gastroenteritis were recruited for the SIFT study through cooperating public health clinics and visited in their homes on two occasions 3 months apart. At each visit, all household members, whether reporting symptoms of gastroenteritis or not, were invited to enroll. Beginning in January 2003, SIFT study household participants of >2 years of age were offered testing with the first-generation QFT using purified protein derivative (PPD) antigen (31), and in September 2003, individuals were offered testing with the second-generation QFT-G-IT.

Subjects who consented completed a structured interview, which was followed by whole-blood collection and TST placement. Individuals of <2 years of age, those reporting histories of active TB, immune-modulating conditions (human immunodeficiency virus infection, malignancy, chemotherapy, or pregnancy) or adverse reaction to TST, or those who had received a TST within the previous 5 months were excluded. Within 48 to 72 h, a return home visit was made to read the TST. Approximately 3 months later, a second home visit was scheduled at which blood collection followed by TST testing and a TST reading visit within 48 to 72 h were repeated. Subjects reporting a prior positive TST at the first visit, or who tested positive at the first visit, were offered a choice of repeating the TST; however, all subjects were asked to complete follow-up blood collection for the QFT-G-IT.

The study was approved by the Institutional Review Boards of Stanford University and the Santa Clara Valley Medical System.

Risk factor assessment. A structured interview was administered to collect demographic information, including country of birth, travel to countries of endemicity, and time in the United States. In addition, information was collected on history of BCG vaccination (including visual inspection for scar), TB exposure, social and medical risk factors for TB, prior diagnosis or treatment of TB, and recent gastrointestinal or pulmonary symptoms.

TST. Tuberculin PPD (0.1 ml; Tuberculin Avenis) was applied intradermally at the inner volar surface of the forearm under supervision of a licensed health care professional as required by California law. The size of the induration was recorded within 48 to 72 h, measured across the wide diameter. Per local public health authority guidelines (41), an induration of ≥10 mm was classified as positive. A TST conversion was defined as an increase in induration of at least 6 mm, accompanied by a qualitative change from negative (<10 mm) to positive (≥10 mm) (27). Participants received a written notice of their TST results, and those with a conversion were offered assistance with referral to a primary health care provider.

QFT-G-IT collection. Prior to placement of the TST, 1 ml whole blood was drawn into one (each) of three manufacturer-precoated heparinized tubes containing mitogen (phytohemagglutinin A), TB peptide antigens ESAT-6/CFP-10/ TBT7.7, and a negative control (phosphate-buffered saline with 0.1% thimerosal), respectively. Tubes were transported to the research laboratory within 6 h and incubated at 37°C for 16 to 24 h. Harvested plasma was stored at 2 to 8°C for up to 3 days or frozen at −20°C for up to 8 weeks pending enzyme-linked immunosorbent assay (ELISA) and at −80°C for periods longer than 2 months. ELISA was used to detect levels of IFN-γ in response to TB antigens according to manufacturer’s instructions and quality control criteria. Absorbance values were converted to international units of concentration using manufacturer’s standard curve methodology. Results were interpreted according to manufacturer’s guidelines: an IFN concentration of ≥0.35 IU/ml over the nil concentration was classified as positive. Nil concentrations of ≥0.8 IU/ml and mitogen differences of <0.5 IU/ml were considered indeterminate. Values greater than 10 IU/ml were assigned a value of 10 IU, the upper limit of the standard curve. All tests were conducted by the same operator, who was blinded to subject characteristics. Conversions and reversions were defined qualitatively as a change from negative (<0.35 IU/ml above background) to positive (≥0.35 IU/ml above background) and as a change from positive to negative between visits, respectively.

Analysis. The McNemar test and the κ coefficient were used to characterize baseline discordance with TST and the reproducibility of QFT-G-IT, respectively. QFT-G-IT conversion and reversion rates and 95% confidence intervals (95%CI) were computed. We also evaluated variation in absolute IFN response levels, using Student’s t test for group comparisons. Fisher’s exact test was used to compare categorical outcomes between groups. Statistical analysis was conducted using SAS 9.3.

Monte Carlo simulation (CrystalBall; Decisioneering, Denver, CO) was used to predict apparent conversion and reversion rates based on assumptions regarding accuracy of testing (38). A base model was constructed using observed test results for QFT at the baseline visit. After one additional testing cycle, the expected rate of conversion in initial negatives can be predicted as follows: Expected conversion % = (1 − PV−)×(1 − Se) + (1 − Sp), where PV− and PV+ are the predictive value negative and positive of the baseline test and Se and Sp are the constant specificity and sensitivity of the test, respectively.

Similarly, the expected reversion rate in initial positives can be predicted as follows:

Expected reversion % = PV+ + (1 − Se) × (1 − PV−) × (1 − Sp).

Prevalence, sensitivity, and specificity were modeled with beta distributions (0.1) using credible estimates from a recent meta-analysis (29). The analysis assumes that repeated test results are independent (conditional only on the true infection state) and that the “true” rates of conversion and reversion are zero over 3 months (18). A three-way sensitivity analysis, varying sensitivity and specificity over the apparent prevalence of infection, was used to evaluate uncertainty in the model’s assumptions. Each Monte Carlo simulation was seeded with the baseline data and sampled for 1,000 trials, using a fixed random number seed.

RESULTS

Baseline characteristics. Between September 2003 and April 2004, 52 SIFT study households were invited to be tested by QFT-G-IT, of which 39 agreed (75%) and 13 declined (25%). Participating households were similar to other SIFT study households in terms of size (median, five persons), education of adults (median, 12 years), and primary language (85% Spanish speaking).

Within the 39 participating households, 88 SIFT study participants over the age of 2 were enrolled. Of these, 87 had baseline QFT-G-IT results (1 individual had an indeterminate QFT-G-IT result). Of 87 subjects with baseline results, 65 (75%) completed blood collection at the 3-month follow-up visit, and 22 (25%) were not available. Two baseline-TST-negative subjects did not complete the follow-up TST. Thus, a total of 63 subjects (49 adults and 14 children ranging in age from 7 to 17 years) were included in the serial testing analysis, of whom 49 were concurrently skin tested at both visits.

Nearly 80% (50/63) of study subjects were born in a high-prevalence country (98% from Latin America), 37% of whom had resided in the United States for less than 5 years; 15 (24%) reported history of exposure to an active TB case (median, 3 years before baseline visit). Over two-thirds (68%) reported BCG vaccination, of whom 31 (72%) were confirmed by observation of scar and 47 (75%) reported having been skin tested before (median, 2.8 years before baseline visit), including 14 (21%) who reported a prior positive TST result. The 63 subjects completing both visits did not differ in characteristics from those who were not retested (Table 1), and there was no difference in the proportions of baseline QFT-positive and QFT-negative subjects who did not complete the second blood collection (25% versus 25%, respectively).

Baseline visit. At the baseline visit, 15/63 subjects (24%; 95%CI, 15 to 36%) were QFT positive at the 0.35 IU/ml difference cutoff, including 15/49 (24%) adults and 0/14 (0%) children (Fisher’s P = 0.03). By comparison, 23/63 (37%; 95%CI, 25 to 49%) subjects were considered TST positive at the 10-mm cutoff, including 20/49 (41%) adults and 3/14 (21%) children (Fisher’s P = 0.22). Of 23 considered TST positive, 15 (including 5 reporting prior positive TST results) were concurrently skin tested by our study personnel, and 8 were not retested.

Agreement between QFT-G-IT and TST results at the baseline visit was moderate (κ = 0.48 [±0.11]). Overall, 14 (22%) subjects had discordant results (McNemar P = 0.03), of whom 11 (79%) were TST+/QFT− and 3 (21%) were QFT+/TST−.
TABLE 1. Characteristics of subjects participating at baseline and 3-month follow-up visits

| Characteristic                          | At baseline (total) | Value for subject group<sup>a</sup>: |
|----------------------------------------|---------------------|-------------------------------------|
|                                        | (n = 88)            | At completion of 3-mo. tests         |
|                                        |                     | Total (n = 63) | TST<sup>+</sup> (n = 23) | QFT<sup>+</sup> (n = 15) |
| Age range (yr)                         | 6–80                | 6–80                  | 7–80                   | 20–80                  |
| Between 2 and 17 yr                    | 18 (21)             | 14 (22)               | 3 (13)                 | 0                      |
| ≥18                                    | 69 (79)             | 49 (78)               | 20 (87)                | 15 (100)               |
| Sex                                    |                     |                       |                        |                        |
| Male                                   | 36 (41)             | 25 (40)               | 8 (35)                 | 6 (40)                 |
| Female                                 | 52 (59)             | 38 (60)               | 15 (65)                | 9 (60)                 |
| Country of birth                       |                     |                       |                        |                        |
| United States                          | 19 (22)             | 13 (21)               | 0                      | 0                      |
| Latin American country                 | 68 (77)             | 49 (78)               | 22 (96)                | 14 (93)                |
| Other country of TB endemicity         | 1 (1)               | 1 (2)                 | 1 (4)                  | 1 (7)                  |
| <5 yr in the United States             | 20 (22)             | 16 (25)               | 10 (43)                | 6 (40)                 |
| BCG report                             | 54 (62)             | 43 (68)               | 22 (96)                | 14 (93)                |
| Visible scar                           | 37 (42)             | 31 (49)               | 18 (78)                | 10 (67)                |
| TB exposure                            | 19 (22)             | 15 (24)               | 9 (39)                 | 6 (40)                 |
| Median no. (range) of yr since TB exposure | 3 (1–19)           | 3 (1–19)              | 6 (1–19)               | 15 (1–19)              |
| Prior positive history                 | 68 (77)             | 47 (75)               | 18 (78)                | 9 (60)                 |
| Median no. (range) of yr since TST     | 2.9 (0.3–18)        | 2.8 (0.3–18)          | 2.2 (0.83–18)          | 2.9 (1.3–18)           |
| Prior positive                         | 21 (24)<sup>b</sup> | 14 (22)<sup>b</sup>   | 13 (56)                | 6 (40)                 |
| Visible scar                           | 37 (42)             | 31 (49)               | 18 (78)                | 10 (67)                |
| Prior TST                              | 19 (22)             | 15 (24)               | 9 (39)                 | 6 (40)                 |
| Median no. (range) of yr since TB exposure | 3 (1–19)           | 3 (1–19)              | 6 (1–19)               | 15 (1–19)              |
| Country of birth                       | 19 (22)             | 13 (21)               | 0                      | 0                      |
| Latin American country                 | 68 (77)             | 49 (78)               | 22 (96)                | 14 (93)                |
| Other country of TB endemicity         | 1 (1)               | 1 (2)                 | 1 (4)                  | 1 (7)                  |
| United States                          | 19 (22)             | 13 (21)               | 0                      | 0                      |
| Latin American country                 | 68 (77)             | 49 (78)               | 22 (96)                | 14 (93)                |
| Other country of TB endemicity         | 1 (1)               | 1 (2)                 | 1 (4)                  | 1 (7)                  |
| <5 yr in the United States             | 20 (22)             | 16 (25)               | 10 (43)                | 6 (40)                 |
| BCG report                             | 54 (62)             | 43 (68)               | 22 (96)                | 14 (93)                |
| Visible scar                           | 37 (42)             | 31 (49)               | 18 (78)                | 10 (67)                |
| TB exposure                            | 19 (22)             | 15 (24)               | 9 (39)                 | 6 (40)                 |
| Median no. (range) of yr since TB exposure | 3 (1–19)           | 3 (1–19)              | 6 (1–19)               | 15 (1–19)              |
| Prior positive history                 | 68 (77)             | 47 (75)               | 18 (78)                | 9 (60)                 |
| Median no. (range) of yr since TST     | 2.9 (0.3–18)        | 2.8 (0.3–18)          | 2.2 (0.83–18)          | 2.9 (1.3–18)           |
| Prior positive                         | 21 (24)<sup>b</sup> | 14 (22)<sup>b</sup>   | 13 (56)                | 6 (40)                 |

<sup>a</sup> Unless otherwise indicated in the leftmost column, data are expressed as numbers of subjects (with percentages given in parentheses). For TST, a ≥10-mm result was considered positive. For QFT, a result of ≥0.35 IU/ml was considered positive.

<sup>b</sup> TST-positive history and not retested with TST (10 at baseline, 8 in retested series).

(Tables 2 and 3). All 14 discordances occurred for subjects reporting histories of BCG vaccination; results from 20 subjects reporting no history of vaccination were 100% concordant. Among those reporting BCG vaccination, agreement was 0.35 (± 0.13) by BCG recall and 0.39 (± 0.14) by observation of scar. Of the 23 subjects considered TST positive at baseline, 4/8 (50%) of those not retested were discordant versus 7/15 (47%) of those testing positive at the baseline visit.

Among the 55 individuals with concurrent TST at the baseline visit, agreement levels were similar (κ = 0.50 ± 0.13). Seven of 15 (47%) with TST indurations of ≥10 mm had QFT values of <0.35 IU/ml (range, –0.07 to 0.30 IU/ml), including 3/9 (33%) of those with indurations of ≥15 mm.

**Three-month visit.** At the second visit 3 months later, 8 (13%) of 63 subjects had different QFT-G-IT results, including 5/15 (33%); 95 CI, 15 to 58%) initially positive subjects who reverted and 3/48 (6.3%; 95 CI, 2.2 to 16.8%) initially negative subjects who converted (κ = 0.63 ± 0.12). There was a nonsignificant trend for subjects with inconsistent QFT-G-IT results to have been discordant by TST at baseline (21% versus 10%; overall Fisher’s P = 0.36). Of 12 subjects whose QFT-G-IT and TST results were concordantly positive, 3 (25%) had reversions of QFT-G-IT, while there were two reversions (66%) among the 3 QFT positives who initially had a negative TST. Among 37 subjects whose QFT-G-IT and TST results were concordantly negative, 2 (5%) had QFT-G-IT conversions, compared to 1 (9%) among the 11 QFT-G negatives who initially had a positive TST.

For 49 individuals concurrently skin tested at both visits (Table 3), including 7 of the 8 with different QFT-G-IT results, QFT-G-IT reproducibility was lower (κ = 0.38 ± 0.19), and 11/49 (22%) QFT-G-IT2 results were discordant with TST2

**TABLE 2. Baseline agreement of TST and QFT-G-IT**

| QFT-G-IT1 result<sup>a</sup> | No. of subjects with indicated TST1 result<sup>a</sup>: | Total |
|-----------------------------|---------------------------------|-------|
|                             | <10 mm                          | ≥10 mm|       |
| <0.35 IU/ml                 | 37                              | 11    | 48    |
| ≥0.35 IU/ml                 | 3                               | 12    | 15    |
| Total                       | 40                              | 23    | 63    |

<sup>a</sup> QFT-G-IT1 and TST1 results are first-visit results: κ = 0.48 (±0.11), McNemar’s P = 0.03.

**TABLE 3. Prospective agreement of QFT-G-IT and TST for 49 concurrently skin-tested subjects**

| QFT1/QFT2 result<sup>a</sup> | No. of subjects with indicated TST1/TST2 result<sup>a</sup>: | Total no. with QFT1/QFT2 result<sup>a</sup>: |
|------------------------------|---------------------------------|---------------------------------|
| −/−                          | 32                              | 3                               | 3                |
| +/−                          | 2                               | 0                               | 0                |
| +/+                          | 1                               | 1                               | 2                |
| +/+                          | 0                               | 1                               | 2                |
| Total                       | 35                              | 5                               | 7                |

<sup>a</sup> TST1/2 and QFT1/QFT2, skin test and QFT-G-IT result at first and 3-month follow-up visit, respectively. Boldface indicates TST1/QFT1 concordance at baseline.
results (McNemar \( P = 0.03 \)). Overall, the two tests agreed on none of seven possible conversions (two QFT-G-IT and five TST) and on one of seven possible reversions (five QFT-G-IT and two TST).

Among the eight individuals with inconsistent QFT-G-IT results, the maximum IFN response at either visit was 0.68 IU/ml (Fig. 1 and Table 4). Compared with what was seen for 10 persistent positives, the 5 reversions had a significantly low mean IFN response at the first visit (0.56 [± 0.12] versus 4.99 [± 3.74]; \( P = 0.001 \)). Similarly, the mean IFN response of the three converters was significantly lower at the second visit (0.39 ± 0.03) than that seen for persistent positives (5.76 ± 3.9) (\( P = 0.001 \)).

All eight individuals with QFT-G-IT results that differed at first and 3-month testing were adults, and seven, including five reverters and two of the converters, reported histories of BCG vaccination (Table 4). Compared to 55 subjects with persistent QFT-G-IT results, the 8 with inconsistent results were, as evidenced by the interview responses of 2 of 3 converters and 3 of 5 reverters, more likely to recall TB exposure (63% versus 18%; Fisher’s \( P = 0.015 \)), including more-recent exposure (38% within 3 years versus 13%; Fisher’s \( P = 0.11 \)), and less likely to have been skin tested previously (25% versus 82%; Fisher’s \( P = 0.002 \)). Two QFT-G-IT reversions occurred in the same home.

Role of testing error. A test with an average sensitivity and specificity of 70% (beta, 70,29) and 98% (beta, 98,1), respectively, would be expected to have a median false-reversion rate of 31.72% (90CI, 25.4 to 38.5%) among initially positive subjects and a median false-conversion rate of 6.95% (90CI, 4.6 to 9.9%) when the apparent prevalence of infection is approximately 24% (beta, 15,47). Figure 2 shows the effect of varying these assumptions. When the specificity of a test is high, the reversion rate depends predominantly on the sensitivity of testing and is relatively stable for the given test characteristics and an apparent prevalence of infection of >10%. Conversely, the expected conversion rate is more dependent on the specificity of testing as well as the prevalence of infection. Conversion rates may also be difficult to interpret due to uncertainty about the difference between the apparent and “true” risk of infection.

DISCUSSION

Although the QFT-G assay using TB-specific RD1 antigens has many attractive features for U.S. health departments, prospective studies are needed to establish its utility in screening U.S. foreign-born and other targeted risk groups. In this small prospective series among immigrant households of northern California, 13% of subjects had inconsistent QFT-G-IT results over 3 months, including reversion and conversion rates of 33% and 6%, respectively. We also found that the inconsistent results had low positive absolute IFN responses (<0.69 IU/ml). These results support the concern that reproducibility, particularly of low positive results, may be an issue in screening high-risk U.S. immigrant populations.

To our knowledge, this is the first study examining the reproducibility of the QFT-G assay with an intended U.S. screening population. Similar results were reported for a cohort of

![FIG. 1. IFN-\( \gamma \) responses to QFT-G-IT by study visit. Mean IFN-\( \gamma \) responses at two study visits 3 months apart for subjects who were persistently negative (Pers. Neg.) or persistently positive (Pers. Pos.) and who converted (Convert) or reverted (Revert) according to QFT-G-IT. 0.35 IU/ml, the manufacturer’s recommended positive cutoff, is marked with a dotted line.](http://cvi.asm.org/)

### TABLE 4. Characteristics of eight subjects with inconsistent QFT-GOLD-IT results

| Status and subject no. | Age (yr) | Sex | BCG report | Time in United States (yr) | TB experienced (no. of yr prior) | Maximum response | IFN1 (IU/ml) | IFN2 (IU/ml) | TST1 (mm) | TST2 (mm) |
|------------------------|----------|-----|------------|---------------------------|----------------------------------|------------------|-------------|-------------|----------|----------|
| Reversion              |          |     |            |                           |                                  |                  |             |             |          |          |
| 358803                 | 42       | M   | Y          | >5                        | Y (1)                            | 0.68             | 0.18        | 0           | 0        | 0        |
| 358903                 | 38       | M   | Y          | <5                        | N                                | 0.6              | 0.18        | 17          | 12       |          |
| 360605                 | 40       | M   | Y          | >5                        | N                                | 0.66             | 0.13        | 0           | 17       |          |
| 373902*                | 20       | F   | Y          | <5                        | Y (18)                           | 0.49             | 0.26        | 10          | 0        |          |
| 373903*                | 41       | F   | Y          | <5                        | Y (18)                           | 0.38             | −0.12       | 16          | 11       |          |
| Conversion             |          |     |            |                           |                                  |                  |             |             |          |          |
| 372005                 | 46       | F   | Y          | >5                        | Y (2)                            | −0.02            | 0.42        | 12          | —d       |          |
| 384602                 | 20       | F   | N          | <5                        | Y (3)                            | 0.19             | 0.35        | 0           | 0        |          |
| 387001                 | 40       | M   | Y          | >5                        | N                                | 0.05             | 0.39        | 7           | 6        |          |

* Same household.

** M, male; F, female.

† Y, yes; N, no.

—d, TST not repeated.

428 PERRY ET AL. CLIN. VACCINE IMMUNOL.
Indian health care workers who were retested with the QFT-G-IT after 18 months. In that study, 24% of subjects reverted by QFT-G-IT, including 55% of those with IFN response levels of 0.35 to 0.69 IU/ml, and conversions ranged from 7 to 12%, depending on the stringency of conversion criteria (33). In 341 Gambian household contacts of TB cases followed by ELISPOT assay, reversion rates were 41% and 36% at 3 and 18 months, respectively, and corresponding conversion rates were 25% and 27% (versus 50% at 18 months by TST) (16). Although reproducibility appears to be somewhat better when the TST is concordant at baseline, this has been difficult to calibrate (35). Despite different IGRA formats, tuberculin formulations, and exposure gradients, serial testing studies to date have consistently reported that variable IFN-γ responses in assays based on the RD1 antigens appear to be common.

Possible explanations for variable responses, as well as discordance with TST, include biological factors, laboratory artifact, and test-related error. While the higher specificity of the QFT is an important explanation of cross-sectional discordance with TST, responses to early secreted antigens over the evolution of latent infection are not well understood and may have cycling patterns (10, 16). Most of our subjects were adult emigrants from Latin America who reside in multigenerational Hispanic communities, where travel to and from Latin America is common. Nearly one-fourth, including five-eighths of inconsistent responders, reported prior TB exposure (16% within 3 years), and over two-thirds reported BCG vaccination. It is possible that fluctuations in response to unknown environmental stimuli contributed to nonspecific variation in test results. Although the assay is theoretically invariant to BCG vaccination, this does not rule out confounding due to concomitant infections, such as NTM or helminths and other host factors that may also be present in BCG-vaccinated populations (4, 13). We were unable to evaluate reproducibility in
children due to low numbers; however, poor cross-sectional agreement with TST (9) and, in the Gambian study (16), higher reversion rates among children may also reflect the variability of IFN-γ responses during the adaptation of the cellular immune system (23). Individuals with more-remote exposure may also respond to the whole-blood incubation protocol differentially. Because the QFT-G assay captures recently activated lymphocytes, it has also been suggested that the 24-hour incubation period misses some dormant infections (6).

Other laboratory parameters could account for minor variations in results around the positive threshold. For replicate serum samples, the manufacturer estimated the overall test-related coefficient of variation for QFT-G to range from 7.6 to 12.3 (mean, 8.7 ± 0.7) for concentrations of 0.33 to 7.7 IU/ml, with higher coefficients of variation at lower concentrations (7). In our study, two of three conversions had follow-up values of <10% from the positive threshold of 0.35 IU/ml, and the third was within 20% of the threshold. Given current plate configurations, running paired serial samples may not always be feasible in consecutive enrollment series, although inclusion of reference samples on every plate would help to quantify this problem. For the in-tube collection system, the degeneracy of the four-point dilution curve after logarithmic curve transformation may also play a role in the nonspecific variation of positive results of <1 IU/ml. In our series, approximately one-half of positive results fell within the range of 0.35 to 1 IU/ml, including 40% between 0.35 and 0.7 IU/ml. Although the test is currently packaged for qualitative interpretation, the use of absolute values and the establishment of an equivocal or “borderline” zone—as is common with many ELISAs—would be useful to physicians.

Another important source of variability in repeated testing studies is test-related error. In the absence of a “gold standard” for latent TB infection, the contribution of test operating characteristics to inconsistent test results is difficult to estimate; however, because serial testing studies will propagate test-related error according to predictable Bayesian laws (38), they can be useful staging grounds for exploratory probabilistic models (2). Although the specificity of the QFT is considered excellent, the sensitivity of the test, particularly for screening populations, is less well understood. While our model was greatly simplified, our estimate of QFT-G-IT sensitivity at around 70% is consistent with estimates from clinically based studies (11, 29, 32). As illustrated by our uncertainty analysis, reversion rates in a high-sensitivity scenario depend predominantly on the sensitivity of testing. Thus, one explanation for the consistently high reversion rates observed in studies to date is the relative sensitivity of the test in a high-risk context. Conversely, the model also predicted that false-conversion rates, which depend on the negative predictive value of an antecedent test, might vary more widely across studies due to differences in underlying risk, such as exposure gradient and age, as well as with use of different cut points that alter relative sensitivity and specificity locally (15, 20, 22). Probabilistic models can be useful adjuncts in diagnostic technology assessment for approaching such questions as verification bias, precision, and sample size requirements for “two-stage” designs (19).

Because the QFT-G does not cross-react with BCG or most NTM, it has been considered a candidate for use in screening protocols such as two-step testing or repeated screening, where TST boosting is also a concern (26). Although the foreign-born are not candidates per se for regular repeated screening programs, our population was fairly typical of groups with ongoing indications for targeted screening in our area, and also of the testing limitations posed. In this series, if the QFT-G replaced TST outright, nearly one-half (48%) of TST positives, including one-third of those with indurations of ≥15 mm, would not be referred for prophylaxis evaluation, a discrepancy that is also noted in cross-sectional studies (21, 25). Of those referred, one-third would not be confirmed if the test were repeated as part of a clinical workup. Alternatively, if the QFT-G were used to confirm TST converters, two of five would have been detected at baseline, although one of these two would have been shown to revert if the QFT-G-IT were repeated. In a larger series from The Gambia, 25% of TST converters with positive baseline ELISPOT results reverted (17). In our study, among the subjects who were both QFT-G-IT and TST negative at baseline, not one of the seven converters identified by either test was identified by the other test, despite the fact that four of five TST converters had increments of greater than 10 mm. Most of our vaccinees were adults from Mexico, where BCG vaccination is usually given once at birth. Significantly boosted reactions in adults are unlikely to be explained by BCG vaccination at birth or NTM (14). The reasons for TST boosting in high-risk populations, including why IGRAs appear to anticipate this phenomenon (17, 31, 40), are not straightforward. Although the tests have excellent agreement for unvaccinated populations, the adoption of highly specific antigens in the screening of more immunologically complex populations is likely to be complex as well. Although our 3-month testing interval does not distinguish whether based on self-report or observation of scar. Second, although false-conversion rates at the second visit could have been caused by the antecedent TST at the first visit 3 months earlier, a recent study suggests this is unlikely (24). A third limitation may have been loss to follow-up, inasmuch as 22 (25%) did not complete the second QFT-G-IT. However, there was no difference in the proportions of baseline positives and negatives who were not available for the second visit. Finally, because of small sample size we could not adequately evaluate risk factors associated with conversions and rever-
sions, indicating the need for larger studies examining the reproducibility of the QFT-G in foreign-born and other high-risk screening populations.

Acknowledging these limitations, our study offers additional information about the reproducibility of the QFT-G-IT and discordance with TST. For a high-risk, BCG-vaccinated population, the agreement of QFT-G-IT and TST was moderate to poor, and nonspecific threshold variation in QFT-G results was common. Although the specificity of the test is important an explanation of cross-sectional TST discordance, much remains to be learned about the kinetics of diagnostic responses to early and late antigens in populations with different patterns of remote and recent exposure to mycobacterial antigens—a research agenda the IGRA’s are revealing at the same time they are being adopted. Although the CDC has recommended use of the QFT-G “in all circumstances” where the TST is indicated, additional studies are needed to understand the reproducibility and relative accuracy of the test before its utility in prospective screening programs can be defined.

ACKNOWLEDGMENTS

This work was supported by the Thrasher Research Fund and NIH-NIAID K23 AI054443. We gratefully acknowledge the assistance of the following community clinics with recruitment for this study: Santa Clara Valley Health Center Emergency Department (San Jose, CA), Santa Clara Valley Health Center Pediatric Clinic (San Jose, CA), East Valley Pediatric and Urgent Care Clinics (San Jose, CA), and Santa Clara County Office of Environmental Health (San Jose, CA).

REFERENCES

1. Andersen, P., M. E. Munk, J. M. Pullock, and T. M. Doherty. 2000. Specific immune-based diagnosis of tuberculosis. Lancet 356:1099–1104.
2. Begg, C. B., and R. A. Greens. 1983. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 39:207–215.
3. Berthet, F. X., P. B. Rasmussen, I. Rosenkrands, P. Andersen, and B. Giephe. 1998. A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). Microbiology 144:3195–3203.
4. Black, G. F., R. E. Weir, S. Floyd, L. Bliss, D. K. Warndorff, A. C. Crampin, M. Klein, T. H. Ottenshoff, P. Anderssen, and T. M. Doherty. 2002. BCG-induced increase in interferon-gamma response to mycobacterial antigens and efficacy of BCG vaccination in Malawi and the UK: two randomised controlled studies. Lancet 359:1303–1401.
5. Brock, I., M. Ruhwald, B. Lundgren, H. West, L. R. Mathiesen, and P. Ravn. 2005. Latent tuberculosis in HIV positive, diagnosed by the M. tuberculosis specific interferon-gamma test. Respir. Res. 7:56.
6. Cehovin, A., J. M. Cliff, P. C. Hill, R. H. Brookes, and H. M. Dockrell. 2007. Extended culture enhances sensitivity of a gamma interferon assay for latent Mycobacterium tuberculosis infection. Clin. Vaccine Immunol. 14:796–798.
7. Celestis. 2007. Quantiferon TB GOLD: precision and reproducibility report. Celestis, Valencia, CA.
8. Centers for Disease Control and Prevention. 2007. Trends in tuberculosis incidence, United States, 2006. MMWR Recomm. Rep. 56:245–250.
9. Connell, T. G., N. Curtis, S. C. Ranganathan, and J. P. Buttery. 2006. Performance of a whole blood interferon gamma assay for detecting latent infection with Mycobacterium tuberculosis in children. Thorax 61:616–620.
10. Demissie, A., E. M. Leyten, M. Abebe, L. Wassie, A. Assefa, G. Abate, H. Farhat, M., C. Greenaway, P. A. Aiken, R. A. Mcdonnell, L. M. Fabbri, and L. Richeldi. 2005. Identification of ELISPOT test for Mycobacterium tuberculosis infection. PLoS Med. 2:e192.
11. Hill, P. C., R. H. Brookes, A. Fox, D. Jackson-Silhah, D. J. Jeffries, M. D. Lugos, S. A. Donkor, M. L. Adefita, B. C. de Jong, J. M. Aiken, A. R. Adegboya, and K. P. Mcdonnell. 2007. Longitudinal study of an ELISPOT test for Mycobacterium tuberculosis infection. Lancet 367:1328–1334.
12. Hill, P. C., D. J. Jeffries, R. H. Brookes, A. Fox, D. Jackson-Silhah, D. M. Lugos, S. A. Donkor, B. C. de Jong, T. Corrah, R. A. Adegboya, and K. P. Mcdonnell. 2007. Using ELISPOT to expose false positive skin test conversion in tuberculosis contacts. PLoS ONE 2:e183.
13. Hui, S. L., and X. H. Zhou. 1998. Evaluation of diagnostic tests without gold standards. Stat. Methods Med. Res. 7:534–570.
14. Irwig, L., P. P. Glazou, G. Berry, C. Chock, P. Mock, and J. M. Simpson. 1994. Efficient study designs to assess the accuracy of screening tests. Am. J. Epidemiol. 140:759–769.
15. Jeffries, D. J., P. C. Hill, A. Fox, M. Lugos, D. J. Jackson-Silhah, R. A. Adegboya, and R. H. Brookes. 2006. Identifying ELISPOT and skin test cut-offs for diagnosis of Mycobacterium tuberculosis infection in The Gambia. Int. J. Tuberc. Lung Dis. 10:192–198.
16. Kang, Y. A., H. W. Lee, H. I. Youn, B. Cho, S. K. Han, Y. S. Shim, and J. J. Yim. 2005. Discrepancy between the tuberculin skin test and the whole-blood interferon gamma assay of latent tuberculosis infection in an intermediate tuberculosis-burden country. JAMA 293:2756–2761.
17. Laivan, A. A. A. Pathan, H. McShane, R. J. Wilkinson, M. Latif, C. P. Conlon, G. Pasvol, and A. V. Hill. 2001. Rapid detection of Mycobacterium tuberculosis infection by enumeration of antigen-specific T cells. Am. J. Respir. Crit. Care Med. 163:824–828.
18. Lewinsohn, D. A., M. L. Gennaro, L. Scholvinck, and D. M. Lewinsohn. 2004. Tuberculosis immunology in children: diagnostic and therapeutic challenges and opportunities. Int. J. Tuberc. Lung Dis. 8:659–674.
19. Leyten, E. M., C. Prins, A. W. Bossink, S. Thijssen, H. T. Ottenshoff, J. T. van Dissel, and S. M. Arend. 2007. Effect of tuberculin skin testing on a Mycobacterium tuberculosis-specific interferon- gamma assay. Eur. Respir. J. 30:1201–1216.
20. Mahomed, H., E. J. Hughes, T. Hawridge, D. Minnies, E. Simon, F. Little, W. A. Hanelekom, L. Geiter, and G. D. Hussey. 2006. Comparison of Mantoux skin test with three generations of a whole blood IFN-gamma assay for tuberculosis infection, Int. J. Tuberc. Lung Dis. 10:319–316.
21. Menzies, D. 1999. Interpretation of repeated tuberculin tests. Boosting, conversion, and reversion. Am. J. Respir. Crit. Care Med. 159:15–21.
22. Menzies, D. 2000. Tuberculin skin testing, p. 279–322. In L. Reichman and E. Hershfield (ed.), Tuberculosis: a comprehensive international approach, 2nd ed. Marcel Dekker, Inc., New York, NY.
23. Menzies, D., M. Pai, and G. Comstock. 2007. New tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. Ann. Intern. Med. 146:340–354.
24. National Institute for Health and Clinical Excellence. 2006. Tuberculosis: clinical diagnosis and management of tuberculosis, and measures for its prevention and control, p. 1-66. National Institute for Health and Clinical Excellence, London, United Kingdom.
25. Nguyen, M., S. Perry, and J. Parsonnet. 2005. Quantiferon(TB) predicts tuberculosis skin test boosting in U.S. foreign-born. Int. J. Tuberc. Lung Dis. 9:985–991.
26. Pai, M., R. Joshi, M. Bandyopadhyay, P. Narang, S. Dogra, B. Taksande, and S. Kalantri. 2007. Performance of a whole-blood interferon-gamma assay among patients with pulmonary tuberculosis and variations in T-cell responses during anti-tuberculosis treatment. Infection 35:98–103.
27. Pai, M., R. Joshi, S. Dogra, D. K. Mendiarrata, P. Narang, S. Kalantri, A. L. Reingold, J. M. Colford, Jr., L. W. Riley, and D. Menzies. 2006. Serial testing of health care workers for tuberculosis using interferon-gamma assays. Am. J. Respir. Crit. Care Med. 174:349–355.
28. Pai, M., S. Kalantri, and D. Menzies. 2006. Discordance between tuberculin skin test and interferon-gamma assays. Int. J. Tuberc. Lung Dis. 10:942–943.
29. Pai, M., and R. O’Brien. 2007. Serial testing for tuberculosis: can we make sense of T cell assay conversions and reversions? PLoS Med. 4:e208.
30. Pai, M., L. W. Riley, and J. M. Colford, Jr. 2004. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect. Dis. 4:761–766.
31. Perry, S., P. de la Luz Sanchez, S. Yang, T. D. Haggerty, P. Hurst, G. Perez-Perez, and J. Parsonnet. 2006. Gastroenteritis and transmission of
38. Perry, S., and J. Parsonnet. 2005. H. pylori infection in early life and the problem of imperfect tests. Int. J. Epidemiol. 34:1356–1358.

39. Porsa, E., L. Cheng, M. M. Seale, G. L. Delclos, X. Ma, R. Reich, J. M. Musser, and E. A. Graviss. 2006. 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. 13:53–58.

40. Pottumarthy, S., A. J. Morris, A. C. Harrison, and V. C. Wells. 1999. Evaluation of the tuberculin gamma interferon assay: potential to replace the Mantoux skin test. J. Clin. Microbiol. 37:3229–3232.

41. Santa Clara County Tuberculosis Prevention and Control Program. 2000. Guidelines for targeted skin testing of latent tuberculosis infection in Santa Clara County. Santa Clara County Disease Prevention and Control Program, San Jose, CA.

42. van Pinxteren, L. A., P. Ravn, E. M. Agger, J. Pollock, and P. Andersen. 2000. Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. Clin. Diagn. Lab. Immunol. 7:155–160.