Serial Testing of Refugees for Latent Tuberculosis Using the QuantiFERON-Gold In-Tube: Effects of an Antecedent Tuberculin Skin Test

Cristina A. Baker,* William Thomas, William M. Stauffer, Phillip K. Peterson, and Dean T. Tsukayama
Infectious Diseases and International Medicine Division, University of Minnesota, Minneapolis, Minnesota; Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota; Infectious Diseases Division, Hennepin County Medical Center, Minneapolis, Minnesota

Abstract. Screening for latent tuberculosis infection (LTBI) in refugee populations immigrating to low-incidence countries remains a challenge. We assessed the characteristics of the QuantiFERON-Gold In-Tube (QFT-GIT) compared with the tuberculin skin test (TST) in 198 refugees of all ages from tuberculosis-endemic countries. Diagnostic agreement between the first QFT-GIT and simultaneous TST was 78% (κ = 0.56) and between serial QFT-GITs was 89% (κ = 0.76). In serial QFT-GIT testing, 70% of subjects had an increased QFT-GIT value, perhaps the result of an antecedent TST in the setting of previous TB exposure. This boosting seemed to become less prevalent with time from TST and occurred less frequently in those with negative first QFT-GIT readings. Despite small changes in the quantitative results caused by nonspecific variation and boosting, the diagnostic result of the QFT-GIT was reliable. The QFT-GIT shows the potential to replace the TST for LTBI screening in refugees from tuberculosis-endemic areas.

INTRODUCTION

Over the last decade, the number of immigrants and refugees entering the United States from countries with a high prevalence of Mycobacterium tuberculosis (MTB) has continued to increase, leading to concerns about the potential public health impact. Currently, the rate of active tuberculosis among foreign-born persons is 9.7 times greater than among US-born persons. The majority of these cases are believed to repre- sent reactivation of infection acquired before immigration, highlighting the inadequacies of our current prevention and control strategies.

The tuberculin skin test (TST), long the diagnostic method for diagnosing latent tuberculosis infection (LTBI), is hindered by suboptimal sensitivity and specificity. A number of factors may adversely affect the performance characteristics of the skin test among refugees, including previous vaccination with bacillus Calmette-Guérin (BCG) or the presence of non-tuberculous mycobacterial infection. Newer interferon-γ release assays (IGRAs) have the potential to address many of the shortcomings of the TST in refugee populations such as eliminating false positives caused by BCG vaccination, distinguishing between tuberculous and non-tuberculous mycobacterial infection (NTB), and providing test results after a single patient visit.2–4

A recent meta-analysis found insufficient data to adequately characterize the performance of IGRAs in high-risk populations, such as refugees, and limited data regarding the tests’ reproducibility.5 Previous studies have shown that IGRAs perform well in healthy individuals from tuberculosis-endemic areas.6–8 One study looked at the performance of an IGRA for LTBI screening of asylum seekers in Norway.9 However, to our knowledge, no studies have examined the reproducibility of these tests in refugees.

An issue unique to refugees resettling to the United States that has potential to alter the IGRA assay’s performance and utility in this population is the fact that TST testing is frequently repeated during relocation. During resettlement, refugees are often repeatedly skin tested for LTBI because previous screening test results are lost or unavailable because of the transient medical care provided to this population during this time. It is unclear whether TST testing performed weeks earlier results in boosting of an IGRA result in this population.

It is well known that serial tuberculin skin testing can result in a boosting of the skin test result, with the greatest effect seen when the interval between tests is 1–5 weeks.10,11 However, it is unclear what role timing of TST placement plays in the boosting of an IGRA result. One study found no evidence of boosting caused by TST testing performed 3 days before IGRA testing.12 Conversely, two recent studies reported potential evidence for boosting of the QuantiFERON Gold (the earlier version of the test) in subjects between 2–4 and 6 weeks after TST testing.13,14

Domestic medical screening of refugees, which is encouraged by the Office of Refugee Resettlement for all newly arriving refugees, provides logistic challenges for the organizations assisting with resettlement. One practical challenge for clinics providing this care is coordinating tuberculin skin testing, which requires two clinic visits. A single-step test for TB that requires only one visit and performs at least as well as the TST would be desirable for many organizations and clinics serving this population.

In this study, we assessed the performance of the QuantiFERON-TB Gold In-Tube (QFT-GIT) in a refugee population relocating to the United States. The goals of this study were 2-fold: to evaluate the utility of the QFT-GIT as a screening tool for LTBI in a refugee population from tuberculosis-endemic countries and to assess the reproducibility of the test in the setting of an antecedent TST.

MATERIALS AND METHODS

Study participants. Refugees who presented for their domestic refugee medical screening examination at the Hennepin Public Health Clinic (Minneapolis, MN) were recruited for the study between July 2006 and March 2007. Only individuals who had been in the United States for < 6 months were eligible for inclusion. Exclusion criteria included suspected or confirmed active tuberculosis, HIV infection, malignancy, liver or kidney disease, and treatment with immunosuppressive drugs. Informed consent was obtained according to

* Address correspondence to Cristina A. Baker, Infectious Diseases and International Medicine Division, University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455. E-mail: bake0088@umn.edu

628
or a technical problem with the test (nil > 8.0 IU/mL). Statistical software was provided by the manufacturer. Values > 10 IU/mL are reported by the software as ≥ 10 IU/mL.

TSTs were performed using 0.1-mL test doses of tuberculin-purified protein derivative (bio-equivalent 5 TU of PPD-S; Aventis Pasteur, Toronto, Ontario, Canada) injected intradermally into the volar aspect of the forearm. The transverse diameter of the induration was measured 48–72 hours later by a trained healthcare provider. A TST was considered positive if ≥ 10 mm of skin induration was observed. TST conversion was defined as a skin test that went from negative (< 10 mm) to positive (≥ 10 mm) on repeat testing. Individuals performing TST measurements were blinded to the QFT-GIT results.

Data analysis. χ² tests or Fisher exact tests were used to compare categorical characteristics, and t tests were used to compare continuous characteristics. Association between dichotomized TST and QFT-GIT results was assessed by positive and negative agreement¹⁵ and κ.¹⁶ Proportion of positive agreement is defined as twice the number of individuals with positive results on both tests, divided by the sum of those positive on one of the tests plus those positive on the other test; negative agreement is defined in a parallel way. Test–retest reliability for IFN-γ responses from the first and second QFT-GIT was measured using an intraclass correlation coefficient. A nonparametric regression smoother, lowess, was used to estimate mean change over time in Figure 4. Data analyses were performed with SPSS (version 14.0; SPSS, Chicago, IL), SAS (SAS version 9.1; SAS Institute, Cary, NC), and R (R Development Core Team; http://www.R-project.org). Graphics were produced in R.

RESULTS

A total of 198 refugees arriving in the United States between April 2006 and January 2007 were enrolled in the study, with an average time between arrival and enrollment of 62 days (range, 19–163 days). Three refugees did not return for the second visit. A total of 159 refugees (80%) attended their third visit for repeat QFT-GIT testing. The most common reason for failure to return for Visit 3 was loss to follow-up. Five (3%) subjects refused repeat testing.

Of the 87 subjects eligible for a repeat TST (first TST < 10 mm), 41 (47%) returned to have the test interpreted. The 45 patients who did not complete repeat TST testing did not differ in characteristics (age, sex, country of origin) from those who returned for retesting (data not shown). The low TST retesting rate was caused by a combination of patient fatigue and inadequate clinic resources to ensure a fourth patient visit. The average time between testing at Visits 2 and 3 was 7.5 weeks (range, 2–52 weeks). The majority of subjects were young, female, and Somali. The likelihood of testing positive for LTBI with both tests differed across categories of age and sex (Table 1).

Comparison of simultaneous TST and QFT-GIT. Of the 195 subjects who completed both tests, 55.4% were TST positive and 53.8% were QFT-GIT positive. No initial QFT-GIT tests were indeterminate. Diagnostic agreement between the two tests was 78% (Table 2). Of refugees who were initially TST negative and QFT-GIT positive (N = 20), 12 (60%) had a QFT-GIT that narrowly exceeded the diagnostic cut-point for positivity of 0.35 IU/mL (range, 0.35–1.0 IU/mL).

To evaluate the effect of the value of the QFT-GIT cut-point on diagnostic agreement with the TST, we calculated

![Study flow diagram of refugee screening and testing for latent tuberculosis infection with the QFT-GIT and the TST.](image-url)
and positive and negative agreement with TST for cut-points ranging from 0 to 2.0 IU/mL (Figure 2). All measures of agreement were found to increase as the cut-point rose from 0.35 IU/mL, reaching their maximum at a cut-point of 1.325 IU/mL.

Serial QFT-GIT testing. Serial testing was completed in 159 subjects. One repeat test was indeterminate. Observed diagnostic agreement between the initial and repeat QFT-GIT was 89% ($\kappa$ = 0.76; 95% CI: 0.66–0.86) and the intraclass correlation of QFT-GIT values was 0.832 (95% CI: 0.78–0.87). Changes from the first to second QFT-GIT are shown in Figure 3.

Did the initial TST increase or “boost” the second QFT-GIT, and if so, was this effect related to time from TST? Of 159 subjects with repeated QFT-GIT measurements, 122 subjects had a first QFT-GIT IFN-$\gamma$ value < 10.0, so that an increase was possible, and had a recorded interval from TST. These intervals ranged from 2 to 52 weeks, but only eight intervals were > 16 weeks; because these are too sparsely scattered to provide much information, we restricted attention to the 114 subjects with intervals of 16 weeks or less. Figure 4 shows the change from the first QFT-GIT plotted against time from TST for these 114 subjects with a nonparametric-regression estimate of mean change as a function of time. As is visually evident in Figure 3, subjects with a positive first QFT-GIT reading tended to experience larger increases.

Our null hypothesis was that changes in serial QFT-GIT IFN-$\gamma$ values are random with median zero, that is, that the chance of an increase is 50%. Boosting would be shown by a significantly higher proportion of increases than decreases. To test this, we divided the 114 subjects into two groups as to whether their interval after the TST was < 5 or > 5 weeks; there is a gap in the data at Week 5 (Figure 4). In the 38 subjects retested 2–5 weeks after TST, the second QFT-GIT increased 87%, with a mean increase of 2.8 ± 0.5 (SE) IU/mL, whereas for 76 subjects retested 5–16 weeks after TST, 69% increased with mean change 0.8 ± 0.2 IU/mL. Both proportions of increases were larger than the null value of 50% ($P < 0.002$), and the first was also significantly greater than the second ($P = 0.033$). In addition, both mean changes were significantly greater than zero. In this group of refugees, there was strong evidence for boosting in the second QFT-GIT, with the rate and size of increase diminishing with time from TST.

### Table 1

Demographic characteristics for 195 refugees with TST and at least one QFT-GIT and comparison of subgroups with jointly positive and negative test results among 195 refugees enrolled in the study

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Frequency (%) (N = 195)</th>
<th>TST+/QFT−GIT+ (%) (N = 85)</th>
<th>TST−/QFT−GIT− (%) (N = 67)</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (median)</td>
<td>19 (range 1–81)</td>
<td>72 (85)</td>
<td>40 (60)</td>
<td>0.001</td>
</tr>
<tr>
<td>Adult</td>
<td>140 (72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child (&lt; 18 years of age)</td>
<td>55 (28)</td>
<td>13 (15)</td>
<td>27 (40)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td>0.011</td>
</tr>
<tr>
<td>Male</td>
<td>67 (34)</td>
<td>44 (52)</td>
<td>20 (30)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>128 (66)</td>
<td>41 (48)</td>
<td>47 (70)</td>
<td></td>
</tr>
<tr>
<td>Country of origin</td>
<td></td>
<td></td>
<td></td>
<td>0.8785</td>
</tr>
<tr>
<td>Somalia</td>
<td>149 (76)</td>
<td>65 (76)</td>
<td>54 (81)</td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>26 (13)</td>
<td>11 (13)</td>
<td>9 (13)</td>
<td></td>
</tr>
<tr>
<td>Liberia</td>
<td>6 (3)</td>
<td>5 (6)</td>
<td>2 (3)</td>
<td></td>
</tr>
<tr>
<td>Eritrea</td>
<td>1 (0.5)</td>
<td>1 (1)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>1 (0.5)</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>1 (0.5)</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ukraine</td>
<td>5 (2.5)</td>
<td>1 (1)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Russian Federation</td>
<td>2 (1)</td>
<td>1 (1)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>4 (2)</td>
<td>1 (1)</td>
<td>2 (3)</td>
<td></td>
</tr>
</tbody>
</table>

*P value was calculated using $\chi^2$ for age and sex and Fisher exact test for country of origin.

### Table 2

Agreement between initial QFT-GIT and TST

<table>
<thead>
<tr>
<th>TST negative</th>
<th>TST positive*</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT negative</td>
<td>67 (34)†</td>
</tr>
<tr>
<td>QFT-GIT positive‡</td>
<td>20 (10)</td>
</tr>
</tbody>
</table>

* TST positive: ≥ 10 mm of induration.
† Percent of total (number of refugees = 195).
‡ QFT-GIT positive: ≥ 0.35 IU/mL IFN-$\gamma$.

Overall agreement 78% (67 + 85)/195
Proportion of tests with positive agreement 0.798 (95% CI: 0.74–0.86)
Proportion of tests with negative agreement 0.757 (95% CI: 0.69–0.83)
$\kappa$ 0.56 (95% CI: 0.44–0.67)

Figure 2. Three measures of agreement ($\kappa$, positive agreement, negative agreement) between TST and initial QFT-GIT, computed as the minimum value for a positive reading on the QFT-GIT which varies from 0 to 2.0. The recommended value is 0.35, indicated by a vertical line.
Ignoring time from TST, 86% of subjects with a positive first QFT-GIT reading increased on the second QFT-GIT, whereas 68% of those with negative first QFT-GIT readings increased ($P = 0.022$ for the difference). Both of these proportions were significantly greater than the null value of 50% ($P < 0.004$).

Rates of increases were similar for those positive and negative on TST, 82% and 68%, respectively, but the difference did not reach significance ($P = 0.059$). However, because there was strong evidence for a time effect on boosting, averaging these rates over time may introduce bias.

Among the 117 participants with initial QFT-GIT IFN-γ values ≤ 1.0 IU/mL, 18 had qualitatively discordant results (12 converted, 6 reverted). Discordance in serial testing was not related to sex or TST status or to time from TST.

There were no cases of discordance observed among refugees with initial QFT-GIT IFN-γ values > 1.0.

**DISCUSSION**

Refugees are a high-risk group for which an accurate diagnosis of LTBI is of paramount importance. Using the TST as a screening test in this population is fraught with diagnostic challenges such as false positives with BCG immunization or NTM infection. In addition, skin testing requiring two clinic visits is a logistic challenge for refugees and public health clinics. Within our programmatic setting, only 47% of eligible patients completed repeat TST testing for the purpose of this study. This finding suggests that similar rates of return for follow-up testing during contact studies may be expected in comparable patient populations.

In addition to the practical issues outlined above, our study supports previous findings that the QFT-GIT may have slightly better agreement with the TST than other generations of the test used in similar high-risk situations. In the study of Calvalho and others of immigrants in Italy, agreement between the TST and QuantiFERON-TB Gold was 71% ($\kappa = 0.37$). However, in the report of Winjie and others of asylum seekers in Norway, agreement between the two tests was better at 79% ($\kappa = 0.51$). Similarly, we found positive and negative agreement of nearly 80% ($\kappa = 0.56$). Furthermore, we observed fewer indeterminate ($N = 1$) results than in other studies.

Screening with the QFT-GIT is not free of challenges. One concern is how to interpret a minimally positive QFT-GIT (0.35–1.0 IU/mL) in the setting of a negative TST. Without a true gold standard or longitudinal studies following subjects with negative TSTs and positive QFT-GITs to determine the likelihood of developing active TB, it is difficult to know whether the QFT-GIT is in fact more sensitive and how to proceed clinically in terms of offering treatment of LTBI.

Another potential issue with the QFT-GIT is reproducibility. As in serial skin testing, we observed nonspecific variation in serial QFT-GIT testing. Such nonspecific variation leading to discordance in serial testing has been observed in other studies. We found that this nonspecific variation resulted in a small proportion of cases (11%) in which the diagnosis of LTBI based on QFT-GIT results could be called into question.
This seemed to occur only in refugees with initial QFT-GIT IFN-γ values < 1.0 IU/mL, suggesting that a QFT-GIT result with a value > 1.0 IU/mL may be more likely to indicate a true positive.

Within our population of refugees, diagnostic agreement between simultaneous TST and QFT-GIT was found to increase as the cut-point for QFT-GIT test positivity was raised, reaching a maximum at a cut-point of 1.325 IU/mL. This suggests that, in a population with presumed high rates of LTBI, a higher cut-point might result in fewer instances in which refugees without LTBI are misclassified as having LTBI (false positive). On the other hand, raising the cut-point for the sake of better agreement may increase the likelihood of a false negative. From a public health and patient perspective, this may not be wise, because the potential for a missed diagnosis of LTBI would likely present a greater hazard to the patient and the public. Given that there is no gold standard for LTBI testing, longitudinal studies looking at the likelihood of developing active TB after a positive QFT-GIT will ideally provide more information on the appropriate cut-point for the test.

We observed that, during serial testing, a significant proportion (2/3) of refugees had a higher second QFT-GIT test value. Those with a positive first QFT-GIT reading tended to experience larger increases. We speculate that this difference may be the result of boosting of a true-positive QFT-GIT from an antecedent skin test. Although the presence of a possible boosting phenomenon was qualitatively present in some cases, it did not frequently affect the diagnostic results of the QFT-GIT. In the 70% of refugees that had a larger second QFT-GIT value, only 7% had tests that qualitatively converted from negative to positive.

Previous studies examining whether TST testing results in boosting of QFT-GIT results were inconclusive and suggested that the presence of boosting may be related to the time interval between testing with the two modalities.12,13 We found that there was strong evidence for boosting of the second QFT-GIT but that the rate and size of increase diminished with time from TST. Therefore, it seems that skin testing performed > 6 weeks before screening with the QFT-GIT would be less likely to result in boosting of the QFT-GIT result.

There were some important limitations to our study. We did not collect individual data on BCG vaccination because confirmation of immunization was unavailable. However, we felt confident that the literature supported the QFT-GIT’s high specificity with respect to BCG,7 and the majority of subjects in our study were Somali, a group with one of the highest annual TB case rates (889/100,000) among the foreign-born in the United States.19 As well, all patients received a TST, which limited our ability to definitively conclude whether the increased IFN-γ values seen with the second QFT-GIT were caused by boosting or other host factors.

In conclusion, the QFT-GIT holds promise to replace the TST for LTBI screening in refugee populations resettling to low-incidence countries. Boosting of a QFT-GIT result because of an antecedent TST in the context of prior tuberculosis exposure may occur in refugees who are screened with both modalities during the resettlement process or during contact studies. Such boosting should be considered to be caused by prior infection with TB because the QFT-GIT’s specificity eliminates the possibility of boosting caused by BCG. Although observed, boosting rarely seems to alter the clinical applicability of the test. Furthermore, the rate and size of boosting seems to diminish over time from TST, with little boosting seen after 6 weeks. In a refugee screening setting, it is reasonable to consider a QFT-GIT positive if its value is > 0.35 IU/mL, given the high prevalence of LTBI in refugees from TB-endemic countries and the potentially greater personal and public health risk of a missed diagnosis. Ideally, longitudinal studies looking at the likelihood of developing active TB given a minimally positive QFT-GIT will help to clarify the issue of the appropriate cut-point for the test. Finally, additional research is needed to better understand the reasons for the test’s variability in this and other populations. Studies looking at the interaction between host factors and test reproducibility would be of interest.

Received June 5, 2008. Accepted for publication November 20, 2008.

Acknowledgments: The authors thank Kathy Kennedy from the Hennepin County Medical Center Microbiology Laboratory for technological expertise.

Financial support: This work was funded, in part, by a research grant from the International Society of Travel Medicine. The QuantiFERON-TB Gold In-Tube tests were provided at no cost by Cellestis.

Disclaimer: The authors state that they have no conflicts of interest.

Authors’ addresses: Cristina A. Baker, William M. Stauffer, Phillip K. Peterson, and Dean T. Tsukayama, Infectious Diseases and International Medicine Division, University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455, E-mails: bake0088@umn.edu, stauf005@umn.edu, peter137@umn.edu, and tsuka001@umn.edu. William Thomas, Division of Biostatistics, School of Public Health University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455, E-mail: thoma003@umn.edu.

REFERENCES

8. Pai M, Gokhale K, Joshi R, Dogra S, Kalantri S, Mendiratta DK, Narang P, Daley CL, Granich RM, Mazurek GH, Reingold AL,


