Discordance Among Commercially-Available Diagnostics for Latent Tuberculosis Infection

<table>
<thead>
<tr>
<th>Journal:</th>
<th>American Journal of Respiratory and Critical Care Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>Blue-201107-1244OC.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>OC - Original Contribution</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>07-Nov-2011</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Mancuso, James; WRAIR, PM; USUHS, PMB</td>
</tr>
<tr>
<td></td>
<td>Mazurek, Gerald; CDC, DTBE</td>
</tr>
<tr>
<td></td>
<td>Tribble, David; USUHS, PMB</td>
</tr>
<tr>
<td></td>
<td>Olsen, Cara; USUHS, PMB</td>
</tr>
<tr>
<td></td>
<td>Aronson, Naomi; USUHS, Medicine</td>
</tr>
<tr>
<td></td>
<td>Geiter, Lawrence; Otsuka Pharmaceuticals</td>
</tr>
<tr>
<td></td>
<td>Goodwin, Donald; USAFSAM</td>
</tr>
<tr>
<td></td>
<td>Keep, Lisa; USUHS, PMB</td>
</tr>
<tr>
<td>Keywords:</td>
<td>tuberculosis screening, interferon gamma release assays,</td>
</tr>
<tr>
<td></td>
<td>tuberculosis diagnostics, non-tuberculous mycobacteria</td>
</tr>
</tbody>
</table>
REVISION 1: Discordance among Commercially-Available Diagnostics for Latent Tuberculosis Infection

James D. Mancuso, MD DrPH†*  
Gerald H. Mazurek, MD§  
David Tribble, MD DrPH*  
Cara Olsen, DrPH MS *  
Naomi E. Aronson, MD §§  
Lawrence Geiter, PhD††  
Donald Goodwin, DrPH**  
Lisa W. Keep, MD MPH*

† Walter Reed Army Institute of Research  
Division of Preventive Medicine  
503 Robert Grant Rd.  
Silver Spring, MD 20910

* Uniformed Services University of the Health Sciences  
Department of Preventive Medicine and Biometrics  
4301 Jones Bridge Rd.  
Bethesda, MD 20814

§ Centers for Disease Control and Prevention  
Division of TB Elimination  
1600 Clifton Rd.  
Atlanta, GA 30333

†† Otsuka Pharmaceuticals  
TB Products Unit  
Rockville, Maryland 20850

** US Air Force School of Aerospace Medicine  
2947 5th Street  
Wright-Patterson AFB, OH  
45433-7212

§§ Uniformed Services University of the Health Sciences  
Department of Medicine  
4301 Jones Bridge Rd.  
Bethesda, MD 20814

Author Contributions: JM, DT, GM, CO, NA, LG, DG, and LK—all had substantial participation in conception and design of the study, acquisition or analysis of data, interpretation of the data, and revision of the article.

Subject Category Descriptor Number: 11.1—Diagnosis of Tuberculosis or Latent Infection

At a glance commentary
Scientific knowledge on the subject: There is substantial discordance between the tuberculin skin test (TST) and the interferon gamma release assays (IGRAs) in populations with low prevalence of tuberculosis, and most positive results from the three tests identify different people.

What this study adds to the field: This study suggests that the majority of positives from any of these tests are false positives in low prevalence populations. To support the current recommendations to treat TB diagnostics interchangeably, targeted testing using risk stratified interpretation should be used for the IGRAs as with the TST.

Clinical trial registered at: www.clinicaltrials.gov, Identifier: NCT00804713

Sources of support: US Army Public Health Command
5158 Blackhawk Rd.
Aberdeen Proving Ground, MD 21010-5403

Infectious Diseases Clinical Research Program
Uniformed Services University
4301 Jones Bridge Rd.
Bethesda, MD 20814

Disclaimer:
Support for this work (IDCRP #021) was provided by the Infectious Disease Clinical Research Program (IDCRP), a Department of Defense (DoD) program executed through the Uniformed Services University of the Health Sciences. This project has been funded in whole, or in part, with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), under Inter-Agency Agreement Y1-AI-5072. The content of this publication is the sole responsibility of the authors and does not necessarily reflect the views or policies of the NIH, the Department of Health and Human Services, the Centers for Disease Control and Prevention, the DoD or the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Keywords: Tuberculosis screening, Interferon Gamma Release Assays, Nontuberculous mycobacteria

Running head: Analysis of Discordance between TST and IGRAs

Corresponding author and reprint requests information:
James Mancuso, MD DrPH
Walter Reed Army Institute of Research
Preventive Medicine Residency Program
503 Robert Grant Rd.
Silver Spring, MD 20910
Phone: 301-319-9218
FAX: 301-319-9104
Email: james.mancuso@us.army.mil

Word Count: 3669
Abstract (word count=250)

Rationale
There is uncertainty regarding how to interpret discordance between tests for latent tuberculosis infection.

Objective
The objective of this study was to assess discordance between commercially-available tests for latent tuberculosis in a low prevalence population, including the impact of nontuberculous mycobacteria.

Methods
This was a cross-sectional comparison study among 2,017 military recruits at Fort Jackson, South Carolina from April to June 2009. Several tests were performed simultaneously with a risk factor questionnaire, including: 1) QuantiFERON®-TB Gold In-Tube test, 2) T-SPOT®.TB test, 3) tuberculin skin test, and 4) Battey skin test using purified protein derivative from the Battey bacillus.

Measurements and Main Results
In this low-prevalence population, the specificities of the three commercially-available diagnostic tests were not significantly different. Of the 88 subjects with a positive test, only 10 (11.4%) were positive to all three tests; 20 (22.7%) were positive to at least two. Bacille Calmette Guerin vaccination, tuberculosis prevalence in country of birth, and Battey skin test reaction size were associated with tuberculin skin test positive, interferon gamma release assay negative test discordance. Increasing agreement between the three tests was associated with both epidemiologic criteria indicating risk of infection and with quantitative test results.
Discussion

For the majority of positive results the three tests identified different people, suggesting that in low prevalence populations the majority of discordant results are due to false positives. False positive tuberculin skin test reactions associated with reactivity to nontuberculous mycobacteria and Bacille Calmette Guerin vaccination may account for a proportion of test discordance observed.
Introduction

There is continued uncertainty as to which diagnostic test for latent tuberculosis infection (LTBI) is most accurate in the US population, the tuberculin skin test (TST) or interferon gamma release assays (IGRAs), including the QuantiFERON®-TB Gold In-Tube test (QFT-GIT) and T-SPOT®.TB test (T-Spot). There is no gold diagnostic standard for evaluating the performance of the IGRAs in comparison with the TST other than the long-term progression to active TB in cohort studies (1). In the absence of a gold standard, IGRAs are routinely compared in practice to the TST in cross-sectional evaluation studies, using active TB cases to assess sensitivity and low-risk populations to assess specificity (2, 3). In these studies, significant discordance is often found between IGRA and TST results. In a study of Navy recruits, 11 of 15 (73%) of the highest risk individuals—whose country of birth had a rate of active TB of >100 per 100,000 person-years and who had TST reactions of at least 15 mm—had negative QFT-Gold tests (4). There are several explanations for these discordant results, including the use of region of difference one (RD1) antigens in the IGRAs, which might result in greater specificity. However, it is also possible that the TST may have greater sensitivity, that the IGRAs may detect only unresolved or more recent infections (5), or that TST and IGRAs provide complementary measures of immune response (6).

Non-tuberculous mycobacteria (NTM) may be an important potential source of false-positive tests for *M. tuberculosis* infection in areas where the likelihood of infection is very low (7), such as the southeast US (8). The late Dr. George Comstock remarked in 1975 that “the frequency of cross-reactions to tuberculin in this [Navy recruit] population is sufficiently great that the prevalence of true tuberculous infections among white recruits may already be approaching zero (9).” The prevalence of sensitization to NTM in the US population increased
from 11% in 1972 to 17% in 2000 (10). Military recruits are an excellent population to explore NTM sensitization as a potential source of TST/IGRA discordance, since BCG and waning sensitivity to TST due to age are uncommon and recruits originate from a wide geographic area.

The impact of cross-reactivity on TST results has been previously investigated by comparing results of skin tests performed with purified protein derivative (PPD) made from *M. tuberculosis* (PPD-Seibert, or PPD-S) and several NTM, including *Mycobacterium intracellulare*. PPD-Battey (PPD-B) is a skin test antigen made from the Boone strain of *Mycobacterium intracellulare* in a manner similar to how PPD-S is made from *M. tuberculosis*. A skin test performed with PPD-B is referred to as a Battey Skin Test (BST). The BST has been used as an aid in the differentiation of reactivity to *M. tuberculosis* from reactivity to NTM in Navy recruit (8, 11) and National Health and Nutrition Examination Survey (NHANES) studies (10, 12, 13). It has also been used in many other smaller epidemiologic studies (14-19). The objectives of this study were to compare commercially available tests for LTBI in a heterogeneous, low LTBI prevalence US population and to assess the impact of NTM reactivity on test discordance.

**Methods**

**Study Enrollment**

After providing written informed consent, recruits originating from all areas of the US, age 18 years or older, undergoing routine entry-level medical processing at Fort Jackson were screened for participation in the study. Recruits were excluded from participating if they: 1) had a history of severe reaction to the TST, 2) were pregnant by urine human chorionic gonadotropin (HCG) testing, 3) had received a live virus vaccine within the past 30 days, or 4) had a major viral infection at the time of screening.
Regulatory Information

PPD-B was used as a skin test antigen under an Investigational New Drug (IND) Protocol sponsored by the Uniformed Services University (USU) in Bethesda, MD. The Infectious Diseases Institutional Review Board (IRB) at USU provided approval and oversight of the study.

Study Design

This cross-sectional comparison study among Army recruits at Fort Jackson, SC consisted of five elements: 1) a TB risk factor questionnaire (RFQ), 2) T-Spot, 3) QFT-GIT, 4) BST, and 5) TST.

Study Methodology/Procedures

Risk Factor Questionnaire (RFQ)

The RFQ contained questions about demographics, TB exposure, work history, location of residence, and other factors shown in Table 1. This questionnaire was developed from the risk factors previously identified in the military and non-military literature (20-25), as well as other factors considered candidates for causal relationships with LTBI.

Interferon Gamma Release Assays

Blood for QFT-GIT and T-Spot was collected at the time of routine phlebotomy for recruit in-processing. Personnel performing IGRAs were blinded to all patient data. QFT-GIT was performed according to package insert instructions, including incubation and centrifugation of blood within the prescribed times at Fort Jackson, NC and completion of ELISAs at the US Air Force School of Aerospace Medicine, Brooks City-Base, TX and the Centers for Disease
Control and Prevention (CDC), Atlanta, GA (26). ELISAs were performed with the aid of Triturus automated ELISA workstations (Grifols USA, LLC, Los Angeles, CA). T-Spot was performed per package insert instructions (27) at the Oxford Immunotec, Ltd. Laboratory, Marlborough, MA with the addition of T-cell Xtend (Oxford Immunotec, Ltd., Oxfordshire, UK) immediately prior to peripheral blood mononuclear cell recovery. IGRAs were interpreted according to published guidelines (36); however, in the analysis of quantitative responses, borderline T-Spot results (i.e. TB Response of 5, 6 or 7 spots) were coded as “negative.”

**Skin testing**

Both TST and BST were placed by study personnel after the blood draw. All personnel involved in placement and reading of the skin test were trained and monitored to strictly adhere to standard operating procedures based on published methods for skin test administration and interpretation (20, 28). The Mantoux technique was used to intradermally administer 0.1 mL (5 TU) of Tubersol® tuberculin PPD (Sanofi Pasteur Ltd., Toronto, Ontario, Canada) and 0.1 mL (0.01 mcg) of PPD-B at the same sitting. One skin test was placed on each forearm. A random number table for each recruitment day determined which PPD was placed on each arm. The transverse diameter of induration at each skin test site was measured 2 days after PPD injection. Participants and those administering and reading the skin tests were blinded to which skin test antigen was administered on each arm.

**Definitions**

Recruits were categorized using a Risk Stratified Interpretation (RSI), as previously described by the Centers for Disease Control and Prevention (CDC) (29). The only modifications to the CDC criteria were that no time limitations were placed on 1) contact with an
active TB case or 2) immigration from a high prevalence country. The TB prevalence reported by the World Health Organization in 1990 was used to estimate exposure risk by country using groups of: 1) less than 20 per 100,000, 2) 20 to 100 per 100,000, and 3) greater than 100 per 100,000 (4, 30). BCG status was determined by self report. There was a strong correlation between reported history of BCG vaccination, presence of BCG scar, and foreign birth in this population. There was no significant difference in the results when using history of BCG vaccination or BCG scar (data not shown). Test specificity was estimated by assuming that recruits with no risk factors for \textit{M. tuberculosis} exposure were uninfected. An invalid test was defined as those with insufficient blood, misplaced or dislodged caps, an insufficient number of PBMCs recovered, or other laboratory errors. Test discordance was categorized as "TST positive / IGRA negative" or "TST negative / IGRA positive" for both the QFT-GIT and the T-Spot. BST induration size was categorized into four 5 mm intervals and one \( \geq 20 \text{ mm} \). A dominant BST reaction was defined as a BST reaction of at least 2 mm greater than the TST reaction.

**Statistical Considerations**

The proportion of recruits with a positive TST, T-Spot, and QFT-GIT were compared using McNemar’s test for correlated proportions, as were specificity and the proportion of indeterminate and invalid results for each test. The proportions of discordant and concordant results were also measured, as well as test agreement using kappa (\( \kappa \)) coefficient. Factors associated with discordance were evaluated using standard chi-square bivariate statistics, stratified analyses, and multivariate analysis. Prevalence ratios were directly estimated for both bivariate and multivariate analyses. As the log-binomial model failed to converge due to
numerical instability, Poisson regression with robust variance estimation was used to calculate multivariate prevalence ratios \((31)\). The variables evaluated are listed in Table 1.

Discordance between TST and IGRA was further assessed using associations between demographic and exposure variables including category of BST induration. TST positive / IGRA negative discordance was assessed separately from TST negative / IGRA positive discordance. The comparison group used for both of these analyses was the group of concordant negatives.

Results

Figure 1 depicts subject participation and follow-up in a flow chart. Of the 3,095 recruits approached from April 1 to June 11, 2009, 2,697 were eligible to participate in the study, of which 2,017 subjects (75%) enrolled. Of the thirty-nine recruits who withdrew prior to blood collection or completion of skin testing, 30 were for administrative reasons unrelated to the study. Characteristics of the remaining 1,978 study participants are shown in Table 1. TST results were available for all of the remaining 1,978 participants, and were read a mean of 45 hours after PPD injection (range of 40 to 50 hours). TST induration was detected in 122 (6.2%) participants and ranged from 2 to 80 mm. No significant digit preference was identified on inspection of the histogram of reaction size (see online supplement). T-Spot and QFT-GIT results were available for 1,913 (96.7%) and 1,850 (93.5%), respectively. QFT-GIT was invalid for 128 (6.5%) subjects, and 17 (0.9%) of the valid QFT-GIT gave indeterminate results. T-Spot was invalid for 65 (3.3%) subjects, 6 (0.3%) of the valid T-Spots were indeterminate and 23 (1.2%) had borderline results with a TB Response between 5 and 7 spots. The relatively high proportion of subjects with invalid tests was due to a need for numerous tubes of blood for
routine recruit inprocessing and investigational tests, and an IRB restriction against additional phlebotomy solely to collect blood for investigational tests.

Of the 1,803 subjects who had valid positive, negative, or borderline results for all three tests, 1,373 were classified as low-risk for \textit{M. tuberculosis} infection based on history, but 19 of them had borderline T-Spot results. Among the 1,354 recruits without identifiable risks and with determinate results for all three tests, estimates of TST specificity were 99.3\% (95\% CI: 98.7\%, 99.7\%) when using the 15 mm cutoff for positive recommended by the CDC for persons at low risk of exposure \cite{29}, or 98.6\% (95\% CI: 97.8\%, 99.2\%) when using a 10 mm cutoff. The specificity of the IGRAs was 98.7\% for the T-Spot (1,336 negatives among 1,354 low-risk recruits, 95\% CI: 97.9\%, 99.2\%); and 98.8\% for the QFT-GIT (1,338 negatives among 1,354 low-risk recruits, 95\% CI: 98.1\%, 99.3\%). Estimates of specificity were unchanged when borderline T-Spot results were coded as negative and included in the analysis (data not shown). None of the differences were statistically significant.

There were 1,781 subjects who had valid positive or negative result, excluding subjects with indeterminate or borderline results by any test. Table 2 shows the number and proportion of positive tests by test type, as well as the prevalence of BST reactions among the positives. An analysis of risk factors for positive tests such as BCG vaccination and foreign birth is presented in another recent publication \cite{32}. The proportion of subjects with a 10 mm or greater TST reaction was significantly larger than with any other test or TST cutoff (p < 0.05), and the proportion of subjects with a 15 mm or greater TST reaction was significantly smaller than that found by RSI or a 10 mm cutoff (p < 0.0001). None of the other differences in proportions was statistically significant. 19 of 57 (33\%) recruits with 10 mm or greater TST reactions did not have identifiable risks for \textit{M. tuberculosis} infection. When using RSI as suggested by the CDC
(29), 2.7% were positive, a similar proportion of positive results as was observed for both the T-Spot (1.9%) and QFT-GIT (2.0%).

Using the RSI for TST, 88 (4.9%) had a positive result to at least one of the three tests. Of these, only 10 (11.4%) were positive to all three tests; 20 (22.7%) were positive to at least two of the tests. Modest agreement between TST and the two IGRAs was seen in Tables 3a, 3b, and 3c. In contrast, good agreement was seen with TST when using different blinded readers (kappa=0.79, see online supplement).

Of the 48 subjects with a positive TST, 9 (18.8%) had a dominant BST reaction, defined as a BST reaction of at least 2 mm greater than the TST, as shown in Table 2. Table 4 further examines the associations of potential risk factors for TST positive, IGRA negative discordance. Strong dose-response relationships were observed between discordance and BST reaction size, TB prevalence in country of birth, and BCG vaccination. No significant associations were seen between any variables and IGRA-positive / TST-negative discordance or T-Spot / QFT-GIT discordance (data not shown).

Among the 1,803 subjects with valid tests and determinate results, Table 5 shows the agreement between the 3 tests by quantitative result of each test. Subjects with borderline T-Spot results were included in this analysis to assess a continuum of TB Responses including 5-7 spots. This shows an association of increased proportion of greater quantitative test results with increased concordance between the tests. This dose-response relationship was highly significant for all three tests. Table 6 shows the quantitative test results for each test according to risk strata. The association of increasing risk for infection with \textit{M. tuberculosis} with increasing proportion of IGRA response suggests a similar relationship between the quantitative test results of the
IGRAs as is seen with the TST. The dose-response relationship between risk of infection with *M. tuberculosis* and quantitative test result was also highly significant for each test. Similarly, Table 7 shows the association of higher TB risk strata with greater test concordance; this dose-response relationship was also statistically significant.

**Discussion**

This study suggests that the three commercially-available TB diagnostics have similar results in US populations with low TB prevalence. IGRAs were designed to increase specificity, but in this study specificity for the IGRAs was no better than TST specificity among low-risk recruits when interpreted using a TST cutoff of 15 mm according to published guidelines. The prevalence of positive results and dose-response relationships with TB exposure were also similar for the three tests. Despite these areas of agreement, the three tests identified different people for the majority of positive test results. In this trial, TST positive, IGRA negative discordance was strongly associated with Battey skin test results, supporting other evidence that NTM sensitization can cause false positive TST results. Conversely, the IGRAs showed little evidence of cross-reactivity to NTM by the BST. Although this suggests that NTM and BCG sensitization cause false positive TST results and that this contributes to discordance, but these factors do not explain the etiology of most of the discordance encountered.

Other aspects of test discordance examined in this study include the dose-response associations seen between the TB exposure risk, quantitative results of the TST and IGRA testing, and degree of concordance between the 3 tests. These data suggest that in low-prevalence populations, the majority of positives resulting from any of the three commercially-available diagnostic tests are false positives since: 1) 77% of subjects with positive test results were positive by only one test,
2) lower quantitative results were associated with smaller risk for TB exposure, 3) lower quantitative results were associated with single positive tests, and 4) lower risk for TB exposure was associated with decreasing test agreement.

The problem of low positive predictive value (PPV) is well known and understood with the TST (33). Use of risk stratification is currently recommended to guide the interpretation of the TST as a way to increase PPV and reduce false positivity (29); this is not employed for the IGRAs. This study suggests that performance of the IGRAs would also benefit from the use of a risk-stratified interpretation, as it would increase PPV and reduce the number of false positives. These findings support the CDC’s recommendation that people at minimal risk of infection (who are at greatest risk of a false positive result) should not be targeted for LTBI testing, regardless of whether a TST or IGRA is used (34).

This study provides reliable estimates of specificity in a low-risk population. Although both IGRAs are generally reported to have specificity higher than the TST (35), there was surprisingly little difference in specificity between TST and either IGRA seen in this study. The specificity estimates for TST and IGRA found in this study are similar to those found in previous studies of Navy recruits (4). Although the specificity of QFT-GIT is sometimes thought to be higher than that of T-Spot (2, 3), the estimated specificities of the two tests were not different in this study. The strong dose-response relationships between TB exposure and positive TST and IGRA results were also similar to those reported previously (2, 3). These findings further support the CDC’s recommendation that IGRAs may be used in place of the TST, but that testing should be targeted to avoid false positive results (34).
While IGRAs and TST may be used in the diagnosis of LTBI, they do not give equivalent information and often have discordant results. Several studies have compared results from different IGRAs and from TST “head-to-head” (36-43), and although the agreement between QFT-GIT and T-Spot has generally been very good, discordant results between the IGRA and TST have been found in up to 20-30% of subjects (3). The magnitude of discordance is demonstrated in this study by the low kappa values and the high proportion of discordance seen among positives, as 68 of 88 (77%) individuals with at least one positive test were positive to only one of the three tests. The frequency of test discordance has varied among studies, leading some authors to conclude that the IGRAs have lower sensitivity (37), while others have concluded that the IGRAs have better specificity due to less cross-reactivity with BCG vaccine and to waning sensitivity due to age (36). The differences may also be due to differences in the populations studied.

A few studies have provided evidence that NTM contribute to discordance between the TST and IGRA (4, 40), but none have used the BST. In this study, the strong dose-response relationship between increasing BST reaction size and increasing prevalence of discordance provide additional evidence that false-positive TSTs contribute to this discordance. BCG vaccination was also strongly associated with discordance in this study. On the other hand, risk-stratified TST positive, IGRA negative discordance was also associated with TB prevalence in country of birth and being Asian or from the Pacific Islands, traditionally factors associated with high risk of developing disease if infected. Thus, some of the discordance also may be attributable to lower sensitivity of the IGRAs compared to TST, or a combination of these two factors.
A limitation of this study is the lack of a gold standard for determining the presence of *M. tuberculosis* infection, making it difficult to assess the true significance of discordance between TST and IGRAs. The significance of reactivity to BST also has some uncertainty. Although it has previously been shown to assist in differentiating between LTBI and cross-reactivity due to NTM (8, 11), BST reactivity also may be due to cross-reactivity following *M. tuberculosis* infection (16, 44). Furthermore, there are other mycobacteria that contain RD1 antigens, such as *M. kansasii, M. szulgai, or M. marinum*; infection with these NTM may cause false-positive reactions to both TST and IGRAs (2, 45). There is potential for misclassification of several variables, including the recall of BCG vaccination among recruits, history of prior TB or LTBI diagnosis or treatment, and contact with a TB case. Although samples were sent blinded to all participating laboratories, the potential still exists for other residual sources of misclassification bias. Recruits are a low-risk population and may not represent the causes of test discordance in other higher-risk populations. Furthermore, since this research was performed in the high-throughput basic training setting, the administrative limitations imposed resulted in larger proportions of inadequate blood draws and TST reading times which were slightly shorter than optimal.

This study highlights the need for better understanding of the significance of test discordance, particularly the need for longitudinal data on progression to active TB among those with discordant test results. Applying the methodology used in this study to other populations (11, 12) may provide a more complete understanding of the test interpretation and test discordance. Finally, further research is needed to better characterize the most appropriate cutoffs to be used for the risk-stratified interpretation of the IGRAs, in order to maximize sensitivity and specificity in different risk groups and populations.
References

47. 21CFR610.11 (General Biological Products Standards—General Safety). 2009.
Acknowledgements

This study would not have been possible without the support of the Uniformed Services University of the Health Sciences, the US Army Public Health Command, the medical and installation leadership at Fort Jackson, and the Soldiers who participated in the study. The United States Army Public Health Command (Edgewood, MD) played no role in the design, conduct, collection, management, analysis, interpretation of the data; or in preparation of the manuscript. However, the organization did review and approve the manuscript. The Infectious Disease Clinical Research Program (Bethesda, MD) participated in all phases of the study, including design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. Oxford Immunotec (Marlborough, MA) performed T-Spot testing (blinded) as an in-kind contribution, but played no other role in the design, conduct, collection, management, analysis, interpretation of the data, preparation, review or approval of the manuscript. Laboratory and technical support was also provided by the US Air Force School of Aerospace Medicine and the Centers for Disease Control and Prevention’s Division of TB Elimination.

We thank Christine Anderson, PhD (Food and Drug Administration) for graciously supplying her expertise in preparing the Battey antigen and testing it for human use. She also provided valuable comments on the manuscript during preparation. This study was greatly assisted by the incredible energy and expertise of Ms. Carey Schlett of the Infectious Disease Clinical Research Program. Her guidance and constant supervision were invaluable to the completion of the study. Finally, we thank Dr. Richard Menzies, who provided invaluable advice and expertise in designing and setting up this study.
Figure 1. Flow chart of study comparing the tuberculin skin test with two interferon-gamma release assays in 2009 at Fort Jackson, SC.

* Note: 1 of these subjects had a borderline T-Spot and an invalid QFT-GIT
### Table 1. Characteristics of Study Participants (n=1978)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number*</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1294</td>
<td>65.5</td>
</tr>
<tr>
<td>Female</td>
<td>681</td>
<td>34.5</td>
</tr>
<tr>
<td>Age (SD)</td>
<td>1974</td>
<td>21.8 years (S.D. 4.6 years)</td>
</tr>
<tr>
<td>Race/ethnic group**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1298</td>
<td>65.6</td>
</tr>
<tr>
<td>Black</td>
<td>459</td>
<td>23.2</td>
</tr>
<tr>
<td>Hispanic</td>
<td>221</td>
<td>11.4</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>117</td>
<td>5.7</td>
</tr>
<tr>
<td>Other</td>
<td>63</td>
<td>3.2</td>
</tr>
<tr>
<td>Prevalence of TB in country of birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20 per 100,000</td>
<td>1873</td>
<td>94.7</td>
</tr>
<tr>
<td>20-100 per 100,000</td>
<td>35</td>
<td>1.8</td>
</tr>
<tr>
<td>&gt;100 per 100,000</td>
<td>70</td>
<td>3.5</td>
</tr>
<tr>
<td>BCG vaccinated</td>
<td>69</td>
<td>3.5</td>
</tr>
<tr>
<td>Greatest prevalence of TB among countries the subject lived in or traveled to for &gt; 1 month</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20 per 100,000</td>
<td>1811</td>
<td>91.6</td>
</tr>
<tr>
<td>20-100 per 100,000</td>
<td>62</td>
<td>3.1</td>
</tr>
<tr>
<td>&gt;100 per 100,000</td>
<td>105</td>
<td>5.3</td>
</tr>
<tr>
<td>Contact with someone with TB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In same household</td>
<td>24</td>
<td>1.2</td>
</tr>
<tr>
<td>Casual contact</td>
<td>73</td>
<td>3.7</td>
</tr>
<tr>
<td>Health care work</td>
<td>232</td>
<td>11.7</td>
</tr>
<tr>
<td>Lived or worked in congregate setting</td>
<td>120</td>
<td>6.1</td>
</tr>
<tr>
<td>Farm work or residence</td>
<td>383</td>
<td>19.4</td>
</tr>
<tr>
<td>Current Residence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast US</td>
<td>337</td>
<td>17.0</td>
</tr>
<tr>
<td>Southeast US</td>
<td>657</td>
<td>33.2</td>
</tr>
<tr>
<td>Western US</td>
<td>706</td>
<td>35.7</td>
</tr>
<tr>
<td>Other</td>
<td>278</td>
<td>14.1</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1480</td>
<td>75.1</td>
</tr>
<tr>
<td>&lt; 1 pack per day</td>
<td>395</td>
<td>20.0</td>
</tr>
<tr>
<td>1+ pack per day</td>
<td>97</td>
<td>4.9</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 12</td>
<td>257</td>
<td>13.0</td>
</tr>
<tr>
<td>12</td>
<td>1095</td>
<td>55.4</td>
</tr>
<tr>
<td>13-15</td>
<td>468</td>
<td>23.7</td>
</tr>
<tr>
<td>16+</td>
<td>158</td>
<td>8.0</td>
</tr>
<tr>
<td>Prior TB treatment</td>
<td>45</td>
<td>2.3</td>
</tr>
<tr>
<td>Prior TB skin test performed</td>
<td>710</td>
<td>35.9</td>
</tr>
<tr>
<td>Prior positive skin test</td>
<td>24</td>
<td>3.4 (of those with a prior test)</td>
</tr>
<tr>
<td>Unknown result</td>
<td>54</td>
<td>7.6 (of those with a prior test)</td>
</tr>
</tbody>
</table>

* Note: some cells do not total to 1978 due to missing data
** Note: Recruits could choose more than one group; Other includes 39 American Indian, 8 Bi- or multi-racial S.D.=Standard Deviation; TB=tuberculosis; BCG=Bacille Calmette Guerin Vaccine
Table 2. Number and Proportion of Positive TB Tests, and Proportions of Positives Reacting to the Battey Skin Test (BST)

<table>
<thead>
<tr>
<th>TB TEST TYPE</th>
<th>Number positive (% of total)^</th>
<th>Number (%) of positives with BST ≥ 10 mm</th>
<th>Number (%) of positives with dominant BST ≥ 10 mm**</th>
</tr>
</thead>
<tbody>
<tr>
<td>TST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 10 mm</td>
<td>57 (3.2%)*</td>
<td>33 (58%)</td>
<td>12 (21%)</td>
</tr>
<tr>
<td>≥ 15 mm</td>
<td>25 (1.4%)†</td>
<td>16 (64%)</td>
<td>3 (12%)</td>
</tr>
<tr>
<td>RSI§</td>
<td>48 (2.7%)</td>
<td>28 (58%)</td>
<td>9 (19%)</td>
</tr>
<tr>
<td>T-Spot</td>
<td>34 (1.9%)</td>
<td>11 (32%)</td>
<td>4 (12%)</td>
</tr>
<tr>
<td>QFT-GIT</td>
<td>36 (2.0%)</td>
<td>8 (22%)</td>
<td>3 (8%)</td>
</tr>
</tbody>
</table>

TST=tuberculin skin test
T-Spot = T-SPOT®.TB test
QFT-GIT=QuantiFERON®-TB Gold In-Tube test
BST= Battey skin test using PPD-B
TB=Tuberculosis
§ RSI=TST positive defined by Risk Stratified Interpretation(29)
* TST with 10 mm cutoff had a statistically significant increased proportion of positives (p<0.05) compared to all other TST cutoffs, QFT-GIT, and T-Spot
† TST with 15 mm cutoff had a statistically significant decreased proportion of positives compared to the RSI and 10 mm TST cutoffs only (p<0.0001)
^ Total population = 1781 with determinate results by all 3 tests
** Dominant BST reaction is defined as BST reaction at least 2 mm larger than the TST reaction
### Table 3a. Agreement between T-Spot and TST

<table>
<thead>
<tr>
<th></th>
<th>TST positive*</th>
<th>TST negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-Spot positive</td>
<td>15 (0.8%)</td>
<td>19 (1.1%)</td>
<td>34 (1.9%)</td>
</tr>
<tr>
<td>T-Spot negative</td>
<td>33 (1.9%)</td>
<td>1714 (96.2%)</td>
<td>1747 (98.1%)</td>
</tr>
<tr>
<td>Total</td>
<td>48 (2.7%)</td>
<td>1733 (97.3%)</td>
<td>1781</td>
</tr>
</tbody>
</table>

TST = tuberculin skin test  
T-Spot = T-SPOT®.TB test  
* TST positive defined by Risk Stratified Interpretation(29)  
§ Includes 23 subjects with borderline TB Response of 5 spots (11 subjects), 6 spots (11 subjects), or 7 spot (1 subject)  
% agreement = 97.1%  
K (95% CI) = 0.35 (0.22, 0.49)

### Table 3b. Agreement between QFT-GIT and TST

<table>
<thead>
<tr>
<th></th>
<th>TST positive*</th>
<th>TST negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFT-GIT positive</td>
<td>11 (0.6%)</td>
<td>25 (1.4%)</td>
<td>36 (2.0%)</td>
</tr>
<tr>
<td>QFT-GIT negative</td>
<td>37 (2.1%)</td>
<td>1708 (95.9%)</td>
<td>1745 (98.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>48 (2.7%)</td>
<td>1733 (97.3%)</td>
<td>1781</td>
</tr>
</tbody>
</table>

TST = tuberculin skin test  
QFT-GIT = QuantiFERON®-TB Gold In-Tube test  
* TST positive defined by Risk Stratified Interpretation(29)  
% agreement = 96.5%  
K (95% CI) = 0.24 (0.12, 0.37)

### Table 3c. Agreement between T-Spot and QFT-GIT

<table>
<thead>
<tr>
<th></th>
<th>QFT-GIT positive</th>
<th>QFT-GIT negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-Spot positive</td>
<td>14 (0.8%)</td>
<td>20 (1.1%)</td>
<td>34 (1.9%)</td>
</tr>
<tr>
<td>T-Spot negative</td>
<td>22 (1.2%)</td>
<td>1725 (96.9%)</td>
<td>1747 (98.1%)</td>
</tr>
<tr>
<td>Total</td>
<td>36 (2.0%)</td>
<td>1745 (98.0%)</td>
<td>1781</td>
</tr>
</tbody>
</table>

T-Spot = T-SPOT®.TB test  
QFT-GIT = QuantiFERON®-TB Gold In-Tube test  
% agreement = 97.6%  
K (95% CI) = 0.39 (0.24, 0.54)
Table 4. Associations between selected characteristics and discordance between positive TST (≥ 10mm) and negative IGRA

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Recruits with a negative QFT-GIT result (n=1745)</th>
<th>Recruits with a negative T-Spot result (n=1747)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (of 1745)</td>
<td>N (of 1747)</td>
</tr>
<tr>
<td></td>
<td># with TST positive§ (n=37)</td>
<td># with TST positive§ (n=33)</td>
</tr>
<tr>
<td></td>
<td>Bivariate Prevalence Ratio (95% CI)</td>
<td>Bivariate Prevalence Ratio (95% CI)</td>
</tr>
<tr>
<td></td>
<td>Multivariate Prevalence Ratio (95% CI)</td>
<td>Multivariate Prevalence Ratio (95% CI)</td>
</tr>
<tr>
<td>Age (years)^</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>1.1 (1.1, 1.2) *</td>
<td>1.1 (1.1, 1.2) *</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>1136 607</td>
<td>1132 613</td>
</tr>
<tr>
<td></td>
<td>21 16</td>
<td>19 14</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>1.0 (REF)</td>
</tr>
<tr>
<td></td>
<td>2.7 (1.1, 6.2)</td>
<td>2.6 (1.1, 6.1)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>607 16</td>
<td>613 14</td>
</tr>
<tr>
<td></td>
<td>14 (0.7, 2.7)</td>
<td>14 (0.7, 2.7)</td>
</tr>
<tr>
<td>Race/Ethnic Group</td>
<td>White</td>
<td>White</td>
</tr>
<tr>
<td></td>
<td>1158 397 60 194</td>
<td>1157 400 59 197</td>
</tr>
<tr>
<td></td>
<td>11 10 12 7</td>
<td>11 10 8 7</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>1.0 (REF)</td>
</tr>
<tr>
<td></td>
<td>2.7 (1.1, 6.2)</td>
<td>2.6 (1.1, 6.1)</td>
</tr>
<tr>
<td></td>
<td>Asian or PI</td>
<td>Asian or PI</td>
</tr>
<tr>
<td></td>
<td>60 12</td>
<td>59 8</td>
</tr>
<tr>
<td></td>
<td>21.1 (9.7, 45.7)</td>
<td>14.2 (6.0, 34.1)</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>Hispanic</td>
</tr>
<tr>
<td></td>
<td>194 7</td>
<td>197 7</td>
</tr>
<tr>
<td></td>
<td>3.8 (1.5, 9.7)</td>
<td>3.7 (1.5, 9.5)</td>
</tr>
<tr>
<td></td>
<td>TB prevalence in country of birth or long-term residence</td>
<td>TB prevalence in country of birth or long-term residence</td>
</tr>
<tr>
<td></td>
<td>&lt;20 per 100,000</td>
<td>&lt;20 per 100,000</td>
</tr>
<tr>
<td></td>
<td>1659 19</td>
<td>1662 18</td>
</tr>
<tr>
<td></td>
<td>8 7</td>
<td>18 10</td>
</tr>
<tr>
<td></td>
<td>8.7 (2.7, 27.9)</td>
<td>9.2 (2.9, 29.7)</td>
</tr>
<tr>
<td></td>
<td>20-100 per 100,000</td>
<td>20-100 per 100,000</td>
</tr>
<tr>
<td></td>
<td>56 15</td>
<td>55 12</td>
</tr>
<tr>
<td></td>
<td>23.3 (12.6, 43.6)</td>
<td>20.1 (10.2, 39.7)</td>
</tr>
<tr>
<td></td>
<td>7.7 (3.0, 20.1)</td>
<td>7.9 (2.8, 22.5)</td>
</tr>
<tr>
<td></td>
<td>&gt; 100 per 100,000</td>
<td>&gt; 100 per 100,000</td>
</tr>
<tr>
<td></td>
<td>BCG vaccination</td>
<td>BCG vaccination</td>
</tr>
<tr>
<td></td>
<td>No 1690 20</td>
<td>No 1695 19</td>
</tr>
<tr>
<td></td>
<td>26.1 (14.5, 47.0)</td>
<td>24.0 (12.8, 45.2)</td>
</tr>
<tr>
<td></td>
<td>4.0 (1.6, 9.7)</td>
<td>4.4 (2.1, 9.5)</td>
</tr>
<tr>
<td></td>
<td>Yes 55 17</td>
<td>Yes 52 14</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>1.0 (REF)</td>
</tr>
<tr>
<td></td>
<td>26.1 (14.5, 47.0)</td>
<td>24.0 (12.8, 45.2)</td>
</tr>
<tr>
<td></td>
<td>4.0 (1.6, 9.7)</td>
<td>4.4 (2.1, 9.5)</td>
</tr>
<tr>
<td></td>
<td>BST reaction</td>
<td>BST reaction</td>
</tr>
<tr>
<td></td>
<td>0-4 mm 1417 8</td>
<td>1423 7</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>1.0 (REF)</td>
</tr>
<tr>
<td></td>
<td>5.9 mm 133 7</td>
<td>132 6</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>9.2 (3.2, 27.1)</td>
</tr>
<tr>
<td></td>
<td>5.5 (2.2, 13.5)</td>
<td>5.4 (2.0, 14.5)</td>
</tr>
<tr>
<td></td>
<td>10-14 mm 145 11</td>
<td>142 9</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>12.9 (4.9, 34.1)</td>
</tr>
<tr>
<td></td>
<td>6.0 (2.3, 15.5)</td>
<td>5.2 (1.9, 14.2)</td>
</tr>
<tr>
<td></td>
<td>15-19 mm 34 4</td>
<td>35 5</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>29.0 (9.7, 87.0)</td>
</tr>
<tr>
<td></td>
<td>20+ mm 16 7</td>
<td>15 6</td>
</tr>
<tr>
<td></td>
<td>77.5 (31.9, 188)</td>
<td>81.3 (31.0, 213)</td>
</tr>
<tr>
<td></td>
<td>37.5 (11.0, 128)</td>
<td>86.1 (32.6, 227)</td>
</tr>
<tr>
<td></td>
<td>Region of birth</td>
<td>Region of birth</td>
</tr>
<tr>
<td></td>
<td>NE 295 7</td>
<td>NE 297 7</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>1.0 (REF)</td>
</tr>
<tr>
<td></td>
<td>0.6 (0.2, 1.7)</td>
<td>0.6 (0.2, 1.7)</td>
</tr>
<tr>
<td></td>
<td>SE 591 9</td>
<td>SE 590 9</td>
</tr>
<tr>
<td></td>
<td>0.6 (0.2, 1.7)</td>
<td>0.6 (0.2, 1.7)</td>
</tr>
<tr>
<td></td>
<td>West 613 10</td>
<td>West 614 9</td>
</tr>
<tr>
<td></td>
<td>0.6 (0.2, 1.7)</td>
<td>0.6 (0.2, 1.7)</td>
</tr>
<tr>
<td></td>
<td>Other 246 11</td>
<td>Other 246 8</td>
</tr>
<tr>
<td></td>
<td>0.6 (0.2, 1.7)</td>
<td>1.4 (0.5, 3.8)</td>
</tr>
<tr>
<td></td>
<td>Farm work</td>
<td>Farm work</td>
</tr>
<tr>
<td></td>
<td>No 1401 34</td>
<td>No 1406 3</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>1.0 (REF)</td>
</tr>
<tr>
<td></td>
<td>0.4 (0.1, 1.2)</td>
<td>0.4 (0.1, 1.3)</td>
</tr>
<tr>
<td></td>
<td>Yes 344 3</td>
<td>Yes 341 30</td>
</tr>
<tr>
<td></td>
<td>1.0 (REF)</td>
<td>1.0 (REF)</td>
</tr>
</tbody>
</table>
* Multivariate models did not include variables with p>0.05

^ Note: Age was modeled as a continuous variable and all observations were used in the analysis; all others were categorical

§ TST positive defined by Risk Stratified Interpretation(29)

BST = Battey skin test, TST = tuberculin skin test, T-Spot = T-SPOT®.TB test, QFT-GIT = QuantiFERON®-TB Gold In-Tube test, IGRA = interferon-gamma release assay, BCG = Bacille Calmette Guerin Vaccine
### Table 5. Comparison of the Quantitative Results of Discordant and Concordant Specimens

<table>
<thead>
<tr>
<th>Test Results</th>
<th>N</th>
<th>Quantitative TST result</th>
<th>Quantitative QFT-GIT result</th>
<th>Quantitative T-Spot result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0-4 mm</td>
<td>5-9 mm</td>
<td>10-14 mm</td>
</tr>
<tr>
<td>All tests negative</td>
<td>1713</td>
<td>1676 (97.8%)</td>
<td>30 (1.8%)</td>
<td>7 (0.4%)</td>
</tr>
<tr>
<td>One test positive</td>
<td>69</td>
<td>35 (50.7%)</td>
<td>19 (27.5%)</td>
<td>14 (20.3%)</td>
</tr>
<tr>
<td>TST only</td>
<td>33</td>
<td>0</td>
<td>18 (54.6%)</td>
<td>14 (42.4%)</td>
</tr>
<tr>
<td>QFT-GIT only</td>
<td>21</td>
<td>21 (100%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T-Spot only</td>
<td>15</td>
<td>14 (93.3%)</td>
<td>1 (6.7%)</td>
<td>0</td>
</tr>
<tr>
<td>Two tests positive</td>
<td>11</td>
<td>1 (9.1%)</td>
<td>1 (9.1%)</td>
<td>4 (36.4%)</td>
</tr>
<tr>
<td>TST &amp; QFT-GIT</td>
<td>2</td>
<td>0</td>
<td>1 (50.0%)</td>
<td>1 (50.0%)</td>
</tr>
<tr>
<td>TST &amp; T-Spot</td>
<td>5</td>
<td>0</td>
<td>1 (20.0%)</td>
<td>4 (80.0%)</td>
</tr>
<tr>
<td>QFT-GIT &amp; T-Spot</td>
<td>4</td>
<td>1 (25.0%)</td>
<td>1 (25.0%)</td>
<td>2 (50.0%)</td>
</tr>
<tr>
<td>All three tests positive</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>2 (20.0%)</td>
</tr>
</tbody>
</table>

^ Note: The 22 borderline T-Spot results were added into this analysis in order to have results for the 5-7 spots category. A borderline T-Spot test result was coded as negative for the purpose of this analysis.
* $\chi^2$ (trend) for 15 mm compared to the other groups = 725.7, p<0.0001
† $\chi^2$ (trend) for 1.0 compared to the other groups = 578.0, p<0.0001
§ $\chi^2$ (trend) for 8 spots compared to the other groups = 1005.2, p<0.0001
TST = tuberculin skin test
T-Spot = T-SPOT®.TB test
QFT-GIT = QuantiFERON®-TB Gold In-Tube test
### Table 6. Risk Stratification of TB Exposure by Quantitative Result of TST, QFT-GIT, and T-Spot

<table>
<thead>
<tr>
<th>Risk Stratification†</th>
<th>N</th>
<th>Quantitative TST result</th>
<th>Quantitative QFT-GIT result</th>
<th>Quantitative T-Spot result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0-4 mm 5-9 mm 10-14 mm ≥ 15 mm*</td>
<td>&lt;0.35 0.35-0.99 ≥ 1.0†</td>
<td>≤ 4 spots 5-7 spots ≥ 8 spots§</td>
</tr>
<tr>
<td>High risk</td>
<td>21</td>
<td>18 (85.7%) 1 (4.8%) 1 (4.8%) 1 (4.8%)</td>
<td>18 (85.7%) 2 (9.5%) 1 (4.8%)</td>
<td>20 (95.2%) 0 1 (4.8%)</td>
</tr>
<tr>
<td>(5 mm criteria)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate risk</td>
<td>409</td>
<td>362 (88.5%) 10 (2.4%) 21 (5.1%) 16 (3.9%)</td>
<td>392 (95.8%) 7 (1.7%) 10 (2.4%)</td>
<td>391 (95.6%) 3 (0.7%) 15 (3.7%)</td>
</tr>
<tr>
<td>(10 mm criteria)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>1373</td>
<td>1332 (97.0%) 21 (1.5%) 10 (0.7%) 10 (0.7%)</td>
<td>1356 (98.8%) 13 (1.0%) 4 (0.3%)</td>
<td>1336 (97.3%) 19 (1.4%) 18 (1.3%)</td>
</tr>
<tr>
<td>(15 mm criteria)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^ Note: The 22 borderline T-Spot results were added into this analysis in order to have results for the 5-7 spots category.
* χ^2^ (trend) for 15 mm compared to the other groups = 24.6, p<0.0001
† χ^2^ (trend) for 1.0 compared to the other groups = 24.4, p<0.0001
§ χ^2^ (trend) for 8 spots compared to the other groups = 11.5, p=0.0007
TST=tuberculin skin test
T-Spot = T-SPOT®.TB test
QFT-GIT=QuantiFERON®-TB Gold In-Tube test
† Risk stratification as per CDC Risk Stratified Index(29)
Table 7. Risk Stratification of TB Exposure by Test Agreement†

<table>
<thead>
<tr>
<th>Test Results</th>
<th>N</th>
<th>High risk (5 mm criteria)</th>
<th>Moderate risk (10 mm criteria)</th>
<th>Low risk (15 mm criteria)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tests negative</td>
<td>1693</td>
<td>16 (1.0%)</td>
<td>359 (21.2%)</td>
<td>1318 (77.9%)</td>
</tr>
<tr>
<td>One test positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TST only</td>
<td>68</td>
<td>4 (5.9%)</td>
<td>33 (48.5%)</td>
<td>31 (45.6%)</td>
</tr>
<tr>
<td>QFT-GIT only</td>
<td>32</td>
<td>2 (6.3%)</td>
<td>23 (71.9%)</td>
<td>7 (21.9%)</td>
</tr>
<tr>
<td>T-Spot only</td>
<td>21</td>
<td>2 (9.5%)</td>
<td>8 (38.1%)</td>
<td>11 (52.4%)</td>
</tr>
<tr>
<td>TST only</td>
<td>15</td>
<td>0 (0%)</td>
<td>2 (13.3%)</td>
<td>13 (86.7%)</td>
</tr>
<tr>
<td>Two tests positive</td>
<td>10</td>
<td>0 (0%)</td>
<td>7 (70.0%)</td>
<td>3 (30.0%)</td>
</tr>
<tr>
<td>TST &amp; QFT-GIT</td>
<td>1</td>
<td>0 (0%)</td>
<td>1 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>TST &amp; T-Spot</td>
<td>5</td>
<td>0 (0%)</td>
<td>5 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>QFT-GIT &amp; T-Spot</td>
<td>4</td>
<td>0 (0%)</td>
<td>1 (25.0%)</td>
<td>3 (75.0%)</td>
</tr>
<tr>
<td>All three tests positive</td>
<td>10</td>
<td>1 (10%)</td>
<td>7 (70.0%)</td>
<td>2 (20.0%)</td>
</tr>
</tbody>
</table>

* $\chi^2$ (trend) for low risk compared to the other groups = 61.6, p<0.0001
† Risk stratification as per CDC Risk Stratified Index(29)
TB = tuberculosis
TST = tuberculin skin test
T-Spot = T-SPOT®.TB test
QFT-GIT = QuantiFERON®-TB Gold In-Tube test
**ONLINE Supplement:**

**Information on the Battey Skin Test Antigen**

PPD-B was prepared from the same stock concentrate used in NHANES and Navy surveys (8, 12), having been obtained from the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration (FDA). The product was prepared in standard 0.1 mcg/mL doses by the Aeras Global TB Vaccine Foundation in Rockville, MD. After dilution and filling of vials, animal testing for potency and general safety was performed according to FDA regulations (46, 47).
Supplemental Figure 1. Distribution of Tuberculin Skin Test Reaction Sizes

Note: Tuberculin skin test (TST) readings in millimeters (mm). 1863 skin test readings had a reading of 0 mm, and 1 had a reading of 80 mm; these were excluded from the figure to improve clarity.
Supplemental Figure 2. Battey skin test (BST) readings in millimeters (mm). 1509 skin test readings had a reading of 0 mm; these were excluded from the figure to improve clarity.
Supplemental Table 1. Inter-rater Agreement of the TST§

<table>
<thead>
<tr>
<th></th>
<th>TST Reader 2 positive</th>
<th>TST Reader 2 negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TST Reader 1 positive</td>
<td>35 (10.6%)</td>
<td>9 (2.7%)</td>
<td>44 (13.3%)</td>
</tr>
<tr>
<td>TST Reader 1 negative</td>
<td>7 (2.1%)</td>
<td>280 (84.6%)</td>
<td>287 (86.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>42 (12.7%)</td>
<td>289 (87.3%)</td>
<td>331</td>
</tr>
</tbody>
</table>

TST= tuberculin skin test
§ Includes 331 subjects performed as a sample to assess inter-rater reliability of the TST
% agreement = 95.2%
K (95% CI) =0.79 (0.69, 0.89)
Supplemental Figure 3. TB Response by QuantiFERON®-TB Gold In-Tube vs. TST Reaction Size in mm
Supplemental Figure 4. TB Response by T-SPOT®.TB vs. TST Reaction Size in mm
Supplemental Figure 5. Battey Skin Test (BST) Reaction Size vs. Tuberculin Skin Test (TST) Reaction Size (both in mm). The line drawn is the line of equivalence where TST=BST.