Optimal Testing Choice and Diagnostic Strategies for Latent Tuberculosis Infection Among US-Born People Living with Human Immunodeficiency Virus (HIV)

April C. Pettit,[1] Jason E. Stout,[2] Robert Belknap,[3] Constance A. Benson,[4] Marie Nancy Séraphin,[5] Michael Lauzardo,[5] David J. Horne,[6] Richard S. Garfein,[4] Fernanda Maruri,[1] and Christine S. Ho[7]; for the Tuberculosis Epidemiologic Studies Consortium (TBESC)

[1]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [2]Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA, [3]Denver Metro TB Clinic, Denver Health and Hospital Authority, Denver, Colorado, USA, [4]Department of Medicine, University of California San Diego School of Medicine, San Diego, California, USA, [5]Department of Medicine, University of Florida College of Medicine, Gainesville, Florida, USA, [6]Department of Medicine, University of Washington, Seattle, Washington, USA, and [7]Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

(See the Editorial Commentary by MacLean and Dendukuri on pages e2285–6.)

Background. Increased risk of progression from latent tuberculosis infection (LTBI) to tuberculosis (TB) disease among people living with human immunodeficiency virus (HIV; PLWH) prioritizes them for LTBI testing and treatment. Studies comparing the performance of interferon gamma release assays (IGRAs) and the tuberculin skin test (TST) among PLWH are lacking.

Methods. We used Bayesian latent class analysis to estimate the prevalence of LTBI and diagnostic characteristics of the TST, QuantiFERON Gold In-Tube (QFT), and T.SPOT-TB (TSPOT) among a prospective, multicenter cohort of US-born PLWH ≥5 years old with valid results for all 3 LTBI tests using standard US cutoffs (≥5 mm TST, ≥0.35 IU/mL QFT, ≥8 spots TSPOT). We also explored the performance of varying LTBI test cutoffs.

Results. Among 1510 PLWH (median CD4+ count 532 cells/mm³), estimated LTBI prevalence was 4.7%. TSPOT was significantly more specific (99.7%) and had a significantly higher positive predictive value (90.0%, PPV) than QFT (96.5% specificity, 50.7% PPV) and TST (96.8% specificity, 45.4% PPV). QFT was significantly more sensitive (72.2%) than TST (54.2%) and TSPOT (51.9%); negative predictive value of all tests was high (TST 97.7%, QFT 98.6%, TSPOT 97.6%). Even at the highest cutoffs evaluated (15 mm TST, ≥1.00 IU/mL QFT, ≥8 spots TSPOT), TST and QFT specificity was significantly lower than TSPOT.

Conclusions. LTBI prevalence among this cohort of US-born PLWH was low compared to non-US born persons. TSPOT's higher PPV may make it preferable for testing US-born PLWH at low risk for TB exposure and with high CD4+ counts.

Keywords. human immunodeficiency virus (HIV); latent tuberculosis infection (LTBI); latent class analysis (LCA); tuberculin skin test (TST); interferon gamma release assay (IGRA).

Progress toward tuberculosis (TB) elimination in the United States will require retooling of current strategies [1]. Modeling studies demonstrate that identification and treatment of individuals with latent TB infection (LTBI) will have the greatest impact on TB elimination [2, 3]. Because human immunodeficiency virus (HIV) infection has historically been shown to greatly increase the risk of progression from LTBI to TB disease, people living with HIV (PLWH) are prioritized for LTBI testing and treatment [4]. Understanding LTBI test performance is vital to optimizing strategies for preventing TB disease in this

population. Currently, the tuberculin skin test (TST) and 3 interferon gamma release assays (IGRAs) are approved for LTBI testing in the United States; IGRAs include the QuantiFERON (QuantiFERON Gold-in-Tube [QFT], QuantiFERON-TB Gold Plus [QFT-Plus]), and T-SPOT.TB (TSPOT) [5].

No direct tests for LTBI, and therefore no gold standards, are available with which to compare LTBI test characteristics. Latent class analysis (LCA) is a statistical technique that provides an understanding of test characteristics when no gold standard is available [6–8]. It ideally involves ≥3 types of tests for each person and uses the observed test result patterns to calculate the prevalence of the underlying condition, which otherwise cannot be directly observed, along with the test sensitivity and specificity.

This study was conducted by the Centers for Disease Control and Prevention (CDC)-supported Tuberculosis Epidemiologic Studies Consortium (TBESC), a multisite consortium of academic medical centers and local TB programs in the United States. The objective of the parent study was to compare the ability of LTBI tests to predict progression to TB disease in

populations at high risk for LTBI and/or progression to TB disease. The primary objective of this analysis was to estimate the prevalence of LTBI and diagnostic test characteristics of LTBI tests among US-born PLWH enrolled into the parent study. We also explored the impact of varying LTBI test cutoffs and LTBI prevalence estimates on diagnostic test characteristics.

METHODS

Study Population
From July 2012 to April 2017, TBESC-affiliated clinics in 11 states prospectively enrolled participants at high risk for LTBI or progression to TB disease. At baseline, participants were tested with TST, QFT, and TSPOT and evaluated for TB disease. Institutional Review Board (IRB) approval was obtained from the CDC, as well as all local study sites that did not rely on CDC for IRB review. All participants provided written informed consent or assent with parental permission for minors. The study was registered at clinicaltrials.gov (identifier NCT01622140). A detailed description of this cohort has been published previously [9].

The current analysis is restricted to US-born PLWH ≥5 years of age with valid results for all 3 tests at enrollment. Individuals with missing LTBI test results, indeterminate or failed QFT results, or invalid TSPOT results were excluded. Non-US born PLWH were excluded from this analysis because (1) small numbers of non-US born PLWH would lead to highly unstable estimates and (2) test characteristics are fundamentally different (ie, lower TST specificity) in the setting of BCG vaccination.

Data Collection/Study Definitions
Study staff received standardized training on specimen collection for QFT and TSPOT. QFT processing was performed using standard protocols at each TBESC site; a central laboratory was not utilized for pragmatic reasons although quality assurance site visits were conducted to ensure compliance with study and manufacturer guidelines. We used the manufacturer's cutpoint for QFT (≥0.35 international units/milliliter [IU/mL] defined as positive) [10]. TSPOT specimens were processed by Oxford Immunotec (Memphis, TN, USA). We used the manufacturer's US cutpoint for the TSPOT (≥8 spots positive); borderline results (5, 6, or 7 spots) and negative results (≤4 spots) were treated as negative in these analyses [11]. We used CDC guidelines for TST interpretation among PLWH (≥5 mm of induration positive) [4].

Self-reported demographics, HIV status, and CD4+ lymphocyte counts [12, 13] were collected through baseline participant interviews. Race/ethnicity was categorized as non-Hispanic Black, non-Hispanic White, Hispanic, and Other (multiple races, Asian, Pacific Islander, Native American, and other races/ethnicities not listed as an option).

Direct Bayesian Latent Class Analysis
We used LCA to directly estimate the prevalence of LTBI and the test diagnostic characteristics in our study population. Because all 3 tests use immunologically overlapping antigens, we used a modification of the method of Qu et al, including a random effect to account for conditional dependence of the tests [8]; this method provides similar results compared to other methods of modeling conditional dependence among LTBI tests [14]. We used a Bayesian approach with literature-based prior distributions for test sensitivities and broad prior distributions for specificity and prevalence [15]. Full details were published previously [9].

The mean difference in the diagnostic test characteristics between test pairs and associated 95% credible intervals (CrI) were calculated using the distributions of the parameter differences obtained from the Markov Chain Monte Carlo iterations (ie, sampling the posterior parameter distributions). A mean difference 95% CrI that does not include zero was considered statistically significant. Model cross-validation was performed by comparing the frequency of observed to model-predicted LTBI test patterns [15]; full details can be found in Supplementary Table 1.

Indirect Bayesian Latent Class Analysis
The estimated sensitivity and specificity of all 3 tests at varying test cutoffs and the receiver operating characteristic (ROC) curves were generated indirectly from the Bayesian LCA model using the PPV of each test combination. These sensitivity and specificity estimates were used to quantify the under- and overdiagnosis of LTBI per 1000 persons screened using varying LTBI prevalence estimates. We evaluated the following cutoffs: ≥5 mm, ≥10 mm, and ≥15 mm for TST; ≥0.35, ≥0.70, and ≥1.00 IU/mL for QFT; ≥5, ≥6 (international cutoff) [16], ≥7, and ≥8 (US cutoff) [11] spots for TSPOT. Underdiagnosed LTBI cases were defined as the number of falsely negative LTBI diagnoses missed per 1000 screened; overdiagnosed LTBI cases were defined as the number of false positive LTBI diagnoses per 1000 screened. Area under the ROC curves (AUCs) were calculated; 95% confidence intervals (CI) for AUCs were obtained by sampling 1000 iterations of the posterior distribution.

RESULTS

There were 1510 US-born, PLWH ≥5 years of age with valid results for all 3 tests included; median age was 49 years (interquartile range [IQR] 42–55 years), 71% were cisgender male, 70% were Black, and the median self-reported CD4+ lymphocyte count at enrollment among those with nonmissing data was 532 cells/mm^3 (IQR: 355–764) (Table 1). We excluded 88 participants for missing test results (32 TST, 23 QFT, 31 TSPOT); there were some demographic differences between those included and excluded (Supplementary Table 2).

The most frequent combination was a negative result for all 3 tests. The second most frequent combination was an isolated

Table 1. Demographic Characteristics of the Study Population

Characteristic Total = 1510	Number (%) or Median (IQR)
Sex	
Cisgender male	1073 (71%)
Cisgender female	420 (28%)
Transgender	17 (1%)
Race/Ethnicity[a]	
Black, non-Hispanic	1057 (70%)
White, non-Hispanic	361 (24%)
Hispanic	92 (6%)
Other	89 (6%)
Age at enrollment (years)	1 (0.1%)
5–14	1 (0.1%)
15–24	52 (3.4%)
25–44	440 (29.1%)
45–64	982 (65.0%)
65+	35 (2.3%)
Self-reported CD4+ lymphocyte count (cells/mm^3)	
Missing	565 (37.4%)
<200 cells/mm^3	94 (10%)
≥200 cells/mm^3	851 (90%)
Median (IQR)	532 (355–764)
TBESC enrollment site	
Baltimore City TB Clinic (Baltimore, Maryland)	606 (40.1%)
Carolinas Medical Center TB Clinic (Charlotte, North Carolina)	9 (0.6%)
Dekalb County Board of Health (Decatur, Georgia)	78 (5.2%)
Denver Metro TB Clinic (Denver, Colorado)	14 (0.9%)
Fort Lauderdale TB Clinic (Ft. Lauderdale, Florida)	65 (4.3%)
Harborview Medical Center (Seattle, Washington)	1 (0.1%)
Lanakila Health Center (Honolulu, Hawaii)	2 (0.1%)
Maricopa County Department of Public Health (Phoenix, Arizona)	39 (2.6%)
Metro Nashville TB Clinic (Nashville, Tennessee)	143 (9.5%)
Alachua County Health Department (Gainesville, Florida)	72 (4.8%)
Pompano Beach TB Clinic (Pompano Beach, Florida)	22 (1.5%)
San Diego Department of Health (San Diego, California)	4 (0.3%)
San Francisco TB Clinic (San Francisco, California)	18 (1.2%)
Tarrant County Health Department (Fort Worth, Texas)	17 (1.1%)
University of California San Diego, Antiviral Research Center (San Diego, California)	292 (19.3%)
Wake County TB Clinic (Raleigh, North Carolina)	128 (8.5%)

Abbreviations: IQR, interquartile range; TB, tuberculosis; TBESC, Tuberculosis Epidemiologic Studies Consortium.

[a]Race/ethnicity does not add to 1510 as participants could select more than 1 race/ethnicity category.

Table 2. Frequency of Test Result Combinations

Test Combination (TST-QFT-TSPOT[a])	Count (N = 1510)	%
Triple negative (---)	1355	89.7
Isolated QFT positive (-+-)	58	3.8
Isolated TST positive (+--)	46	3.0
Triple positive (+++)	23	1.5
QFT positive and TSPOT positive (-++)	12	0.8
TST positive and QFT positive (++-)	12	0.8
Isolated TSPOT positive (--+)	4	0.3
TST positive and TSPOT positive (+-+)	0	0.0

Abbreviations: QFT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT.TB; TST, tuberculin skin test.

[a]We used the CDC guidelines for the interpretation of TST results (≥5 mm positive) and the manufacturer's US cutoffs for QFT (≥0.35 IU/mL) and TSPOT (≥8 spots positive). Borderline TSPOT was treated as negative.

Using standard US cutoffs for all 3 tests and an estimated LTBI prevalence of 4.7%, the estimated sensitivity and NPV was highest for the QFT; specificity and PPV were highest for TSPOT. The estimated sensitivity and PPV were lowest for the TST, specificity was lowest for the QFT, and NPV was lowest for the TSPOT (Table 3). QFT was significantly more sensitive than the TST and TSPOT; the sensitivity of the TST did not significantly differ from TSPOT. TSPOT was significantly more specific than the QFT and TST, with no significant difference between the specificities of the QFT and TST. Similarly, TSPOT had a significantly higher PPV than the QFT and TST; the difference in PPV between QFT and TST was not statistically significant. Finally, QFT had a significantly higher NPV than the TST and TSPOT; the difference in NPV between the TST and TSPOT was not statistically significant (Table 4). The goodness-of-fit χ^2 was 10.17, with 7 degrees of freedom ($P = .18$), suggesting a reasonable fit of the modeled to the observed data (Supplementary Table 1).

Diagnostic test characteristics were also estimated indirectly from the Bayesian LCA (Table 5). These differ slightly from those estimated directly from the Bayesian LCA model (Table 3) due to the relatively low frequencies of test combinations aside from a negative result for all 3 tests (Table 2), which, in turn, limits the precision of PPVs for these test combinations. In a low LTBI prevalence population (1%), LTBI overdiagnosis overshadowed underdiagnosis with the TST at the 5 mm and 10 mm cutoff (vs 15 mm), with QFT at all evaluated cutoffs, and with TSPOT at the 5 spot cutoff (vs higher cutoffs). In a medium LTBI prevalence population (5%), the lowest cutoffs for both the TST and QFT would result in more LTBI overdiagnosis than underdiagnosis (vs higher cutoffs), although TSPOT would result in more underdiagnosis at all cutoffs. In a high LTBI prevalence population (15%), the lowest cutoff for all 3 tests would yield more LTBI underdiagnosis than overdiagnosis (vs higher cutoffs) (Table 5). The ROC AUCs were 0.74 for TST (95% CI .69–.81), 0.89 for QFT (95% CI .81–.96), and 0.82 for TSPOT (95% CI .76–.90) (Figure 1).

positive QFT. An isolated positive TSPOT was the least frequent combination observed (there were no instances of a positive TST and TSPOT with a negative QFT) (Table 2).

Using Bayesian LCA, the estimated overall LTBI prevalence was 4.7% (95% CrI 3.2–6.7%); it ranged from 0.7% to 14.5% across TBESC sites. LTBI prevalence estimates also varied by sociodemographic factors. For example, LTBI prevalence by race was estimated to be 3.2% for non-Hispanic whites, 5.1% for non-Hispanic Blacks, and 3.6% for Hispanics (Supplementary Table 3).

Table 3. Diagnostic Test Characteristics for US-Born PLWH Estimated Directly From Latent Class Analysis Using Standard US Cutoffs at 4.7% Estimated LTBI Prevalence

	Sensitivity (95% CrI)	Specificity (95% CrI)	PPV (95% CrI)	NPV (95% CrI)
TST (≥5 mm)	54.2% (45.2–64.3)	96.8% (95.7–97.7)	45.4% (33.3–58.4)	97.7% (96.4–98.7)
QFT (≥0.35 IU/mL)	72.2% (58.7–85.4)	96.5% (95.3–97.6)	50.7% (37.1–65.6)	98.6% (97.4–99.4)
TSPOT (≥8 spots)	51.9% (39.3–66.7)	99.7% (99.3–99.9)	90.0% (77.1–98.1)	97.6% (96.1–98.8)

Abbreviations: CrI, credible interval; LTBI, latent tuberculosis infection; NPV, negative predictive value; PLWH, people living with human immunodeficiency virus; PPV, positive predictive value; QFT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT.TB; TST, tuberculin skin test.

DISCUSSION

In this US-born population of PLWH, the estimated LTBI prevalence using Bayesian LCA was 4.7% (95% CrI 3.2–6.7%). An isolated positive QFT was the second most frequently observed test combination (behind 3 negative tests) consistent with the estimated low specificity and PPV of QFT. Although TSPOT was only slightly more specific than the TST (specificity difference 3.0%) and QFT (specificity difference 3.2%), this small specificity difference translates into a large PPV difference given the low LTBI prevalence in this population (PPV difference 44.5% and 39.3% between TSPOT with TST and QFT, respectively). The higher PPV of TSPOT compared with TST and QFT may make it preferable for screening PLWH with relatively high CD4+ counts in low-risk settings.

Current US clinical practice guidelines for LTBI testing recommend taking into account an individual's likelihood of LTBI, likelihood of progression to TB disease, and the benefit of therapy when making testing decisions [17]. PLWH are identified in these guidelines as a group with high likelihood of progression to TB disease who would benefit from therapy, so baseline LTBI testing is recommended [18]. However, epidemiologic risk of TB exposure and immune status (CD4+ counts) are not considered in this risk classification. When LTBI testing occurs among low risk groups, guidelines recommend that all initial positive LTBI tests should by confirmed by a second LTBI test and be considered infected only if both tests are positive. This may also be a reasonable strategy for PLWH at otherwise low risk for TB exposure who have high CD4+ counts, particularly if the first test was a TST or QFT, given their low PPV. Previous studies have demonstrated the cost-effectiveness of a sequential testing strategy in other populations [19, 20], although additional prospective economic studies are needed to determine the incremental cost-effectiveness of such a strategy among US-born PLWH.

The US Food and Drug Administration required a borderline result zone (5, 6, and 7 spots) when it approved TSPOT in 2008. This classification was meant to address test result variability around the cutpoint, and because this range of spot counts represented an area of overlap between patients with culture-confirmed TB and low-risk patients without TB in preclinical studies [21]. The borderline designation was not required for marketing approvals in other countries in which ≥6 spots is positive and ≤5 spots is negative. In our analysis, using the TSPOT international cutoff (≥6 spots) and US cutoff (≥8 spots) resulted in more LTBI underdiagnosis than overdiagnosis, regardless of LTBI prevalence. Assuming a LTBI prevalence of 5% for every 1000 PLWH screened, using the international cutoff in place of the US cutoff would maintain more underdiagnosis than overdiagnosis, although 2 more overdiagnosed cases of LTBI and 1 fewer underdiagnosed case of LTBI would result. Although our study provides quantitative estimates of under- and overdiagnosis, the clinical consequences of underdiagnosis (potential progression to TB disease) must be weighed against the consequences of overdiagnosis (unnecessary LTBI treatment) in the setting of a patient's epidemiologic risk and immune status (CD4+ count). Regardless, these data suggest that the international and US cutoffs may be equivalent and that interpreting a TSPOT result ≥6 spots as positive may be a reasonable clinical decision when testing US-born PLWH residing in the United States.

Table 4. Differences in Diagnostic Test Characteristics for US-Born PLWH Estimated Directly From Latent Class Analysis Using Standard US Cutoffs at 4.7% Estimated LTBI Prevalence

Test Combination	Difference (95% CrI)
Sensitivity:	
TST vs QFT	**−18.0% (−31.8 to −4.1)**
TST vs TSPOT	2.2% (−11.5 to 15.2)
QFT vs TSPOT	**20.2% (5.2 to 35.3)**
Specificity:	
TST vs QFT	0.2% (−1.2 to 1.7)
TST vs TSPOT	**−3.0% (−4.1 to −2.0)**
QFT vs TSPOT	**−3.2% (−4.5 to −2.0)**
PPV:	
TST vs QFT	−5.3% (−18.8 to 7.8)
TST vs TSPOT	**−44.5% (−58.5 to −29.1)**
QFT vs TSPOT	**−39.3% (−54.5 to −22.4)**
NPV:	
TST vs QFT	**−0.9% (−1.7 to −.2)**
TST vs TSPOT	0.1% (−.6 to .7)
QFT vs TSPOT	**0.9% (.2 to 1.9%)**

Bold font indicates a statistically significant mean difference.

Abbreviations: CrI, credible interval; LTBI, latent tuberculosis infection; NPV, negative predictive value; PLWH, people living with human immunodeficiency virus; PPV, positive predictive value; QFT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT.TB; TST, tuberculin skin test.

Table 5. Estimated Test Characteristics and Frequency of Under/Overdiagnosis of LTBI Using Varying Prevalence and Test Cutoffs[a]

	TST			QFT			TSPOT			
	≥5 mm	≥10 mm	≥15 mm	≥0.35 IU/mL	≥0.70 IU/mL	≥1.00 IU/mL	≥5 spots	≥6 spots	≥7 spots	≥8 spots
Sensitivity	51.7	44.8	27.8	77.3	57.1	50.6	56.7	53.2	51.7	51.6
Specificity	97.0	97.8	99.4	96.7	98.0	98.4	98.9	99.4	99.6	99.8
1% LTBI Prevalence										
Underdiagnosed LTBI	4.8	5.5	7.2	2.3	4.3	4.9	4.3	4.7	4.8	4.8
Overdiagnosed LTBI	30.4	22.1	5.6	32.8	19.1	15.4	8.5	4.0	3.6	2.0
5% LTBI prevalence										
Underdiagnosed LTBI	24.2	27.6	36.1	11.3	21.5	24.7	21.7	23.4	24.2	24.2
Overdiagnosed LTBI	29.2	21.2	5.4	31.7	18.7	15.2	8.2	3.8	3.4	1.9
15% LTBI prevalence										
Underdiagnosed LTBI	72.5	82.8	108.2	34.0	64.4	74.0	65.0	70.1	72.5	72.5
Overdiagnosed LTBI	26.1	19.0	4.8	28.4	16.8	13.6	9.3	5.4	3.7	1.3

Abbreviations: LTBI, latent tuberculosis infection; QFT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT; TB; TST, tuberculin skin test.

[a]Estimates of diagnostic test characteristics differ slightly from Table 3 due to differences in analytic techniques as described in the methods section. Underdiagnosed LTBI cases: Number of LTBI diagnoses missed per 1000 screened Overdiagnosed LTBI cases: Number of LTBI misdiagnoses per 1000 screened.

For both the TST and QFT, current US cutoffs (5 mm for TST and 0.35 IU/mL for QFT) resulted in more LTBI overdiagnosis than underdiagnosis in a population with medium (5%) LTBI prevalence (similar to our estimated prevalence). In a very low prevalence population (1%), higher cutoffs for the TST (10 mm) and QFT (1.00 IU/mL) still resulted in more LTBI overdiagnosis than underdiagnosis. Conversely, in a high prevalence population (15%), all cutoffs resulted in more LTBI underdiagnosis than overdiagnosis. These findings underscore the importance of understanding the local prevalence of the population being tested as this may vary geographically, as was the case for our study population (.7–14.5% by site).

LTBI prevalence in the US general population has been estimated using National Health and Nutrition Examination Survey (NHANES) data [22]. Among US-born persons, the estimated LTBI prevalence was 1.5% (95% CI .9–2.6%) using TST and 2.8% (95% CI 2.0–3.8%) using QFT. Among non-US born persons, the estimated LTBI prevalence was 20.5% (95% CI 16.1–25.8%) using TST and 15.9% (95% CI 13.5–18.7%) using QFT. There were too few PLWH with LTBI to make reliable prevalence estimates in NHANES. Our data suggest that the prevalence of LTBI among US-born PLWH in the United States may be higher than the overall US-born population but still relatively low compared to non-US born residents.

LTBI prevalence estimates from NHANES differed by sociodemographic characteristics [22]. Notably, LTBI prevalence estimates among the US-born were highest for non-Hispanic Blacks at 5.1% (95% CI 3.6–7.3%) using TST and 4.4% (95% CI 3.2–6.0%) using QFT. In our population of PLWH, 70% were Black, which is higher than the proportion of PLWH who are Black nationally (42%) [23]. Similar to NHANES, we found that LTBI prevalence estimates varied by race with Black non-Hispanics having the highest LTBI prevalence. If the racial/ethnic breakdown of our study population were to mirror the general US-born population of PLWH, the estimated prevalence of LTBI in our study population would likely be lower with associated larger changes in PPV.

Some limitations in this study should be considered. First, we evaluated the QFT Gold In-Tube assay, as the QFT-Plus assay had not yet been FDA approved until our study enrollment was completed. QFT-Plus incorporates a fourth tube with shorter TB peptides intended to stimulate CD8+ lymphocytes (in addition to longer TB peptides intended to stimulate CD4+ lymphocytes). Several studies have demonstrated strong agreement between QFT and QFT-Plus, including among immunocompromised patients [24–26]. These data suggest that our findings would be similar using QFT-Plus in place of QFT.

Second, our study population is not reflective of the general US-born population of PLWH as indicated by racial/ethnicity differences between our study population (70% Black, non-Hispanic) and the general US-born population of PLWH (42% Black, non-Hispanic) [23]. This is likely due to the urban location of TBESC sites. Although the estimated LTBI prevalence in this US-born population of PLWH was 4.7%, LTBI prevalence differed across TBESC sites. Given that prevalence helps determine the predictive value of diagnostic tests, it is critically important to understand local LTBI prevalence when interpreting diagnostic test results.

Third, data on HIV status was collected by self-report, so only people with diagnosed HIV infection were included. Similarly, CD4+ counts were self-reported, leading to CD4+ data that were not missing at random. However, studies have shown that self-report is a reliable and valid means of CD4+ count collection [12, 13]. Among those with known self-reported CD4+ counts, the median was relatively high (532 cells/mm³); therefore, our results and conclusions may not be generalizable to PLWH and low CD4+ counts, a population in which LTBI
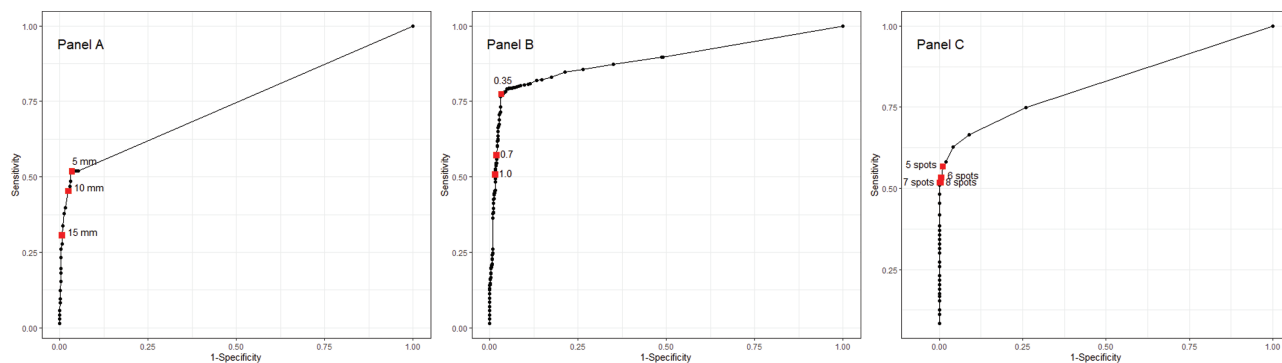
Figure 1. Receiver operator curves for varying test cutoff values. *A*, TST (5 mm, 10 mm, 15 mm cutoffs); *B*, QFT (0.35 IU, 0.70 IU, 1.00 IU cutoffs); *C*, TSPOT (5 spot, 6 spot, 7 spot, 8 spot cutoffs). Abbreviations: QFT, QuantiFERON Gold In-Tube; TSPOT, T.SPOT-TB; TST, tuberculin skin test.

diagnostic tests may perform differently [27]. Data on antiretroviral therapy (ART) and viral suppression were not available.

Fourth, we excluded patients without valid results for all 3 tests. It is likely that test results were not missing at random, making standard imputation procedures problematic. Although the number excluded was relatively small compared with the study sample size, it is possible that results may have differed slightly if complete test results were available.

Strengths of this study include leveraging data from a large prospective cohort of US-born PLWH enrolled across the United States. The availability of results from three LTBI tests allowed us to utilize LCA as a novel approach to estimating LTBI prevalence and evaluating diagnostic characteristics in the absence of a gold standard. An additional strength is the pragmatic nature of the study given that local site staff placed/interpreted TSTs and local laboratories performed QFT testing.

CONCLUSIONS

Among this large population of PLWH in the United States, the estimated prevalence of LTBI using Bayesian LCA is higher than the general US-born population although still quite low compared to non-US born residents. The NPV of all 3 tests were all very high and similar to each other, although the PPV varied widely. The significantly higher PPV of the TSPOT in this population may make it preferable to the TST and QFT for screening US-born PLWH in low-risk settings. However, following an initial positive LTBI test (particularly an initial TST or QFT) with a second confirmatory LTBI test may also be sensible for PLWH at low risk for TB exposure and with high CD4+ counts, similar to recommendations for other low TB risk populations. Among PLWH in the United States, the international TSPOT cutpoint of ≥6 spots maintains more LTBI underdiagnosis than overdiagnosis compared to higher spot counts and may therefore be a reasonable cutpoint in this population. The local LTBI prevalence of the population being tested is critical to interpreting the predictive value of diagnostic tests. Additional prospective studies and cost-effectiveness analyses are needed to support these recommendations.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

References

1. Talwar A, Tsang CA, Price SF, et al. Tuberculosis - United States, 2018. Am J Transplant **2019**; 19:1582–8.
2. Hill AN, Becerra J, Castro KG. Modelling tuberculosis trends in the USA. Epidemiol Infect **2012**; 140:1862–72.
3. Menzies NA, Cohen T, Hill AN, et al. Prospects for tuberculosis elimination in the United States: results of a transmission dynamic model. Am J Epidemiol **2018**; 187:2011–20.
4. American Thoracic Society and Centers for Disease Control and Prevention Statement Committee on Latent Tuberculosis Infection. Targeted tuberculin testing and treatment of latent tuberculosis infection. American Thoracic Society. MMWR Recomm Rep **2000**; 49(RR-6):1–51.
5. Mazurek GH, Jereb J, Vernon A, LoBue P, Goldberg S, Castro K; IGRA Expert Committee; Centers for Disease Control and Prevention (CDC). Updated guidelines for using interferon gamma release assays to detect *Mycobacterium tuberculosis* infection - United States, 2010. MMWR Recomm Rep **2010**; 59:1–25.
6. Johnson WO, Jones G, Gardner IA. Gold standards are out and Bayes is in: Implementing the cure for imperfect reference tests in diagnostic accuracy studies. Prev Vet Med **2019**; 167:113–27.
7. Formann AK, Kohlmann T. Latent class analysis in medical research. Stat Methods Med Res **1996**; 5:179–211.
8. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics **1996**; 52:797–810.
9. Stout JE, Wu Y, Ho CS, et al; Tuberculosis Epidemiologic Studies Consortium. Evaluating latent tuberculosis infection diagnostics using latent class analysis. Thorax **2018**; 73:1062–70.
10. Qiagen. QuantiFERON-TB Gold (QFT) ELISA Package Insert. Germantown, MD, U.S. March 2018. Available at: https://www.quantiferon.com/us/wpcontent/uploads/sites/13/2019/03/L1075116-QuantiFERON-TB-Gold-ELISA-IFU-FDA-rev04.pdf. Accessed 25 June 2019.
11. Oxford Immunotec. T-SPOT.TB Package Insert U.S. version 5. Marlborough, MA, US, March 2015. Available at: http://www.tspot.com/wp-content/uploads/2012/01/PI-TB-US-v5.pdf. Accessed 25 June 2019.
12. Kalichman SC, Rompa D, Cage M. Reliability and validity of self-reported CD4 lymphocyte count and viral load test results in people living with HIV/AIDS. Int J STD AIDS **2000**; 11:579–85.
13. Cunningham WE, Rana HM, Shapiro MF, Hays RD. Reliability and validity of self-report CD4 counts-in persons hospitalized with HIV disease. J Clin Epidemiol **1997**; 50:829–35.
14. Ling DI, Pai M, Schiller I, Dendukuri N. A Bayesian framework for estimating the incremental value of a diagnostic test in the absence of a gold standard. BMC Med Res Methodol **2014**; 14:67.
15. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics **2001**; 57:158–67.
16. Oxford Immunotec. T-SPOT.TB Package Insert UK version 3. Abingdon, Oxfordshire, UK. **2016**. Available at: http://www.oxforddiagnosticlaboratories.eu/wp-content/media/PI-TB-IVD-UK-v3.pdf. Accessed 25 June 2019.
17. Lewinsohn DM, Leonard MK, LoBue PA, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: diagnosis of tuberculosis in adults and children. Clin Infect Dis **2017**; 64:111–5.
18. Panel on Opportunistic Infections in Adults and Adolescents with HIV. Guidelines for the prevention and treatment of opportunistic infections in adults and adolescents with HIV: recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. Available at http://aidsinfo.nih.gov/contentfiles/lvguidelines/adult_oi.pdf. Accessed 12 February 2020.
19. Tasillo A, Salomon JA, Trikalinos TA, Horsburgh CR Jr, Marks SM, Linas BP. Cost-effectiveness of testing and treatment for latent tuberculosis infection in residents born outside the United States with and without medical comorbidities in a simulation model. JAMA Intern Med **2017**; 177:1755–64.
20. Abubakar I, Lalvani A, Southern J, et al. Two interferon gamma release assays for predicting active tuberculosis: the UK PREDICT TB prognostic test study. Health Technol Assess **2018**; 22:1–96.
21. U.S. Food and Drug Administration. Summary of Safety and Effectiveness. **2008**. Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf7/P070006b.pdf. Accessed 25 June 2019.
22. Miramontes R, Hill AN, Yelk Woodruff RS, et al. Tuberculosis infection in the United States: prevalence estimates from the national health and nutrition examination survey, 2011–2012. PLoS One **2015**; 10:e0140881.
23. Centers for Disease Control and Prevention. HIV Surveillance Report, **2017**; vol. 29. Available at: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Published November 2018. Accessed 25 June 2019.
24. Ryu MR, Park MS, Cho EH, et al. Comparative evaluation of QuantiFERON-TB gold in-tube and QuantiFERON-TB gold plus in diagnosis of latent tuberculosis infection in immunocompromised patients. J Clin Microbiol **2018**; 56:e00438–18.
25. Theel ES, Hilgart H, Breen-Lyles M, et al. Comparison of the QuantiFERON-TB gold plus and QuantiFERON-TB gold in-tube interferon gamma release assays in patients at risk for tuberculosis and in health care workers. J Clin Microbiol **2018**; 56:e00614–18.
26. Moon HW, Gaur RL, Tien SS, Spangler M, Pai M, Banaei N. Evaluation of QuantiFERON-TB gold-plus in health care workers in a low-incidence setting. J Clin Microbiol **2017**; 55:1650–7.
27. Cattamanchi A, Smith R, Steingart KR, et al. Interferon-gamma release assays for the diagnosis of latent tuberculosis infection in HIV-infected individuals: a systematic review and meta-analysis. J Acquir Immune Defic Syndr **2011**; 56:230–8.