ORIGINAL ARTICLE

Evaluating latent tuberculosis infection diagnostics using latent class analysis

Jason E Stout,[1] Yanjue Wu,[2] Christine S Ho,[3] April C Pettit,[4] Pei-Jean Feng,[3] Dolly J Katz,[3] Smita Ghosh,[3] Thara Venkatappa,[3] Ruiyan Luo,[5] on behalf of the Tuberculosis Epidemiologic Studies Consortium

[1]Division of Infectious Diseases and International Health, Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA
[2]Northrop Grumman, McLean, Virginia, USA
[3]Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[4]Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[5]Department of Epidemiology and Biostatistics, Georgia State University, School of Public Health, Atlanta, Georgia, USA

Correspondence to
Dr Jason E Stout, Division of Infectious Diseases and International Health, Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA; stout002@mc.duke.edu

ABSTRACT

Background Lack of a gold standard for latent TB infection has precluded direct measurement of test characteristics of the tuberculin skin test and interferon-γ release assays (QuantiFERON Gold In-Tube and T-SPOT. TB).

Objective We estimated test sensitivity/specificity and latent TB infection prevalence in a prospective, US-based cohort of 10 740 participants at high risk for latent infection.

Methods Bayesian latent class analysis was used to estimate test sensitivity/specificity and latent TB infection prevalence among subgroups based on age, foreign birth outside the USA and HIV infection.

Results Latent TB infection prevalence varied from 4.0% among foreign-born, HIV-seronegative persons aged <5 years to 34.0% among foreign-born, HIV-seronegative persons aged ≥5 years. Test sensitivity ranged from 45.8% for the T-SPOT.TB among foreign-born, HIV-seropositive persons aged ≥5 years to 80.7% for the tuberculin skin test among foreign-born, HIV-seronegative persons aged ≥5 years. The skin test was less specific than either interferon-γ release assay, particularly among foreign-born populations (eg, the skin test had 70.0% specificity among foreign-born, HIV-seronegative persons aged ≥5 years vs 98.5% and 99.3% specificity for the QuantiFERON and T-SPOT. TB, respectively). The tuberculin skin test's positive predictive value ranged from 10.0% among foreign-born children aged <5 years to 69.2% among foreign-born, HIV-seropositive persons aged ≥5 years; the positive predictive values of the QuantiFERON (41.4%) and T-SPOT.TB (77.5%) were also low among US-born, HIV-seropositive persons aged ≥5 years.

Conclusions These data reinforce guidelines preferring interferon-γ release assays for foreign-born populations and recommending against screening populations at low risk for latent TB infection.

Trial registration number NCT01622140.

Key messages

What is the key question?
► How do the three existing tests for latent TB perform in a large cohort of persons at high risk in a low-incidence country using latent class analysis, given that there is no gold standard test for latent TB?

What is the bottom line?
► Test sensitivity was moderate at best; specificity was good for the interferon-γ release assays but moderate for the tuberculin skin test, resulting in poor test positive predictive values in key populations including young children and persons infected with the HIV.

Why read on?
► This is the first study that directly examines the test characteristics of the three existing tests for latent TB without using imperfect surrogate markers and provides key data for important populations such as young children.

INTRODUCTION

Targeted testing and treatment of latent TB infection (LTBI) is a key component of TB elimination efforts in countries with a low incidence of TB disease. However, LTBI diagnosis is hampered by the imperfect nature of the three existing diagnostic tests: the Mantoux tuberculin skin test (TST), the QuantiFERON Gold In-Tube (Qiagen, Germantown, Maryland) and the T-SPOT.TB (Oxford Immunotec, Marlborough, Massachusetts). A

substantial challenge to understanding how to best use these tests in clinical practice is the absence of a gold standard for LTBI that indicates infection with certainty. This absence prevents both accurate assessment of test characteristics (sensitivity and specificity) and accurate evaluation of LTBI prevalence among populations of interest. Evaluation of test performance has employed: (A) the proportion of positive tests among patients with confirmed TB disease as a measure of sensitivity, (B) the proportion of negative results among populations at low risk as a measure of specificity or (C) the correlation between exposure to infectious TB and the likelihood of a positive test result.[1–3] These surrogates are all imperfect at best and misleading at worst. For example, the ability for the immune system to respond to TB antigens, which must occur to have a positive LTBI test, might differ for someone with TB disease, with actively replicating bacilli, compared with LTBI,[4 5] with quiescent dormant bacilli. Similarly, persons who are at low risk for LTBI reside in low-incidence areas and are often socioeconomically and demographically different from persons residing in high-incidence areas. These differences might be associated with differential exposure to non-tuberculous mycobacteria and *Mycobacterium*

leprae, both of which have potential to cross-react with tests for LTBI and affect the measured specificity of the tests.[6][7] Studies that examine the association between a positive test result and subsequent risk for progression to TB disease are the best gold standard available, although even that type of study is imperfect, limited by low event rates and confounded by differential acceptance of LTBI treatment.[8–10]

Latent class analysis (LCA) is a statistical technique that offers an alternative method for understanding test characteristics when no gold standard is available. LCA has been used extensively in the social sciences but is now increasingly used to examine diagnostic tests in medicine.[11] It uses the observed patterns of test results to calculate the prevalence of the underlying condition, which otherwise cannot be directly observed, as well as the sensitivity and specificity of the tests. We used LCA to understand the three available tests for LTBI among a cohort of persons at high risk for LTBI.

METHODS

We examined a cohort of participants enrolled in a prospective study to assess the predictive ability of available tests for LTBI. The Centers for Disease Control and Prevention (CDC) funded the study through the Tuberculosis Epidemiologic Studies Consortium (TBESC), a partnership of academic institutions and TB control programmes in 11 US states. Sixteen TBESC-affiliated clinics enrolled children and adults at high risk for LTBI and tested them concurrently with a TST, a QuantiFERON Gold In-Tube (QFT-IT) and a T-SPOT.TB (TSPOT) test; a 17th clinic used only TST and QFT-IT because TSPOT was not initially available. All participants had at least one of the following risk factors for LTBI or progression to TB disease: (1) close contact with an infectious TB patient; (2) recent immigration (≤5 years) from a country with moderate rates of TB (eg, Mexico); (3)

immigration at any time from a country with high rates of TB (eg, India); (4) recent (≤5 years) residence for ≥30 days in a country with high rates of TB; (5) member of a group with high (≥25%) local prevalence of LTBI (eg, homeless persons); or (6) a diagnosis of HIV infection.

Trained study personnel collected demographic and LTBI-related risk information at participant enrolment. All participants were evaluated for TB disease at time of enrolment and will be followed up at 6 month intervals for 2 years or until diagnosis of TB; matching of participant names to state TB registries will continue through 2021. Participants found to have TB at the time of enrolment were excluded from this analysis. All participants provided written informed consent, assent or parental permission. The study was registered at clinicaltrials.gov (identifier NCT01622140).

Study enrolment began on 20 July 2012 and ended in April 2017. This LCA includes participants enrolled during 20 July 2012–8 September 2014. A flow chart of enrolled participants is included in figure 1, and comparison with excluded participants is in online supplementary table S1.

Testing for measurement invariance among groups was performed with PROC LCA[12] in SASV.9.4. On the basis of published data,[1][13][14] a strong belief existed that test characteristics would differ among the following groups: (A) HIV-seropositive, compared with HIV-seronegative; (B) foreign-born persons, compared with US-born (born in US states and territories) persons; and (C) aged ≥5 years, compared with aged <5 years. We tested for measurement invariance across these groups by sequentially comparing a model that assumed measurement invariance with one that permitted the ρ parameters (ie, sensitivity and specificity) to vary freely, examining the differences in the G^2 values (a measurement of goodness-of-fit) for each model. We assumed that the differences follow a χ^2 distribution
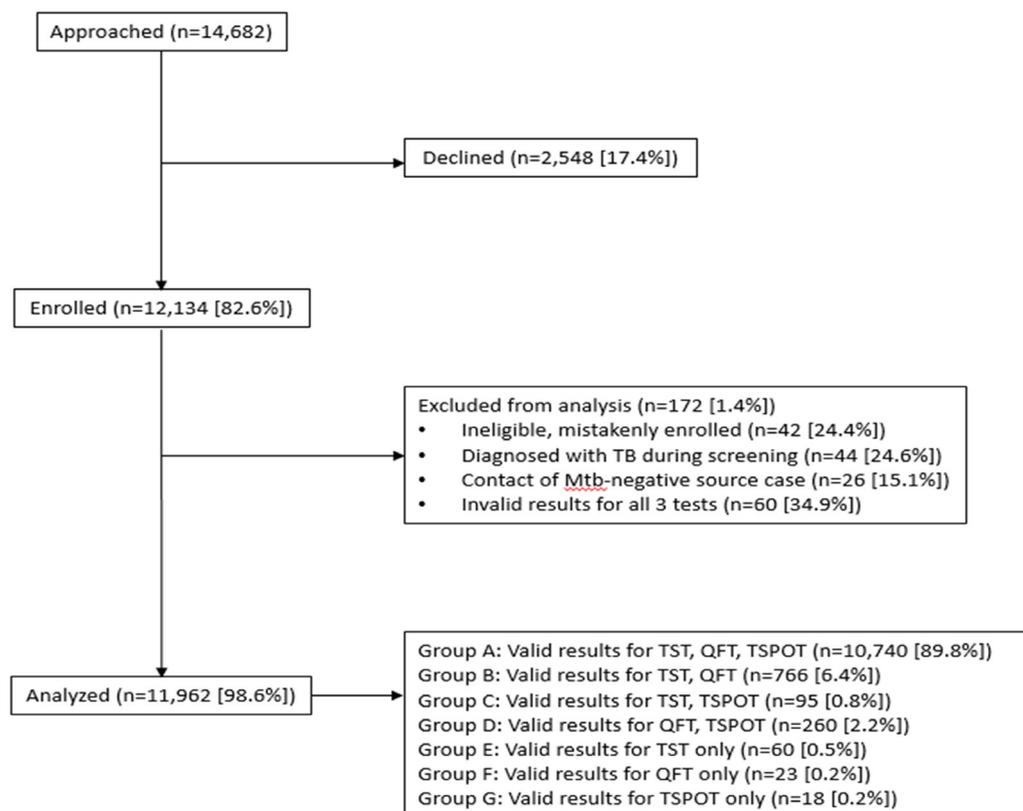
Figure 1 Flow chart of enrolled participants. QFT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT.TB; TST, tuberculin skin test.

with the df equal to the differences in df between models. We rejected the assumption of measurement invariance across all three groups (p<0.05 for all three comparisons) and divided the cohort into analysis groups on the basis of permutations of the three grouping variables (HIV status, foreign birth and aged ≥5 years). Because the numbers of participants in certain groups were limited and unlikely to yield robust results, we focused on five groups with ≥100 participants in each: (A) foreign-born, HIV-seronegative and aged ≥5 years; (B) foreign-born, HIV-seropositive and aged ≥5 years; (C) foreign-born, HIV-seronegative and aged <5 years; (d) US-born, HIV-seronegative and aged ≥5 years; and (e) US-born, HIV-seropositive and aged ≥5 years.

QFT and TSPOT were performed using standard clinical protocols at each site; as this was a pragmatic study, no special efforts were made to standardise the laboratory procedures at each site. We included data only from participants with valid results for all three tests. We used the international cut-off for the TSPOT (six or more spots as positive and five or fewer spots as negative),[15] because using the US interpretation with a borderline zone[16] would have required discarding all results in the borderline zone (five, six or seven spots); we also ran the models using the US cut-off and classifying borderline results as negative. We used current CDC guidelines for TST interpretation.[17] We used the standard manufacturer cut-off for the QFT-IT.[18] Because all of the tests are immunologically based with overlapping antigens, we suspected that the assumption of conditional independence of tests (after taking true latent class status into account) would not be valid. Therefore, we created two latent class models for each group, one using a modification of the method of Qu et al[19] that included a random effect to account for conditional dependence and the other setting the random effect to zero (ie, assuming conditional independence). We used a Bayesian approach for both of these models[20] with literature-based prior distributions for test sensitivities and broad prior distributions for specificity and prevalence (table 1). The literature-based prior distributions were used for sensitivity because the proportion of persons tested who subsequently develop active TB disease is a good proxy for test sensitivity, whereas broad prior distributions were used for other parameters because there is no similarly good proxy for test specificity or LTBI prevalence in the literature. The literature-based prior distributions for test sensitivity represent a weighted average of patients who were tested for LTBI and then subsequently experienced incident TB disease (online supplementary table S2 and S3). For example, the prior distribution for TST sensitivity for HIV-seronegative persons is based on 180 contacts, as reported in the literature, who had a TST, were prospectively followed and subsequently developed TB disease. A total of 143 of these persons had a positive TST; the point estimate for sensitivity would be 143/180, and the prior distribution representing those data would be a beta (143,37) distribution. Prior distributions for HIV-seronegative adults were obtained from a simple summation of studies of contacts of persons with TB disease; prior distributions for other groups (HIV-seropositive adults and children aged <5 years) were obtained from studies of those populations. To further assess the robustness of the latent class models, we split the TSPOT into two separate tests (the panel A and panel B results) with the same cut-off for a positive test applied to each (five or more spots greater than the nil plate). We then ran four test models (TST, QFT-IT, TSPOT panel A and TSPOT panel B) for each group. We used R V.3.3.0 (open-source software, the R Foundation, Vienna, Austria) and JAGS (Just Another Gibbs Sampler, available at http://mcmc-jags.sourceforge.net/) V.4.2.0

(open-source software) through the runjags package (V.3.2) to implement these models, with Markov chain Monte Carlo sampling to estimate parameter distributions. Four independent chains were used for each model. The initial 1000 and subsequent 5000 samples were used for model adaptation and burn-in, with subsequent sampling of a minimum of 20 000 iterations or enough iterations to obtain Gelman-Rubin statistics <1.05 for all sensitivity, specificity and prevalence parameters, whichever was greater.

Models that assumed the tests were conditionally independent (ie, without a random effect) provided higher sensitivity estimates for all three tests and lower prevalence estimates than the models that included a random effect to account for conditional dependence (figure 2 and online supplementary figure S1). Furthermore, the posterior means of the sensitivity estimates for the models that assumed conditional independence were often outside the 95% credible intervals of the literature-based prior estimates for sensitivity. Conversely, the sensitivity estimates for the models that accounted for conditional dependence generally fell within the credible intervals for the evidence-based prior distribution. We therefore focused on the conditional dependence (random effects) model for further reporting.

To assess the validity of the latent class measure, we created an exposure variable for the foreign-born participants as an estimation of the opportunities for a person to be exposed to and infected with TB. This surrogate exposure variable was calculated by multiplying the age at immigration by the WHO estimate of TB incidence in the country of origin in 2012.[21] We assigned each participant to a latent class (LTBI or no LTBI) with probability equal to the Bayesian posterior mean of the probability of belonging to that class derived from the model and examined the relationship between quartile of the exposure variable and estimated LTBI infection status (defined by latent class membership). The χ^2 test for trend was used to assess this relationship.

RESULTS
A total of 12 134 participants were enrolled in the study during 20 July 2012–28 September 2014. Of these, 10 740 had valid results for all three tests and were included in our study. Table 2 provides a summary of demographic characteristics for these participants. Table 3 includes a summary of test combinations for each of the five groups evaluated, and online supplementary table S4 shows the same summary considering US borderline results as negative. Certain test combinations were absent or rarely encountered in some groups (eg, negative TST with positive QFT-IT and TSPOT among foreign born, HIV-seronegative children aged <5 years).

Table 4 lists estimates for the prevalence of LTBI as well as test characteristics among the five groups derived from the three-test random effects model. The prevalence of LTBI in the groups ranged from 4.0% among foreign-born, HIV-seronegative children aged <5 years to 34.0% among foreign-born, HIV-seronegative persons aged ≥5 years. The sensitivity of the tests varied widely across the groups, usually with lower point estimates for sensitivity of all three tests among HIV-seropositive persons, compared with HIV-seronegative persons. Specificity of the TST was higher among US-born than among foreign-born groups, whereas specificities of the QFT-IT and TSPOT did not appear to vary by birthplace. The positive predictive value of the TST was only 10.0% among foreign-born, HIV-seronegative children aged <5 years and ranged from 40.4% among US-born, HIV-seropositive persons aged ≥5 years to 69.2% among foreign-born,

Table 1 Prior distributions used for Bayesian latent class analysis

| Study group	Parameter	Distribution	Reference
Foreign-born, HIV-seronegative, aged ≥5 years	LTBI prevalence	beta(1,1)	Assumption
	TST sensitivity	beta(143,37)	See online supplementary table S2
	TST specificity	beta(7,3)	Assumption
	QFT-IT sensitivity	beta(80,29)	See online supplementary table S2
	QFT-IT specificity	beta(9,1)	Assumption
	TSPOT sensitivity	beta(36,20)	See online supplementary table S2
	TSPOT specificity	beta(9,1)	Assumption
Foreign-born, HIV-seronegative, aged <5 years	LTBI prevalence	beta(1,1)	Assumption
	TST sensitivity	beta(90,41)	See online supplementary table S2
	TST specificity	beta(7,3)	Assumption
	QFT-IT sensitivity	beta(54,22)	See online supplementary table S2
	QFT-IT specificity	beta(9,1)	Assumption
	TSPOT sensitivity	beta(56,31)	See online supplementary table S2
	TSPOT specificity	beta(9,1)	Assumption
Foreign-born, HIV-seropositive, aged ≥5 years	LTBI prevalence	beta(1,1)	Assumption
	TST sensitivity	beta(40,56)	See online supplementary table S2
	TST specificity	beta(7,3)	Assumption
	QFT-IT sensitivity	beta(6,5)	See online supplementary table S2
	QFT-IT specificity	beta(9,1)	Assumption
	TSPOT sensitivity	beta(10,9)	See online supplementary table S2
	TSPOT specificity	beta(9,1)	Assumption
US-born, HIV-seronegative, aged ≥5 years	LTBI prevalence	beta(1,1)	Assumption
	TST sensitivity	beta(143,37)	See online supplementary table S2
	TST specificity	beta(7,3)	Assumption
	QFT-IT sensitivity	beta(80,29)	See online supplementary table S2
	QFT-IT specificity	beta(9,1)	Assumption
	TSPOT sensitivity	beta(36,20)	See online supplementary table S2
	TSPOT specificity	beta(9,1)	Assumption
US-born, HIV-seropositive, aged ≥5 years	LTBI prevalence	beta(1,1)	Assumption
	TST sensitivity	beta(40,56)	See online supplementary table S2
	TST specificity	beta(7,3)	Assumption
	QFT-IT sensitivity	beta(6,5)	See online supplementary table S2
	QFT-IT specificity	beta(9,1)	Assumption
	TSPOT sensitivity	beta(10,9)	See online supplementary table S2
	TSPOT specificity	beta(9,1)	Assumption

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

HIV-seropositive persons aged ≥5 years. The positive predictive value of the QFT-IT was low (41.4%) among US-born, HIV-seropositive persons aged ≥5 years and ranged from 73.1% to 96.4% among other groups. The positive predictive value of the TSPOT ranged from 77.5% to 98.2% across groups. Negative predictive values of all tests ranged from 79.1% to 98.8% across groups. Using the US cut-off for TSPOT instead of the international cut-off had minimal impact on the prevalence, sensitivity and specificity estimates for most groups with the exception of prevalence in foreign-born, HIV-seropositive persons aged ≥5 years, but the credible interval for this group was very wide due to a small number of observations (online supplementary table S5).

Table 5 displays the association between our surrogate exposure variable (age at immigration multiplied by TB incidence in the country of origin) and probability of a positive test or latent class assignment among foreign-born, HIV-seronegative persons aged ≥5 years with valid data for the surrogate exposure variable (n=7880). The probability of a positive result increased substantially by increasing exposure quartile for all three tests as well as for the latent class assignment. The LTBI prevalence by LCA for each quartile of exposure was between the prevalence by interferon-γ release assay (IGRA) and TST.

Table 6 describes the positive predictive values of the various test combinations derived from the three-test, random effects model. Although credible intervals were wide for certain relatively uncommon test combinations, common themes emerged across groups. First, as expected, the positive predictive value of three positive tests was high (approximately 100%). Second, the positive predictive value of three negative tests was low but varied
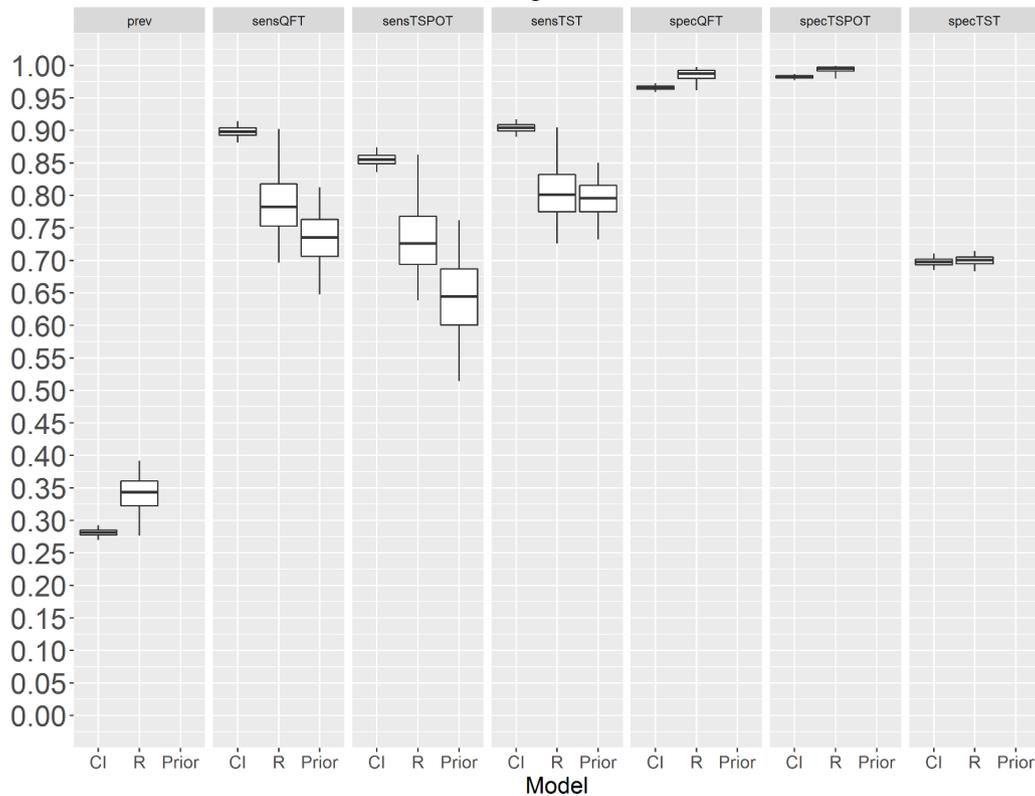
Figure 2 Latent TB infection prevalence and test characteristics (sensitivity and specificity) for foreign-born, HIV-seronegative participants aged ≥5 years. On the x-axis, the label 'CI' stands for the model that assumes conditional independence of the tests; 'R' stands for the model that includes a random effect (to model conditional dependence between the tests); and 'Prior' indicates the scientific literature-based prior distribution (for test sensitivity only). The horizontal lines in the centre of the rectangles are the Bayesian posterior means, and the top and bottom lines in each box are the boundaries of the first and third quartiles, respectively. The tips of the vertical lines span the 95% credible interval for each parameter. For purposes of this analysis, TSPOT was evaluated by using the international criteria for a positive test (≥6 spots is positive; ≤5 spots is negative). CI, conditional independence model; Prev, latent TB prevalence; Prior, scientific literature-based prior distribution (for test sensitivity only); R, random effect model; sensQFT, sensitivity of QuantiFERON Gold In-Tube; sensTSPOT, sensitivity of T-SPOT.TB; sensTST, sensitivity of the tuberculin skin test; specQFT, specificity of QuantiFERON Gold In-Tube; specTSPOT, specificity of T-SPOT.TB; specTST, specificity of the tuberculin skin test.

by underlying prevalence of LTBI within the group. Foreign-born persons aged ≥5 years, who had an underlying LTBI prevalence of 34.0% and 31.7% in HIV-seronegative and HIV-seropositive

groups, respectively, both had non-zero proportions of participants (4.8% and 12.7%) with LTBI, despite having three negative tests. Conversely, having an isolated positive TST (with the

Table 2 Study participant demographic characteristics

Characteristic	Foreign-born, HIV-seronegative, aged ≥5 years (n=7931)	Foreign-born, HIV-seronegative, aged <5 years (n=464)	Foreign-born, HIV-seropositive, aged ≥5 years (n=140)	US-born, HIV-seronegative, aged ≥5 years (n=775)	US-born, HIV-seropositive, aged ≥5 years (n=1226)
Age, years (median, IQR)	28.9 (18.9–40.6)	3.1 (2.2–4.0)	46.4 (37.6–56.0)	36.6 (22.3–51.0)	49.6 (42.6–54.8)
TB incidence/100 000 population in country of birth	n (%)	n (%)	n (%)	n (%)	n (%)
0 – 9.9	30 (0.4)	5 (1.1)	23 (16.4%)	775 (100%)	1226 (100%)
10 – 19	69 (0.9)	19 (4.1)	5 (3.6%)	0	0
20 – 49	1445 (18.2)	60 (12.9)	39 (27.9%)	0	0
50 – 124	1072 (13.5)	213 (45.9)	16 (11.4%)	0	0
125 – 299	3374 (42.5)	139 (30.0)	36 (25.7%)	0	0
300 – 499	1893 (23.9)	15 (3.2)	20 (14.3%)	0	0
≥ 500	48 (0.6)	13 (2.8)	1 (0.7%)	0	0
CD4+ T-lymphocyte count (median, IQR)	NA	NA	555 (255–713)*	NA	544 (355–761)†

*76 foreign-born, HIV-seropositive participants aged ≥5 years had available self-reported CD4+ lymphocyte count data.
†789 US-born, HIV-seropositive participants aged ≥5 years had available self-reported CD4+ lymphocyte count data.
NA, not applicable.

Table 3 Combinations of test results for participants, by study group (by using the international cut-off of ≥6 spots=positive for TST=T-SPOT.TB

Test combination, TST/ QFT–IT/TSPOT	Foreign-born, HIV-seronegative, aged ≥5 Years	Foreign-born, HIV-seronegative, aged <5 years	Foreign-born, HIV-seropositive, aged ≥5 years	US-born, HIV-seronegative, aged ≥5 years	US-born, HIV-seropositive, aged ≥5 years
− − −	3778	330	94	589	1090
− − +	86	1	1	7	9
− + −	162	2	8	20	53
− + +	157	0	0	12	10
+ − −	1666	115	10	53	40
+ − +	200	1	0	5	1
+ + −	314	6	7	14	8
+ + +	1568	9	20	75	15
Total	7931	464	140	775	1226

−, negative test; +, positive test; FB, foreign born; QFT-IT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT.TB; TST, tuberculin skin test.

other two tests negative) had similar positive predictive value to having all three tests negative across all groups.

DISCUSSION

LCA of this diverse cohort enrolled in a low-incidence setting provides insights into LTBI screening test performance. Consistent with other observations,[22–24] we determined that test performance varied substantially, depending on the group tested. Specifically, we noted lower test sensitivity among HIV-seropositive persons than among HIV-seronegative persons and lower specificity for the TST among foreign-born persons (presumably because of a combination of BCG vaccination and exposure to non-tuberculous mycobacteria) than among US-born persons. More importantly, combining these test characteristics with estimated underlying LTBI prevalence enables calculation of the positive predictive value of tests and test combinations when

there is no gold standard, which permits optimum use of these tests. For example, a foreign-born, HIV-seronegative adult in our cohort with a positive TST would have a calculated Bayesian probability of 57.9% of having LTBI, which is only slightly better than a coin toss. If that same person had a positive QFT-IT or TSPOT, the Bayesian probability of LTBI would be 96.4% or 98.2%, respectively. A positive TST with a negative QFT-IT or TSPOT reduces the probability of LTBI to 21%, whereas a positive TST followed by a positive QFT-IT or TSPOT is associated with a 99% probability of LTBI in a foreign-born, HIV-seronegative adult in our cohort (ignoring any potential booster effect of the TST on the QFT-IT or TSPOT). These findings support both CDC recommendations of using an IGRA for foreign-born persons who might have received the BCG vaccine and European guidelines recommending confirmation of a positive TST with an IGRA.[25 26]

Table 4 Estimated prevalence and test characteristics derived from the three-test, random effects model, by using the international TSPOT cut-off among specified groups

Parameter	Foreign-born, HIV-seronegative, aged ≥5 years, % (95% credible interval)	Foreign-born, HIV-seronegative, aged <5 years, % (95% credible interval)	Foreign-born, HIV-seropositive, aged ≥5 years, % (95% credible interval)	US-born, HIV-seronegative, aged ≥5 years, % (95% credible interval)	US-born, HIV-seropositive, aged ≥5 years, % (95% credible interval)
LTBI prevalence	34.0 (27.6 to 39.2)	4.0 (1.9 to 6.7)	31.7 (1.9 to 46.8)	17.5% (13.1 to 22.5)	4.2 (2.6 to 6.3)
TST					
Sensitivity	80.7 (72.6 to 90.5)	69.1 (58.5 to 79.7)	57.0 (49.6 to 65.3)	72.6 (62.2 to 84.3)	54.0 (44.2 to 64.3)
Specificity	70.0 (68.3 to 71.5)	73.9 (69.6 to 77.9)	88.5 (70.7 to 96.0)	92.1 (89.7 to 94.3)	96.5% (95.3 to 97.6)
PPV	57.9 (52.8 to 61.3)	10.0 (4.8 to 16.5)	69.2 (3.7 to 90.3)	66.0 (56.5 to 75.4)	40.4% (27.0 to 55.1)
NPV	87.3 (79.9 to 95.0)	98.3 (96.7 to 99.3)	81.3 (69.1 to 98.9)	93.9 (89.9 to 97.3)	97.9 (96.6 to 98.9)
QFT-IT					
Sensitivity	78.9 (69.6 to 90.2)	71.2 (55.3 to 86.5)	59.1 (47.1 to 72.0)	78.0 (65.0 to 91.0)	67.5 (52.9 to 81.7)
Specificity	98.5 (96.1 to 99.8)	98.9 (97.4 to 99.9)	93.1 (73.9 to 99.1)	97.9 (96.0 to 99.4)	95.8 (94.4 to 97.1)
PPV	96.4 (90.0 to 99.5)	73.1 (41.3 to 95.3)	79.0 (3.9 to 97.7)	88.6 (78.6 to 97.0)	41.4 (27.3 to 56.8)
NPV	89.9 (83.6 to 96.3)	98.8 (97.4 to 99.6)	82.8 (70.1 to 98.9)	95.3 (91.1 to 98.5)	98.5 (97.2 to 99.4)
TSPOT					
Sensitivity	73.5 (63.9 to 86.3)	58.9 (42.7 to 76.2)	45.8 (35.2 to 58.6)	68.5 (55.7 to 83.4)	55.0 (40.7 to 70.6)
Specificity	99.3 (98.0 to 99.9)	99.4 (98.4 to 99.9)	97.2 (84.1 to 99.7)	99.1 (98.1 to 99.8)	99.3 (98.6 to 99.8)
PPV	98.2 (94.2 to 99.8)	79.2 (52.0 to 96.3)	86.9 (5.8 to 99.0)	94.3 (87.2 to 98.9)	77.5 (58.0 to 93.5)
NPV	87.7 (81.1 to 94.9)	98.3 (96.5 to 99.4)	79.1 (65.3 to 98.9)	93.5 (89.0 to 97.3)	98.0 (96.5 to 99.1)

Parameter values are mean estimates with 95% credible intervals in parentheses.
LTBI, latent TB infection; NPV, negative predictive value; PPV, positive predictive value; QFT-IT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT.TB; TST, tuberculin skin test.

Table 5 Association between the exposure variable (age at immigration multiplied by estimated active TB incidence in country of birth) and either test results or latent class status among foreign-born, HIV-seronegative participants aged ≥5 years

Exposure quartile	TST-positive (RR)	QFT-IT-positive (RR)	TSPOT-positive (RR)	Latent class (RR)
1	34.3% (1.0)	14.9% (1.0)	13.2% (1.0)	19.8% (1.0)
2	41.9% (1.22)	22.0% (1.48)	20.8% (1.58)	27.0% (1.36)
3	53.3% (1.55)	34.0% (2.28)	32.5% (2.46)	39.5% (1.99)
4	59.6% (1.74)	40.2% (2.70)	41.4% (3.14)	46.9% (2.37)
P value for trend	$<2\times10^{-16}$	$<2\times10^{-16}$	$<2\times10^{-16}$	$<2\times10^{-16}$

QFT-IT, QuantiFERON Gold In-Tube; RR, relative risk; TSPOT, T-SPOT.TB, classified as positive/negative by using the international standard (≥6 spots=positive; TST, tuberculin skin test; ≤5 spots=negative).

Our data demonstrate certain interesting findings in some of the groups. First, because the estimated prevalence of LTBI was low among the US-born, HIV-seropositive study participants, the positive predictive value of the QFT-IT was low among this group (41.4% overall) despite high test specificity. This finding is consistent with other authors' descriptions of what appear to be relatively frequent false-positive QFT-IT tests among US-born, HIV-seropositive persons.[27] The TSPOT had a higher point estimate of positive predictive value among this population (77.5%), but this still implies that approximately one-fourth of positive TSPOT results among this group are false positives. Conversely, the negative predictive values of all the tests among this group were high (97%–98%); therefore, few HIV-seropositive persons with LTBI would be missed with any of the tests. These findings should be considered in the context of our study population, who had high CD4$^+$ T-lymphocyte counts. Second, the vast majority of positive TSTs among foreign-born children aged <5 years were isolated positive tests, with concurrent negative QFT-IT and TSPOT. Our analysis demonstrates that a limited proportion of children with an isolated positive TST (1.2%) truly have LTBI, indicating that these isolated positive TSTs are almost all false positives, presumably because of recent BCG vaccination. Although concerns have been raised that IGRAs are less sensitive than TSTs among young children,[28 29] our results demonstrate that the sensitivity of all three tests are suboptimal to approximately the same degree; however, the greater specificity of the IGRAs would avoid unnecessary LTBI treatment for many young

foreign-born children. Furthermore, the negative predictive value of all the tests was good (>98%); therefore, few young children with LTBI would be expected to be missed because of false-negative tests. On the basis of these results, using an IGRA for LTBI screening among foreign-born children aged <5 years would be strongly preferred over the TST.

Multiple indirect lines of evidence support our findings. First, the strong association between our surrogate exposure variable (age at immigration multiplied by TB incidence in the country of birth) and LTBI as assigned by the latent class model supports the concept that the model is measuring LTBI. Second, the estimated LTBI prevalence among foreign-born children aged <5 years is in line with the expected prevalence among this group, given other studies that have estimated an annual risk for infection (with serial testing) of 0.2%–6% in TB-endemic countries,[30 31] with the majority of estimates for high-incidence countries being 1%–3%.[32] In this context, the latent class LTBI prevalence (4.2%) is a more plausible estimate than the 28% prevalence (131/464) obtained among our participants if the TST were used to ascertain LTBI. The estimated LTBI prevalence among certain groups was significantly higher than what has been reported among other population-based studies (eg, 34.0% among our foreign-born, HIV-negative participants aged ≥5 years, compared with approximately 9% in the National Health and Nutrition Examination Survey data).[33] However, this difference is to be expected given that (A) our study was not population based but focused on recruiting persons believed to be at high risk for LTBI and (B) the methods used to estimate LTBI prevalence in our study take into account the imperfect sensitivity of currently used tests, which has not yet been done with the National Health and Nutrition Examination Survey data.

This analysis has potential limitations. The study was conducted as a pragmatic trial, so no special efforts were made to synchronise laboratory procedures across sites; this may have increased variability in test performance across sites and contributed to uncertainty in estimates of test characteristics. However, this reflects the 'real-world' performance of the tests, which makes our results more generalisable. LCA by definition examines phenomena (in this case LTBI) that cannot be directly observed. Although strong conceptual and biological plausibility exists for believing that the latent class represents a dichotomy between TB infection and lack of infection, the latent class might represent a subtly different biological phenomenon that is not concordant with the traditional dichotomy of LTBI. In addition,

Table 6 Positive predictive values of test combinations derived from the three-test, random effects model in specified subgroups (TSPOT positive at the international cut-off of ≥6 spots)

Test combination (TST/QFT-IT/TSPOT)	Foreign-born, HIV-seronegative, aged ≥5 years, % (95 % credible intervals)	Foreign-born, HIV-seronegative, aged <5 years, % (95 % credible intervals)	Foreign-born, HIV-seropositive, aged ≥5 years, % (95 % credible intervals)	US-born, HIV-seronegative, aged ≥5 years, % (95 % credible intervals)	US-born, HIV-seropositive, aged ≥5 years, % (95 % credible intervals)
– – –	4.8 (0.1 to 11.8)	0.5 (0 to 2.1)	12.7 (0 to 26.6)	2.1 (0 to 6.3)	0.6 (0 to 1.7)
– – +	58.3 (15.5 to 75.0)	NA	NA	50.2 (0 to 100)	34.7 (0 to 77.8)
– + –	69.6 (16.2 to 96.8)	NA	50.0 (0 to 100)	44.3 (10.0 to 85.0)	12.9 (1.9 to 28.3)
– + +	98.2 (93.3 to 100)	NA	NA	98.7 (91.7 to 100)	93.7 (70.0 to 100)
+ – –	8.2 (1.7 to 13.0)	1.2 (0 to 4.3)	30.1 (0 to 70.0)	11.4 (1.9 to 26.4)	8.4 (0 to 22.5)
+ – +	82.2 (68.0 to 88.3)	NA	NA	92.8 (60.0 to 100)	NA
+ + –	91.1 (73.9 to 99.2)	66.0 (16.7 to 100)	82.2 (0 to 100)	92.0 (71.4 to 100)	75.3 (37.5 to 100)
+ + +	99.6 (98.8 to 100)	99.2 (88.9 to 100)	94.0 (5.0 to 100)	100 (98.7 to 100)	99.7 (93.3 to 100)

The positive predictive values represent the estimated likelihood that an individual with the given test combination has latent TB infection; 95% credible intervals are listed in parentheses after each point estimate; cells in which <5 participants in the group had the listed test combination are listed as NA.
NA, not applicable; QFT-IT, QuantiFERON Gold In-Tube; TSPOT, T-SPOT.TB; TST, tuberculin skin test.

the choice of prior probabilities has significant influence on the model results; in the current manuscript, the sensitivity estimates were chosen based on the best available gold standard in the literature (progression to active TB after a test was performed), but use of different prior probabilities may well have produced different results. Furthermore, whereas adjusting for conditional dependence clearly had important effects on the results, several methods for performing such an adjustment exist, and each method might provide slightly different results, as discussed by Albert and Dodd.[34] However, alternative proposed methods (eg, latent mixture models) have been problematic for some authors, involving a failure to converge.[35] We attempted a latent mixture approach with our data and encountered problems with convergence as well; therefore, we discarded that approach. Finally, limited sample sizes for certain groups (eg, HIV-positive, foreign-born participants aged ≥5 years) and low frequency of observing certain test patterns (eg, negative TST, positive QFT-IT, positive TSPOT among HIV-negative, foreign-born children aged <5 years) limited our ability to precisely estimate selected parameters for these groups; additional data are needed for improving parameter precision for these groups.

In conclusion, Bayesian LCA of a large, prospectively enrolled, US-based cohort highlighted the limitations of LTBI diagnostic tests. Our findings demonstrate that IGRAs should be strongly preferred over the TST for both foreign-born children and adults. Although we hope that these findings can be used to optimise existing tests, they just as clearly demonstrate that improved diagnostic tests for LTBI are urgently needed.

REFERENCES

1 Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Ann Intern Med* 2008;149:177–84.
2 Rutherford ME, Nataprawira M, Yulita I, *et al*. QuantiFERON®-TB Gold In-Tube assay vs. tuberculin skin test in Indonesian children living with a tuberculosis case. *Int J Tuberc Lung Dis* 2012;16:496–502.
3 Diel R, Loddenkemper R, Meywald-Walter K, *et al*. Comparative performance of tuberculin skin test, QuantiFERON-TB-Gold In Tube assay, and T-Spot.TB test in contact investigations for tuberculosis. *Chest* 2009;135:1010–8.
4 Kassa D, Ran L, Jager W, *et al*. Discriminative expression of whole blood genes in HIV patients with latent and active TB in Ethiopia. *Tuberculosis* 2016;100:25–31.
5 Zak DE, Penn-Nicholson A, Scriba TJ, *et al*. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016;387:2312–22.
6 Geluk A, van Meijgaarden KE, Franken KL, *et al*. Immunological crossreactivity of the Mycobacterium leprae CFP-10 with its homologue in Mycobacterium tuberculosis. *Scand J Immunol* 2004;59:66–70.
7 Kuznetcova TI, Sauty A, Herbort CP. Uveitis with occult choroiditis due to Mycobacterium kansasii: limitations of interferon-gamma release assay (IGRA) tests (case report and mini-review on ocular non-tuberculous mycobacteria and IGRA cross-reactivity). *Int Ophthalmol* 2012;32:499–506.
8 Zellweger JP, Sotgiu G, Block M, *et al*. Risk Assessment of Tuberculosis in Contacts by IFN-γ Release Assays. A Tuberculosis Network European Trials Group Study. *Am J Respir Crit Care Med* 2015;191:1176–84.
9 Kik SV, Franken WP, Mensen M, *et al*. Predictive value for progression to tuberculosis by IGRA and TST in immigrant contacts. *Eur Respir J* 2010;35:1346–53.
10 Rangaka MX, Wilkinson KA, Glynn JR, *et al*. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis* 2012;12:45–55.
11 Sobotzki C, Riffelmann M, Kennerknecht N, *et al*. Latent class analysis of diagnostic tests for adenovirus, Bordetella pertussis and influenza virus infections in German adults with longer lasting coughs. *Epidemiol Infect* 2016;144:840–6.
12 Lanza ST, Collins LM, Lemmon DR, *et al*. PROC LCA: A SAS Procedure for Latent Class Analysis. *Struct Equ Modeling* 2007;14:671–94.
13 Aabye MG, Ravn P, PrayGod G, *et al*. The impact of HIV infection and CD4 cell count on the performance of an interferon gamma release assay in patients with pulmonary tuberculosis. *PLoS One* 2009;4:e4220.
14 Seddon JA, Paton J, Nademi Z, *et al*. The impact of BCG vaccination on tuberculin skin test responses in children is age dependent: evidence to be considered when screening children for tuberculosis infection. *Thorax* 2016;71:932–9.
15 O I. T-SPOT.TB Package Insert UK version 3 Abingdon, Oxfordshire, UK2016. 2017 www.oxforddiagnosticlaboratories.eu/wp-content/media/PI-TB-IVD-UK-v3.pdf (accessed 19 Oct 2017).
16 Immunotec O. T-SPOT.TB Package Insert US version 5 Marlborough, MA, USA2015. 2017 www.tspot.com/wp-content/uploads/2012/10/PI-TB-US-v5.pdf (accessed 19 Oct 2017).
17 American Thoracic Society. Targeted tuberculin testing and treatment of latent tuberculosis infection. *MMWR Recomm Rep* 2000;49(RR-6):1–51.
18 Qiagen. QuantiFERON-TB Gold (QFT) ELISA Package Insert Germantown, MD, USA 2016. 2017 www.QuantiFERON.com (accessed 19 Oct 2017).
19 Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996;52:797–810.
20 Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001;57:158–67.
21 Organization WH. *Global Tuberculosis Report 2013*. Geneva: World Health Organization, 2013:289.
22 Mandalakas AM, van Wyk S, Kirchner HL, *et al*. Detecting tuberculosis infection in HIV-infected children: a study of diagnostic accuracy, confounding and interaction. *Pediatr Infect Dis J* 2013;32:e111–8.
23 Weinfurter P, Blumberg HM, Goldbaum G, *et al*. Predictors of discordant tuberculin skin test and QuantiFERON®-TB Gold In-Tube results in various high-risk groups. *Int J Tuberc Lung Dis* 2011;15:1056–61.
24 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.

25 Mazurek GH, Jereb J, Vernon A, *et al*. Updated guidelines for using Interferon Gamma Release Assays to detect Mycobacterium tuberculosis infection - United States, 2010. *MMWR Recomm Rep* 2010;59(RR-5):1–25.

26 Weston R, Stagg HR FD. *Health Protection Agency position statement on the use of Interferon Gamma Release Assay (IGRA) tests for Tuberculosis (TB)*: HPA Tuberculosis Programme Board, 2012.

27 Gray J, Reves R, Johnson S, *et al*. Identification of false-positive QuantiFERON-TB Gold In-Tube assays by repeat testing in HIV-infected patients at low risk for tuberculosis. *Clin Infect Dis* 2012;54:e20–3.

28 Chiappini E, Bonsignori F, Mazzantini R, *et al*. Interferon-gamma release assay sensitivity in children younger than 5 years is insufficient to replace the use of tuberculin skin test in western countries. *Pediatr Infect Dis J* 2014;33:1291–3.

29 Bamford AR, Crook AM, Clark JE, *et al*. Comparison of interferon-gamma release assays and tuberculin skin test in predicting active tuberculosis (TB) in children in the UK: a paediatric TB network study. *Arch Dis Child* 2010;95:180–6.

30 Yuen CM, Krapivina TM, Kazennyy BY, *et al*. Annual risk of tuberculous infection measured using serial skin testing, Orel Oblast, Russia, 1991-2005. *Int J Tuberc Lung Dis* 2015;19:39–43.

31 Shanaube K, Sismanidis C, Ayles H, *et al*. Annual risk of tuberculous infection using different methods in communities with a high prevalence of TB and HIV in Zambia and South Africa. *PLoS One* 2009;4:e7749.

32 Cauthen GM, Pio A, ten Dam HG. Annual risk of tuberculous infection. 1988. *Bull World Health Organ* 2002;80:503–11. discussion 01-2.

33 Mancuso JD, Diffenderfer JM, Ghassemieh BJ, *et al*. The Prevalence of Latent Tuberculosis Infection in the United States. *Am J Respir Crit Care Med* 2016;194:501–9.

34 Albert PS, Dodd LE. A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics* 2004;60:427–35.

35 van Smeden S. *Diagnostic research in the absence of a gold standard*. Utrecht: University Medical Center, 2016.