Published in final edited form as:

*Ann Am Thorac Soc*. 2023 September ; 20(9): 1258–1266. doi:10.1513/AnnalsATS.202212-1005OC.

# Examining Test Cutoffs to Optimize Diagnosis of Latent Tuberculosis Infection in People Born Outside the United States

**Sofia Zavala**[1], **Kathryn Winglee**[2], **Christine S. Ho**[2], **April C. Pettit**[3], **Amina Ahmed**[4], **Dolly J. Katz**[2], **Robert W. Belknap**[5], **Jason E. Stout**[1] **Tuberculosis Epidemiologic Studies Consortium**

[1]Division of Infectious Diseases and International Health, Department of Medicine, Duke University Medical Center, Durham, North Carolina

[2]Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia

[3]Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee

[4]Department of Pediatrics, Atrium Health, Charlotte, North Carolina

[5]Denver Health and Hospital Authority, Denver, Colorado

## Abstract

**Rationale:** Detection of latent tuberculosis infection (LTBI) in persons born in high tuberculosis (TB) incidence countries living in low TB incidence countries is key to TB elimination in low-incidence countries. Optimizing LTBI tests is critical to targeting treatment.

**Objectives:** To compare the sensitivity and specificity of tuberculin skin test (TST) and two interferon-γ release assays at different cutoffs and of a single test versus dual testing.

**Methods:** We examined a subset ($N$ = 14,167) of a prospective cohort of people in the United States tested for LTBI. We included non–U.S.-born, human immunodeficiency virus-seronegative people ages 5 years and older with valid TST, QuantiFERON-TB Gold-in-Tube (QFT), and T-SPOT.*TB* (TSPOT) results. The sensitivity/specificity of different test cutoffs and test combinations, obtained from a Bayesian latent class model, were used to construct receiver operating characteristic (ROC) curves and assess the area under the curve (AUC) for each test. The sensitivity/specificity of dual testing was calculated.

Correspondence and requests for reprints should be addressed to Jason E. Stout, M.D., M.H.S., Division of Infectious Diseases and International Health, Department of Medicine, Duke University Medical Center, Box 102359-DUMC, Durham, NC 27710. jason.stout@dm.duke.edu.

*Author Contributions*: K.W. and J.E.S. conceived and designed the analysis. J.E.S. performed the analysis. S.Z. and J.E.S. wrote the initial version of the manuscript. S.Z., K.W., C.S.H., A.C.P., A.A., D.J.K., R.W.B., and J.E.S. revised the manuscript for critically important intellectual content and approved the manuscript. The findings and conclusions are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention (CDC). References in this manuscript to any specific commercial products, process, service, manufacturer, or company do not constitute its endorsement or recommendation by the U.S. government or CDC.

**Author disclosures** are available with the text of this article at www.atsjournals.org.

**Results:** The AUC of the TST ROC curve was 0.81 (95% credible interval (CrI), 0.78–0.86), with sensitivity/specificity at cutoffs of 5, 10, and 15 mm of 86.5%/61.6%, 81.7%/71.3%, and 55.6%/88.0%, respectively. The AUC of the QFT ROC curve was 0.89 (95% CrI, 0.86–0.93), with sensitivity/specificity at cutoffs of 0.35, 0.7, and 1.0 IU/mL of 77.7%/98.3%, 66.9%/99.1%, and 61.5%/99.4%. The AUC of the TSPOT ROC curve was 0.92 (95% CrI, 0.88–0.96) with sensitivity/specificity for five, six, seven, and eight spots of 79.2%/96.7%, 76.8%/97.7%, 74.0%/98.6%, and 71.8%/99.5%. Sensitivity/specificity of TST-QFT, TST-TSPOT, and QFT-TSPOT at standard cutoffs were 73.1%/99.4%, 64.8%/99.8%, and 65.3%/100%.

**Conclusion:** Interferon-γ release assays have a better predictive ability than TST in people at high risk of LTBI.

## Keywords

latent tuberculosis; tuberculin skin test; interferon γ release assay; test cutoffs; latent class analysis

In countries with a low burden of tuberculosis (TB), reactivated latent TB infection (LTBI) accounts for the majority of cases of TB disease (1, 2). In the United States, an estimated 13 million people have LTBI; on the basis of estimates from interferon-γ release assays (IGRA), prevalence is 15.9% in the non–U.S.-born population and 2.8% in U.S.-born people (3, 4). TB elimination in low-incidence countries relies mainly on accurate identification and treatment of people with LTBI to prevent progression to TB disease (5).

Two types of tests are approved by the FDA (U.S. Food and Drug Administration) for the diagnosis of LTBI: The tuberculin skin test (TST) and two IGRAs: The QuantiFERON Gold in-tube (QFT) and the T-SPOT.*TB* (TSPOT). Both types are indirect tests on the basis of immunologic response to infection with *Mycobacterium tuberculosis*; neither is a reference gold standard.

TST has poor specificity in non–U.S.-born persons because of crossreactivity of its antigens with the *Bacillus* Calmette-Guerin (BCG), administered to 80% of newborns worldwide (6). IGRAs have greater specificity because they are on the basis of the stimulation of T cells by two *Mycobacterium tuberculosis*-specific antigens not found in BCG (7–10). Two strategies have been used to improve the performance of these tests. One strategy, which has been a standard practice for the interpretation of TST but not IGRAs, uses different test cutoffs on the basis of epidemiologic (e.g., country of birth and recent contact with a TB case) and clinical (e.g., immunosuppression) risk factors (11, 12). With the more recent availability of IGRAs, another strategy uses sequential testing with a second test to confirm the initial test result. This strategy most frequently uses an IGRA as the second test after an initial positive TST, reduces false-positive results, and is cost-effective (13–17). The lack of a gold standard for LTBI poses a significant challenge to comparing a risk-stratified test cutoff strategy (applied to IGRAs) and a sequential testing strategy for LTBI screening.

One technique to address this challenge is latent class analysis (LCA), a statistical method that can estimate test characteristics in the absence of a gold standard. LCA has been used to evaluate the diagnostic performance of TST and IGRAs at standard cutoffs (18–20). The aim of this study is to extend previous LCA models to better understand the tradeoffs in

sensitivity and specificity of different test cutoffs and to understand how different test cutoffs compare with dual testing for LTBI diagnosis in a large cohort of non–U.S.-born people residing in the United States.

## Methods

### Study Population and Procedures

The main study has been previously described (21). Briefly, 18 participating clinics in 11 states prospectively enrolled children and adults at relatively high risk of LTBI for the United States (we will refer to this group as “high risk” going forward) or progression to TB disease (close contacts of people with infectious TB disease; non–US-born people whose populations in the United States had TB rates of 10 cases per 100,000 people or higher; people who spent at least 30 days in the previous 5 years in countries with high (>100 cases per 100,000) TB rates; people living with human immunodeficiency virus [HIV]; or members of groups with high [≥25%] local prevalence of LTBI [e.g., people experiencing homelessness]). Participants were interviewed about demographics and past LTBI/TB history, administered all three available LTBI tests, and followed for 2 years to identify incident TB cases.

The study was approved by the CDC (Centers for Disease Control and Prevention) Institutional Review Board (IRB) and local IRBs that did not rely on the CDC IRB. All adult participants provided written informed consent; those under 18 years old had parental permission and provided assent. The study was registered at clinicaltrials.gov (identifier NCT01622140).

### Exclusions

In this analysis, we excluded U.S.-born individuals and people with missing country of birth. We also excluded non–U.S.-born individuals who were also close contacts, HIV-positive, or less than 5 years old because the diagnostic characteristics of LTBI tests are likely to systematically differ in those groups compared with the rest of the study cohort (18). Individuals with missing LTBI test results, indeterminate or invalid QFT results, or invalid TSPOT results were excluded. Participants found to have TB disease at enrollment were excluded as the focus of this study was on LTBI diagnosis.

### Selection of Test Cutoffs

Initial thresholds for test positivity were on the basis of CDC guidelines and FDA-approved labeling: TST ≥ 10 mm of induration, QFT ≥ 0.35 international units/milliliter (IU/mL), and TSPOT ≥ 8 spots (11, 22–24). As the protocol predated FDA recommendations for retesting, borderline TSPOT results (5–7 spots) were not repeated per study protocol and were categorized as negative (24).

Although all possible cutoffs were included in the receiver operator characteristic (ROC) curves, we reported specific sensitivity/specificity values for cutoffs that are either commonly used in U.S. guidelines (e.g., 5, 10, and 15 mm for TST) or have been reported in prior publications (22, 25). For QFT, we reported sensitivity/specificity for 0.35, 0.7, and

1.0 IU/mL, as these were used in a prior analysis (26). TSPOT cutoffs are designated by two separate guidelines: In the United States, FDA-approved cutoffs are fewer than 5 spots for a negative result and at least 8 spots for a positive, with 5–7 spots designated as borderline, with retest recommended; international guidelines use fewer than 6 spots as negative, and at least 6 spots as positive (11, 27). We, therefore, reported sensitivity/specificity of using 5, 6, 7, and 8 spots as cutoffs (28). All values below a cutoff were interpreted as negative, and values at or above a cutoff were interpreted as positive for this analysis.

## Statistical Analysis

LCA was used to estimate the prevalence of LTBI and the test diagnostic characteristics in our study population. Because all three tests are immunologically on the basis of overlapping antigens, the assumption of conditional independence of tests would not be valid. Therefore, we used the Dendukuri modification of the Qu and colleagues method (29, 30) that included a random effect to account for the conditional dependence of the tests. We used a Bayesian approach for the model, with literature-based prior distributions for test sensitivities and broad prior distributions for specificity and prevalence. Additional details on this approach were published previously (18), and data on model fit are provided in the data supplement (*see* Section EI in the data supplement).

Analysis was performed in R version 4.0.5 (31). We estimated the sensitivity and specificity of all three tests at the varying test cutoffs noted above. ROC curves were generated indirectly from the Bayesian LCA model using the positive predictive value (PPV; also known as posterior probability) of each of the eight possible test result combinations (positive or negative TST, QFT, and TSPOT). The prevalence of LTBI on the basis of each of the different test combinations was calculated by multiplying the number of participants with the combination by the corresponding PPV obtained from our LCA model (*see* example calculation in Section EII). The assumption was made that the probability of an individual having LTBI was uniform across all numerical values for a given test within the dichotomous positive or negative category, given fixed values of the other two tests. We performed a limited sensitivity analysis examining the effect of violating this assumption, and it did not substantially change the results (Section EIII). Areas under the ROC curve (AUC) were calculated using the trapezoidal rule implemented via trapz in the R pracma package (32); 95% credible intervals (CrI) for each AUC were obtained by sampling 1,000 iterations of the posterior distribution of PPV that was generated by the Markov chain Monte Carlo analysis.

To understand the tradeoffs in diagnostic errors (i.e., false-positive and false-negative test results), we compared the estimated numbers and types of errors for each of the test cutoffs in three hypothetical populations of 1,000 people with an LTBI prevalence of 5%, 30%, and 50%. These prevalences were chosen to approximately represent the general U.S. population, our study population, and a high-prevalence population of non–U.S.-born people (33, 34).

We applied the LCA model-derived estimates of PPV for each of the eight combinations to simulate the test characteristics of performing two sequential LTBI tests. We examined three possible strategies: TST and QFT (TST–QFT), TST and TSPOT (TST–TSPOT), and

QFT and TSPOT (QFT–TSPOT). Sequential same-test combinations were not explored as participants were tested only once with each test. For purposes of this analysis, the second test was used as a confirmatory test: Participants with two positive tests would be interpreted as positive for LTBI, and discordant results were interpreted as negative for LTBI. Uncertainty associated with estimates of both population parameter values and PPV was incorporated by creating 1,000 bootstrapped datasets of equal size to the original study population by sampling with replacement from that population and associating each dataset with a randomly selected set of PPV drawn from the LCA posterior distribution (*see* Section EIV).

## Results

Among 22,131 total participants in the original cohort, 14,167 (64%) were non–U.S.-born, HIV-negative, 5 years or older, nonclose contacts of TB cases, and had valid results for all three tests (Figure 1). Among those included in the analysis, 51.5% were female, the median age at enrollment was 29 years (interquartile range [IQR], 19–41), and the median time in the United States was 1.5 months (IQR, 1–4.5); 61% were born in Asia, 19% in Africa, and 14.5% in Central America (Table 1).

The ROC for the TST had an AUC of 0.81 (95% CrI, 0.78–0.86) (Figure 2A). The estimated sensitivity/specificity of the TST at cutoffs of 5, 10, and 15 mm were 86.5%/61.6%, 81.7%/71.3%, and 55.6%/88.0%, respectively (Table 2). At an estimated LTBI prevalence of 30.9% in our study population, the 5-, 10-, and 15-mm cutoffs correspond to PPV of 50.2%, 56.0%, and 67.4%. Corresponding negative predictive values (NPVs) for the 5-, 10-, and 15-mm cutoffs were 91.1%, 89.7%, and 81.6%.

The ROC curve for the QFT had an AUC of 0.89 (95% CrI, 0.86–0.93) (Figure 2B). The estimated sensitivity/specificity of the QFT at cutoffs of 0.35, 0.7, and 1.0 IU/mL were 77.7%/98.3%, 66.9%/99.1%, and 61.5%/99.5%, respectively. These cutoffs correspond to PPV of 95.3%, 97.1%, and 97.9%, with NPV of 90.8%, 87.0%, and 85.2% (Table 2).

The ROC curve for the TSPOT had an AUC of 0.92 (95% CrI, 0.88–0.96) (Figure 2C). Estimated sensitivity/specificity of the TSPOT at cutoffs of 5, 6, 7, and 8 spots were 79.2%/96.7%, 76.8%/97.7%, 74.0%/98.6%, and 71.8%/99.5%, respectively. These cutoffs correspond to PPV of 91.2%, 93.7%, 95.9%, and 98.5%, with NPV of 91.2%, 90.4%, 89.5%, and 88.8%. The AUC for TSPOT was significantly greater than that for TST (median difference, 0.10; 95% CrI, 0.096–0.11) and for QFT (median difference, 0.024; 95% CrI, 0.018–0.034); the AUC for QFT was also greater than for TST (median difference, 0.077; 95% CrI, 0.067–0.086). Although the 95% CrI for AUC for TST did not overlap with the 95% CrI for AUCs for QFT or TSPOT, the QFT and TSPOT AUC CrIs did overlap, so the statistically significant difference between the QFT and TSPOT AUCs is unlikely to be clinically meaningful (*see* Figure E1).

For the dual testing strategy using the same underlying 30.9% LTBI prevalence, the sensitivity/specificity of the TST–QFT, TST–TSPOT, and QFT–TSPOT strategies were 73.1%/99.4%, 64.8/99.8% and 65.3%/100.0% respectively. The corresponding PPV/NPV

were 98.2%/89.2%, 99.3%/86.4%, and 100%/86.6% (Table 2). The characteristics of the TST–QFT strategy were similar to the test characteristics of TSPOT using a cutoff of eight spots, and the characteristics of the TST–TSPOT strategy were similar to the test characteristics of QFT at a cutoff of 1.0 IU/mL.

Figures 3–5 illustrate the tradeoffs in the estimated number and type of errors (false positives + false negatives) for hypothetical populations of 1,000 people at LTBI prevalences of 5%, 30%, and 50%. Overall, IGRAs make fewer total errors than the TST at the three explored prevalences. Figure 3 represents the number of total errors and type of errors by diagnostic cutoff at a low LTBI prevalence of 5%. Using a TST cutoff of 15 mm results in 236 fewer total errors than when using a cutoff of 5 mm. QFT performs similarly at this prevalence at all the evaluated cutoffs, although the type of error varies on the basis of cutoff selection. TSPOT at a cutoff of eight spots makes fewer total errors than TSPOT at a cutoff of five spots. At an underlying LTBI prevalence of 5%, TSPOT at a diagnostic cutoff of eight spots and QFT at a cutoff of 1.0 IU/mL make the fewest total errors and minimize the number of false-positive results.

Figure 4 represents the estimated total number and type of errors by cutoff in a population with an LTBI prevalence of 30%. TST at 15 mm makes the fewest errors compared with the 5 and 10 mm cutoffs. QFT at the standard 0.35 IU/mL cutoff minimizes the number of total errors and false-negative results for this test. TSPOT makes a similar number of errors at the four explored cutoffs, but the proportion of false-negative results is lowest at a cutoff of five spots. In summary, at an LTBI prevalence of 30%, QFT at a diagnostic cutoff of 0.35 IU/mL makes the fewest errors among the three tests, followed by TSPOT at a cutoff of five spots. These cutoffs minimized the number of false negatives.

Figure 5 represents the estimated number and type of errors by cutoff in a population with an underlying LTBI prevalence of 50%. TST at a cutoff of 10 mm, QFT at 0.35 IU/mL, and TSPOT at five spots make the fewest errors by reducing the number of false-negative results.

## Discussion

Our study used outputs from a Bayesian LCA model to explore changes in diagnostic accuracy by varying test cutoffs and dual testing in a cohort of non–U.S.-born, HIV-negative participants 5 years and older residing in the United States. Our findings show that IGRAs have a better overall test performance (or higher AUCs) than TST, that alternative IGRA cutoffs depending on estimated LTBI prevalence can be used to minimize errors, and that dual testing strategies are comparable to single testing at higher IGRA cutoffs. These observations support the consideration of lower IGRA cutoffs to improve test accuracy in the diagnosis of LTBI in high-risk populations.

In our study cohort, IGRAs performed better than the TST, and overall, TSPOT used alone was the best predictor of LTBI. The UK Prognostic Evaluation of Diagnostic IGRAs Consortium (UK PREDICT TB) study (35) explored the predictive ability of all three tests for progression to TB disease in a population that included migrants from high-burden countries and contacts of TB cases living in the United Kingdom, a low-burden country.

Similar to our study, UK PREDICT TB found that the TSPOT had the best PPV for prediction of incident TB, which presumably correlates with PPV for LTBI. Test sensitivities for all tests were lower in UK PREDICT TB than in the present study; the reasons for this are unclear but may relate to the large number of recent TB contacts in UK PREDICT TB, who were excluded from our cohort. The inferior performance of TST in both the present study and UK PREDICT TB presumably relates to worse specificity than IGRAs because of BCG crossreactivity in these highly–BCG-vaccinated populations.

Maximizing test accuracy or reducing the number of total errors is as important as balancing the error tradeoffs. Higher cutoffs in a low prevalence population maximize test accuracy by increasing specificity and PPV. This translates into fewer false-positive results and would prevent unnecessary testing, medical evaluations, and treatment. At the same time, the lower NPV results in missing a greater number of true LTBI cases than at lower cutoffs. Contrarily, in high prevalence assumptions, lower cutoffs maximize test accuracy by increasing test sensitivity and NPV. Minimizing the number of false-negative results or missed cases in high-prevalence settings would help ensure that resources are appropriately allocated to the subpopulation of patients requiring additional testing, evaluation, and treatment. However, the tradeoff associated with lowering cutoffs and increasing sensitivity is a decrease in PPV that translates into a greater number of false positives that need to be evaluated, potentially burdening the health system.

Sequential testing is recommended in the United States for persons at low risk, given the high risk of false-positive results. However, although not generally recommended as clinical practice in the United States, sequential testing with the TST and IGRAs has been used in individuals at high risk for LTBI or progression to TB disease to confirm the diagnosis as well. The sequential testing strategy in high-risk populations has proven cost-effective in immigrants (16, 17), healthcare workers (14), and close contacts of TB cases (13). In our analysis, the two-test strategies reduced false-positive results and had similar characteristics to single IGRAs at higher cutoffs: The TST–QFT strategy has a similar sensitivity, specificity, PPV, and NPV to a TSPOT with a cutoff of eight spots, whereas the TST–TSPOT combination has a similar sensitivity, specificity, PPV and NPV to a QFT cutoff of 1.0 IU/mL. These results can be used as model inputs in cost-effectiveness analyses that would formally assess various LTBI test cutoffs and sequential testing strategies.

To our knowledge, this is the first study using LCA to explore the variation in test performance by changing test cutoffs. LCA has been previously used to compare test performance of the TST and IGRAs among hospitalized South African children suspected of having pulmonary TB, people living with HIV in a low-incidence setting, healthcare workers in a low-incidence setting, and a TB and HIV endemic setting (20, 36–38).

## Limitations

Our analysis had several limitations. The participants were not tested in a sequential manner, as all three tests were performed on the same day. However, we used this simultaneous testing as a surrogate for sequential testing by assuming only those with two positive tests have LTBI. This approach is recommended in U.S. practice guidelines for persons at low risk of TB (12). Second, we used the QFT test instead of the currently available version of

the test, QFT-Plus, as QFT-Plus was not available when the study began. However, a study with a subset of the original cohort that received both QFTs demonstrated a 94% agreement between QFT-Plus and QFT (39), and a smaller prospective cohort demonstrated QFT-Plus had roughly comparable performance to prior studies of QFT in predicting incident TB (40). A key (but untestable) assumption used for the ROC curve analysis was that given the fixed values of the other two tests, the numerical value of a third test had no correlation with the risk of LTBI. We did perform a sensitivity analysis to examine this assumption but cannot entirely rule out a residual correlation that might have affected the results. An alternative approach would have been to run separate latent class models for each test at each cutoff; this approach was not used because it was computationally prohibitive. Lastly, our analysis was performed on a non-random sample of non–US-born, HIV-negative individuals aged 5 years or older that was a cohort of relatively young recent immigrants. Although our results are robust in a broad population, as the main principles surrounding changing the diagnostic cutoffs are generalizable, specific performance may vary on the basis of age and the presence of comorbidities. Our results may be less applicable to older populations, individuals who immigrated less recently, or people with immunosuppression (37).

### Conclusions

Our study found that in a cohort of non–U.S.-born, HIV-negative individuals aged 5 years or older using lower IGRA cutoffs, specifically the 0.35 IU/mL cutoff for QFT and five spots for TSPOT, improves test accuracy and that the predictive values of dual testing with TST and an IGRA are comparable to using higher IGRA cutoffs. Cost-effectiveness analyses are needed to evaluate the clinical applicability of these findings.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgment:

The authors would like to acknowledge all of the study participants, without whose generosity this work would not be possible. The authors would also like to acknowledge the hard work of all Tuberculosis Epidemiologic Studies Consortium investigators and study coordinators.

Supported by a contract from the Centers for Disease Control and Prevention to fund the Tuberculosis Epidemiologic Studies Consortium-II and the National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention.

## References

1. Houben RM, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. PLoS Med 2016;13: e1002152. [PubMed: 27780211]
2. Yuen CM, Krapivina TM, Kazennyy BY, Kiryanova EV, Aksenova VA, Gordina A, 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.
3. Haddad MB, Raz KM, Lash TL, Hill AN, Kammerer JS, Winston CA, et al. Simple estimates for local prevalence of latent tuberculosis infection, United States, 2011–2015. Emerg Infect Dis 2018;24:1930–1933. [PubMed: 30226174]

4. Miramontes R, Hill AN, Yelk Woodruff RS, Lambert LA, Navin TR, Castro KG, 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. [PubMed: 26536035]
5. Shrestha S, Cherng S, Hill AN, Reynolds S, Flood J, Barry PM, et al. Impact and effectiveness of state-level tuberculosis interventions in California, Florida, New York, and Texas: a model-based analysis. Am J Epidemiol 2019;188:1733–1741. [PubMed: 31251797]
6. Zwerling A, Behr MA, Verma A, Brewer TF, Menzies D, Pai M. The BCG World Atlas: a database of global BCG vaccination policies and practices. PLoS Med 2011;8:e1001012. [PubMed: 21445325]
7. Gray J, Reves R, Johnson S, Belknap R. 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–e23. [PubMed: 22057704]
8. Weinfurter P, Blumberg HM, Goldbaum G, Royce R, Pang J, Tapia J, et al. ; Tuberculosis Epidemiological Studies Consortium. Predictors of discordant tuberculin skin test and QuantiFERONV® -TB Gold In-Tube results in various high-risk groups. Int J Tuberc Lung Dis 2011;15: 1056–1061. [PubMed: 21740668]
9. Cattamanchi A, Smith R, Steingart KR, Metcalfe JZ, Date A, Coleman C, 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;56:230–238. [PubMed: 21239993]
10. Zellweger JP, Sotgiu G, Corradi M, Durando P. The diagnosis of latent tuberculosis infection (LTBI): currently available tests, future developments, and perspectives to eliminate tuberculosis (TB). Med Lav 2020;111:170–183. [PubMed: 32624559]
11. 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.
12. Lewinsohn DM, Leonard MK, LoBue PA, Cohn DL, Daley CL, Desmond E, 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;64:111–115. [PubMed: 28052967]
13. Nienhaus A, Schablon A, Costa JT, Diel R. Systematic review of cost and cost-effectiveness of different TB-screening strategies. BMC Health Serv Res 2011;11:247. [PubMed: 21961888]
14. del Campo MT, Fouad H, Solís-Bravo MM, Sánchez-Uriz MA, Mahíllo-Fernández I, Esteban J. Cost-effectiveness of different screening strategies (single or dual) for the diagnosis of tuberculosis infection in healthcare workers. Infect Control Hosp Epidemiol 2012;33: 1226–1234. [PubMed: 23143360]
15. Pooran A, Booth H, Miller RF, Scott G, Badri M, Huggett JF, et al. Different screening strategies (single or dual) for the diagnosis of suspected latent tuberculosis: a cost effectiveness analysis. BMC Pulm Med 2010;10:7. [PubMed: 20170555]
16. Serre-Delcor N, Treviño-Maruri B, Tortola MT, Fernández-Quevedo M, Soriano-Arandes A, Oliveira-Souto I, et al. Sequential strategy for the LTBI screening of newly-arrived immigrants in vulnerable social situations. Enferm Infecc Microbiol Clin (Engl Ed) 2018;36: 550–554. [PubMed: 29223317]
17. Salinas C, Ballaz A, Díez R, Aguirre U, Anton A, Altube L. [Tuberculosis screening program for undocumented immigrant teenagers using the QuantiFERON(®)-TB Gold In-Tube test]. Med Clin (Barc) 2015;145: 7–13. [PubMed: 24747025]
18. Stout JE, Wu Y, Ho CS, Pettit AC, Feng PJ, Katz DJ, et al. ; Tuberculosis Epidemiologic Studies Consortium. Evaluating latent tuberculosis infection diagnostics using latent class analysis. Thorax 2018;73: 1062–1070. [PubMed: 29982223]
19. Doan TN, Eisen DP, Rose MT, Slack A, Stearnes G, McBryde ES. Interferon-gamma release assay for the diagnosis of latent tuberculosis infection: a latent-class analysis. PLoS One 2017;12: e0188631. [PubMed: 29182688]

20. Schumacher SG, van Smeden M, Dendukuri N, Joseph L, Nicol MP, Pai M, et al. Diagnostic test accuracy in childhood pulmonary tuberculosis: a Bayesian latent class analysis. Am J Epidemiol 2016; 184:690–700. [PubMed: 27737841]
21. Ho CS, Feng PI, Narita M, Stout JE, Chen M, Pascopella L, et al. ; Tuberculosis Epidemiologic Studies Consortium. Comparison of three tests for latent tuberculosis infection in high-risk people in the USA: an observational cohort study. Lancet Infect Dis 2022;22:85–96. [PubMed: 34499863]
22. American Thoracic Society. Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR Recomm Rep 2000;49:1–51.
23. Food and Drug Administration. QuantiFERON-TB Gold In-Tube P010033/S011. 2001 [accessed 2021 Jul 16]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf/P010033B.pdf.
24. Food and Drug Administration. T-SPOT-TB - P070006. 2008 [accessed 2021 Jul 16]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf7/p070006b.pdf.
25. Gupta RK, Lipman M, Jackson C, Sitch AJ, Southern J, Drobniewski F, et al. Quantitative IFN-g release assay and tuberculin skin test results to predict incident tuberculosis. A prospective cohort study. Am J Respir Crit Care Med 2020;201:984–991. [PubMed: 31825645]
26. Ghassemieh BJ, Attia EF, Koelle DM, Mancuso JD, Narita M, Horne DJ. Latent tuberculosis infection test agreement in the national health and nutrition examination survey. Am J Respir Crit Care Med 2016;194: 493–500. [PubMed: 26890477]
27. European Centre for Disease Prevention and Control. Mastering the basics of TB control: development of a handbook on TB diagnostic methods. Stockholm: European Centre for Disease Prevention and Control; 2011.
28. Rosales-Klintz S, Bruchfeld J, Haas W, Heldal E, Houben RMGJ, van Kessel F, et al. Guidance for programmatic management of latent tuberculosis infection in the European Union/European economic area. Eur Respir J 2019;53:1802077. [PubMed: 30655449]
29. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics 1996;52: 797–810. [PubMed: 8805757]
30. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 2001;57: 158–167. [PubMed: 11252592]
31. R Core Team. R: a language and environment for statistical computing. 2021 [accessed 2021 Jul 16]. Available from: https://www.R-project.org/.
32. Borchers H. Pracma: practical numerical math functions, version 2.3.3. 2021 [accessed 2021 Jul 16]. Available from: https://CRAN.R-project.org/package=pracma.
33. Mancuso JD, Diffenderfer JM, Ghassemieh BJ, Horne DJ, Kao TC. The prevalence of latent tuberculosis infection in the United States. Am J Respir Crit Care Med 2016;194:501–509. [PubMed: 26866439]
34. Collins JM, Stout JE, Ayers T, Hill AN, Katz DJ, Ho CS, et al. ; Tuberculosis Epidemiologic Studies Consortium. Prevalence of latent tuberculosis infection among non-US-born persons by country of birth-United States, 2012–2017. Clin Infect Dis 2021;73: e3468–e3475. [PubMed: 33137172]
35. Abubakar I, Drobniewski F, Southern J, Sitch AJ, Jackson C, Lipman M, et al. ; PREDICT Study Team. Prognostic value of interferon-γ release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. Lancet Infect Dis 2018;18:1077–1087. [PubMed: 30174209]
36. Adams S, Ehrlich R, Baatjies R, Dendukuri N, Wang Z, Dheda K. Evaluating latent tuberculosis infection test performance using latent class analysis in a TB and HIV endemic setting. Int J Environ Res Public Health 2019;16:2912. [PubMed: 31416206]
37. Pettit AC, Stout JE, Belknap R, Benson CA, Seraphin MN, Lauzardo M, et al. ; Tuberculosis Epidemiologic Studies Consortium (TBESC). Optimal testing choice and diagnostic strategies for latent tuberculosis infection among US-born people living with human immunodeficiency virus (HIV). Clin Infect Dis 2021;73:e2278–e2284. [PubMed: 32761083]
38. Girardi E, Angeletti C, Puro V, Sorrentino R, Magnavita N, Vincenti D, et al. Estimating diagnostic accuracy of tests for latent tuberculosis infection without a gold standard among healthcare workers. Euro Surveill 2009;14:19373. [PubMed: 19883555]

39. Venkatappa TK, Punnoose R, Katz DJ, Higgins MP, Banaei N, Graviss EA, et al. Comparing QuantiFERON-TB gold plus with other tests to diagnose Mycobacterium tuberculosis infection. J Clin Microbiol 2019;57:e00985–19. [PubMed: 31462550]

40. Gupta RK, Kunst H, Lipman M, Noursadeghi M, Jackson C, Southern J, et al. Evaluation of QuantiFERON-tb gold plus for predicting incident tuberculosis among recent contacts: a prospective cohort study. Ann Am Thorac Soc 2020;17:646–650. [PubMed: 32083944]

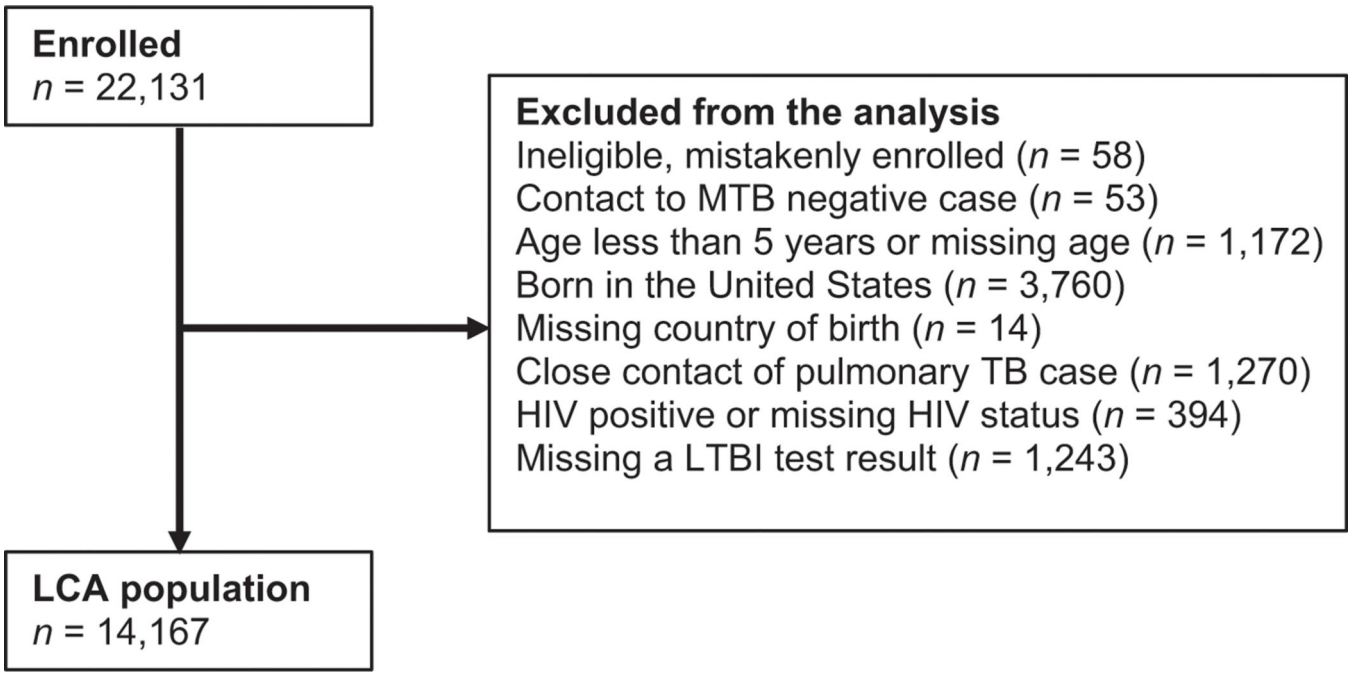


**Figure 1.**

Analysis population. HIV = human immunodeficiency virus; LCA = latent class analysis; LTBI = latent tuberculosis infection; MTB = *Mycobacterium tuberculosis*; TB = tuberculosis.

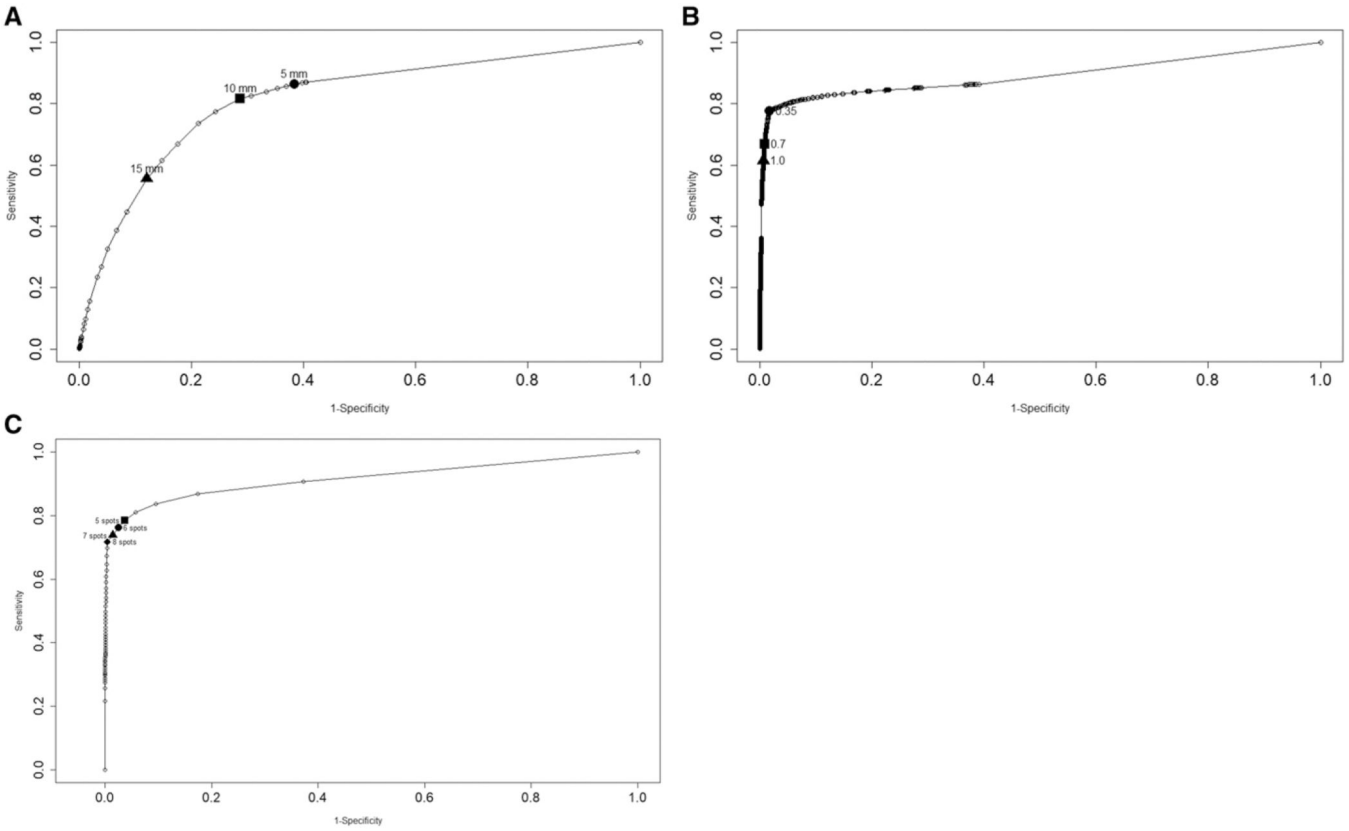


**Figure 2.**

Receiver operating characteristic curves for (*A*) tuberculin skin test, (*B*) QuantiFERON Gold in-Tube, and (*C*) T-SPOT.*TB*.

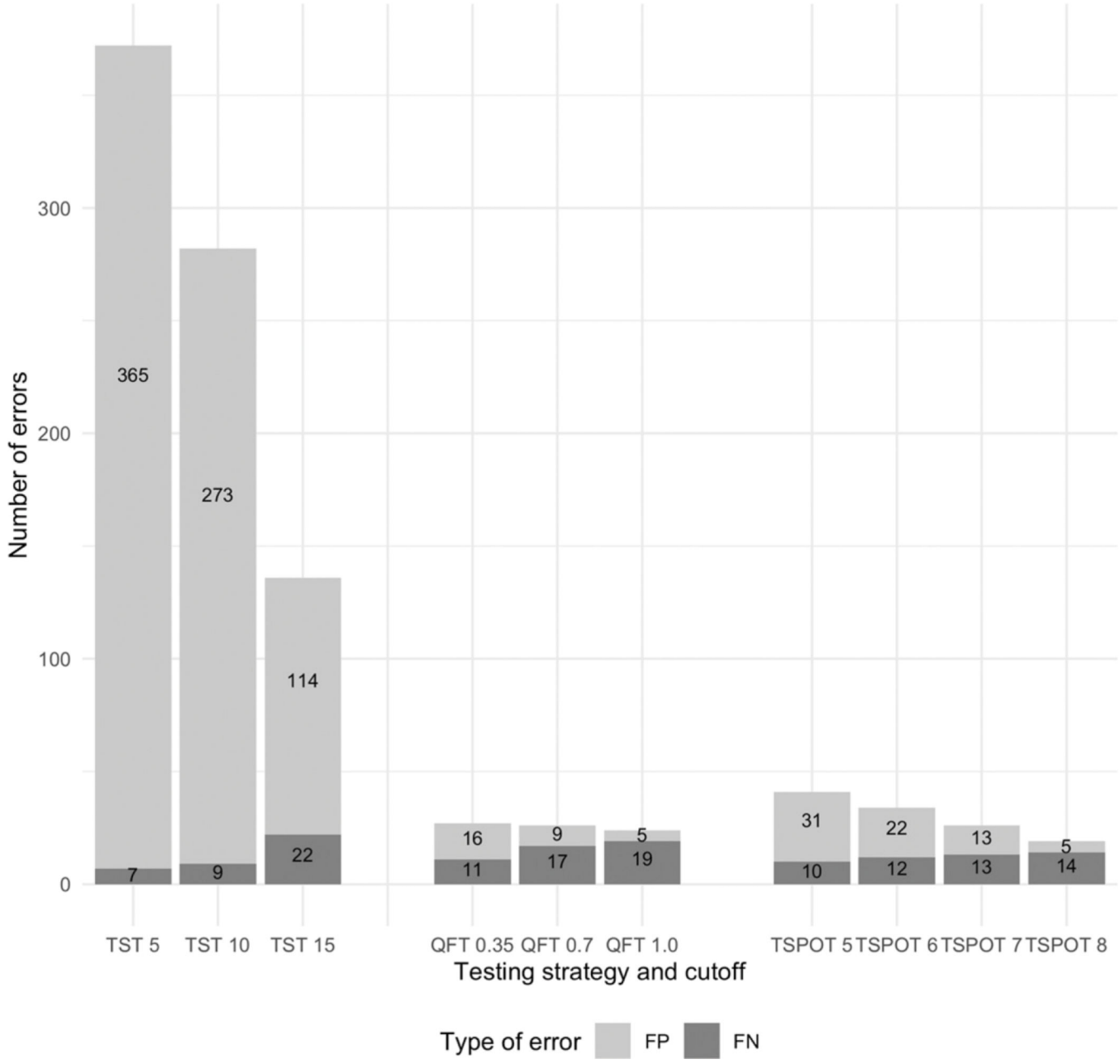


**Figure 3.**

Change in the number and type of diagnostic errors in a hypothetical population with a latent tuberculosis infection prevalence of 5% ($n$ = 1,000). FN = false negative; FP = false positive; QFT = QuantiFERON Gold in-Tube; TSPOT = T-SPOT.*TB*; TST = tuberculin skin test.

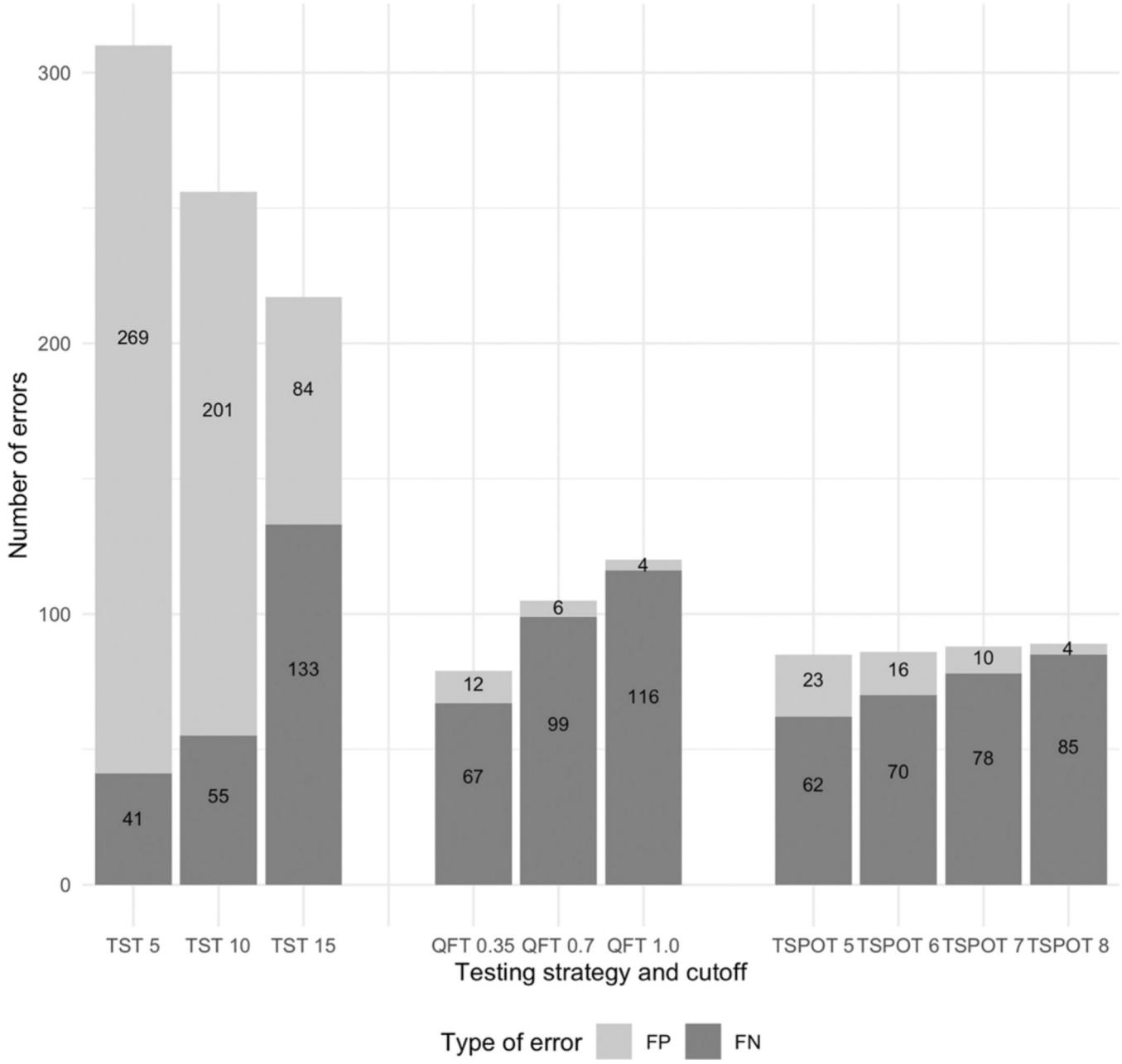


**Figure 4.**

Change in the number and type of diagnostic errors in a hypothetical population with a latent tuberculosis infection prevalence of 30% ($n$ = 1,000). FN = false negative; FP = false positive; QFT = QuantiFERON Gold in-Tube; TSPOT = T-SPOT.*TB*; TST = tuberculin skin test.

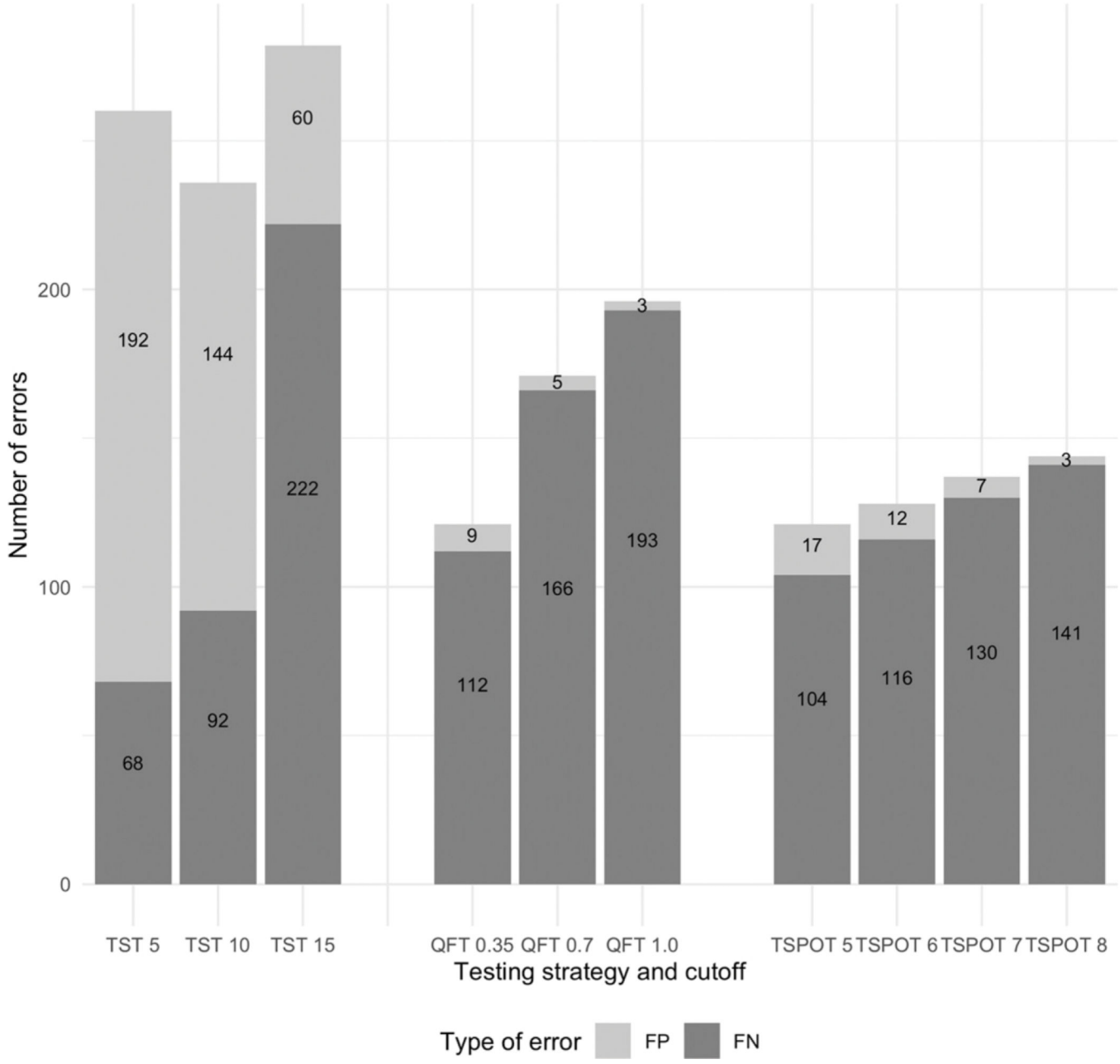


**Figure 5.**

Change in the number and type of diagnostic errors in a hypothetical population with a latent tuberculosis infection prevalence of 50% ($n$ = 1,000). FN = false negative; FP = false positive; QFT = QuantiFERON Gold in-Tube; TSPOT = T-SPOT.*TB*; TST = tuberculin skin test.

**Table 1.**

Demographic characteristics of the study population; non–U.S.-born, human immunodeficiency virus-seronegative people ages 5 years and older with a valid tuberculin skin test, QuantiFERON Gold in-Tube, and T-SPOT.*TB* enrolled between 2012 and 2017

| Characteristic | Total (*n* = 14,167) |
|---|---|
| Female sex, *n* (%) | 7,300 (51.5) |
| Age at enrollment, median (IQR) | 29 (19–41) |
| Time in the United States (mo), median (IQR) | 1.5 (1–4.5) |
| Region of birth, *n* (%) | |
| Asia | 8,709 (61.5) |
| Africa | 269 (18.9) |
| Central America | 1,380 (9.7) |
| South America | 299 (2.1) |
| Oceania | 269 (1.9) |
| Europe | 146 (1.0) |
| North America | 678 (4.8) |
| Not reported | 8 (0.1) |
| Medical history, *n* (%) * | |
| BCG vaccination | 9,165 (72.9) |
| Diabetes | 578 (4.1) |
| Cancer | 58 (0.4) |
| Smoker † | 2,632 (22.5) |
| Recreational drug use ‡ | 280 (2.4) |

*Definition of abbreviations*: BCG = *Bacillus* Calmette-Guérin; IQR = interquartile range.

*Participants who responded “do not know” or refused to provide information are not considered in the total count for each characteristic. Patients may have more than one of the listed medical histories.

†Defined as having smoked 100 cigarettes in life.

‡Ever used recreational drugs.

**Table 2.**

Estimated characteristics of single tests per cutoff value and dual testing for latent tuberculosis infection (cohort prevalence 30.9%)

| Testing Strategy and Cutoff | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) |
|---|---|---|---|---|
| TST | | | | |
| 5 mm | 86.5 | 61.6 | 50.2 | 91.1 |
| 10 mm | 81.7 | 71.3 | 56.0 | 89.7 |
| 15 mm | 55.6 | 88.0 | 67.4 | 81.6 |
| QFT | | | | |
| 0.35 IU/mL | 77.7 | 98.3 | 95.3 | 90.8 |
| 0.7 IU/mL | 66.9 | 99.1 | 97.1 | 87.0 |
| 1.0 IU/mL | 61.5 | 99.5 | 97.9 | 85.2 |
| TSPOT | | | | |
| 5 spots | 79.2 | 96.7 | 91.2 | 91.2 |
| 6 spots | 76.8 | 97.7 | 93.7 | 90.4 |
| 7 spots | 74.0 | 98.6 | 95.9 | 89.5 |
| 8 spots | 71.8 | 99.5 | 98.5 | 88.8 |
| Dual testing strategies * | | | | |
| TST–-QFT | 73.1 | 99.4 | 98.2 | 89.2 |
| TST–TSPOT | 64.8 | 99.8 | 99.3 | 86.4 |
| QFT–TSPOT | 65.3 | 100.0 | 100.0 | 86.6 |

*Definition of abbreviations*: NPV = negative predictive value; PPV = positive predictive value; QFT = QuantiFERON Gold in-Tube; TSPOT = T-SPOT.*TB*; TST = tuberculin skin test.

* Latent tuberculosis infection was diagnosed only if the participant had a positive result for both tests at standard U.S. cutoffs: TST ≥ 10 mm per Centers for Disease Control and Prevention guidelines (22); QFT ≥ 0.35 IU/mL; and TSPOT ≥ 8 spots per U.S. Food and Drug Administration-approved labeling (23, 24).