WHO consolidated guidelines on tuberculosis

Module 3: diagnosis. Tests for TB infection

Web Annex A

Accuracy of *Mycobacterium tuberculosis* antigen-based skin tests: a systematic review and meta-analysis

Review team: Yohhei Hamada[1], Elena Surkova[2], Irina Kontsevaya[3], Ting Ting Wang[1], Victoria Liu[4], Liliya Eugenevna Ziganshina[5,6], Molebogeng X Rangaka[1,7]

1. Institute for Global Health, University College London, UK
2. Royal Brompton Hospital. Part of Guy's and St Thomas' NHS Foundation Trust London, UK
3. Research Center Borstel, Germany
4. London School of Hygiene and Tropical Medicine, UK
5. Institute of Fundamental Medicine and Biology of Kazan Federal University, Russian Federation,
6. Cochrane Russia, Russian Federation
7. Division of Epidemiology and Biostatistics & CIDRI-AFRICA, University of Cape Town, South Africa

Technical advisor: Claudia Denkinger[1], Ibrahim Abubakar[2]

1. Centre of Infectious Disease, University of Heidelberg, Germany
2. Faculty of Population Health Science, University College London, UK

Produced in preparation for the WHO guideline group meeting on "Skin-based tests for TB infection, 4-6 February 2022]".

Report version 2.0, Date 8 January 2022

World Health Organization

WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for TB infection. Web Annex A. Accuracy of *Mycobacterium tuberculosis* antigen-based skin tests: a systematic review and meta-analysis/ Yohhei Hamada, Elena Surkova, Irina Kontsevaya, Ting Ting Wang, Victoria Liu, Liliya Eugenevna Ziganshina et al.

ISBN 978-92-4-005659-6 (electronic version)

Suggested citation. Hamada Y, Surkova E, Kontsevaya I, Wang TT, Liu V, Ziganshina LE et al. Web Annex A. Accuracy of *Mycobacterium tuberculosis* antigen-based skin tests: a systematic review and meta-analysis. In: WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for TB infection. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

This publication forms part of the WHO guideline entitled *WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for TB infection*. It is being made publicly available for transparency purposes and information, in accordance with the *WHO handbook for guideline development*, 2nd edition (2014).

Contents

1. Executive summary

Background

Diagnosis and treatment of TB infection (TBI) is a pillar of the WHO End TB strategy to achieve a global reduction in TB incidence by 2035. Accurate diagnostic tests are needed to better identify individuals who would benefit from preventive treatment. If found to have at least equivalent diagnostic accuracy compared to existing screening tests, the interferon gamma release assays (IGRA) and purified protein derivative (PPD) tuberculin skin test (TST), the new *Mycobacterium tuberculosis (*MTB*)* specific skin-based tests have the potential to improve access and/or reduce cost in low-resource settings. We updated a previous review performed by Krutikov et al. to synthesise current evidence on the diagnostic performance of novel skin-based tests for TB infection (TBST) compared to currently available *in vitro* IGRA tests and TST against a range of pre-defined reference standards.

Methods

We updated the previous search that identified papers published until 20 Oct 2020. The search was carried out in Medline, Embase, e-library, Chinese Biomedical Literature Database, and China National Knowledge Infrastructure for all studies published until 30 July 2021 with no language restrictions. Following title and abstract screening, full texts were reviewed according to eligibility criteria. Included studies reported performance of tests alone or against comparator; these were evaluated against a hierarchy of pre-defined reference standards for TBI; efficacy of TB preventive treatment based on TB test results; predictive performance; correlation with exposure gradient; sensitivity in active TB; specificity in populations at low risk for TB infection; test agreement. Pooled estimates were obtained via random-effects meta-analyses. Study quality was assessed using QUADAS-2.

We conducted sensitivity analyses of the specificity estimates using a less restrictive criteria, we examined the proportion of negative results as a proxy for specificity regardless of background TB incidence and estimated the following measures:

1) Differences in proportion of negative results between index tests and comparator tests evaluated in the same cohort (i.e. 'specificity difference')
2) Proportion of negative TBST results in participants with negative IGRA results (i.e. agreement of negative results).

Our primary analysis prioritised three-way head-to-head assessments of TBST vs comparator tests to facilitate comparability, and consequently did not pool different TBST because of substantial heterogeneity in study designs. However, based on a request from WHO, we also conducted post-hoc analysis combining data across different TBST by ignoring heterogeneity with regards to study designs, cut-off used, and populations, to support WHO's class-based recommendations of Mtb specific skin-based tests.

Results

We identified three novel skin tests using ESAT6 and CFP10 antigens: C-Tb (Serum Institute of India, India), Diaskintest (Generium, Russian Federation), and C-TST (formerly known as ESAT6-CFP10 test, Anhui Zhifei Longcom, China]). Additionally, we identified the DPPD test, which contains a recombinant protein rv0061, named DPPD. The gene coding DPPD is present only in the MTB complex (including *Mycobacterium bovis*-

BCG) and is absent in non-tuberculosis mycobacteria. We identified and included for review five studies for C-Tb, 34 for Diaskintest, four for C-TST. DPPD is still undergoing evaluation and only one study was found. The test is not ready for commercialization and not included in the pooled analysis but included in the main report for completeness.

No longitudinal studies evaluating index test performance were identified. Thresholds for positivity used for the skin tests varied, and included any induration, 5mm, or 7mm for the Diaskintest, 5mm for C-Tb and C-TST. For the TST a 5mm or 15mm cut-off was used according to the risk population tested reported disaggregated or aggregated for a single cohort (TST5mm/15mm).

Three C-Tb studies were conducted in South Africa, while two were in Spain and UK, respectively. 33/34 studies for Diaskintest were done in Russian Federation and the remaining one in Ukraine. All C-TST studies were done in China.

Study quality varied, with low concerns about applicability for all studies but high risk of bias in Diaskintest studies, largely due to the study design (using data collected under routine clinical practice rather than designed as diagnostic accuracy studies) and a lack of clarity on the population selection criteria.

Sensitivity

Two studies assessed test sensitivity of Diaskintest in HIV-negative adults with active tuberculosis with direct comparison with TST and IGRA (i.e. three-way head-to-head). Pooled sensitivity for Diaskintest$_{5\,mm}$ was 91% (95% CI: 82–96%), 88% (95%CI: 78–94%) for TST$_{5\,mm}$, 90% (95%CI: 79–95%) for QFT, and 91% (95%CI: 80–96%) for TSPOT.TB. In four head-to-head studies including both adults and children with and without HIV, C-Tb sensitivity was 75% (95%CI:70-78%) vs TST$^{5mm/15mm}$ 79% (95%CI:68-86%), IGRA 72% (95%CI:63-79%). The pooled sensitivity across 6 studies on C-Tb and Diaskintest was 78% (95%CI : 71-84%) for TBST, 77% (95%CI : 66-85%) for IGRA, and 84% (95%CI : 79-89%) for TST$_{5\,mm}$.

Three studies evaluated the sensitivity of the C-TST. The pooled sensitivity at the ≥5mm induration was 86% (95%CI: 83-89%). In one study, the sensitivity of C-TST (90% [73-98%]) was similar to that of T-SPOT.TB (89% [78-95%] and TST $^{10\,mm}$ (87% [76-94%] and slightly higher than TST $^{10\,mm}$ (82% [69-90%].

When combining all studies on Diaskintest, C-Tb, and C-TST, the pooled sensitivity was 76% (95%CI: 70-81%, 17 studies) in individuals with HIV-negative or unknown status and 63% (95%CI: 53-73%, 5 studies) in HIV-positive individuals.

Specificity

Two studies reported three-way head-to-head comparison of specificity in a low risk population: C-Tb 98% (95%CI:94-99%) vs TST^{15mm} 93% (95%CI:90-95%), IGRA 99% (95%CI:80-100%). There was no data on specificity for other tests using this pre-defined standard criteria and proxy estimation was warranted.

The differences in specificity estimates between Diaskintest^{5mm} and QFT ranged from -1.9 to 10.6% across 3 studies, with a pooled difference of 4.5% (95%CI: -13.1- 22.1). The differences were substantially larger between Diaskintest^{5mm} and TST^{5mm}: 29.9% (95%CI: -3.7- 63.5, 5 studies); The proportion with negative Diaskintest results among QFT-negative individuals in one study was 99.1% (95%CI 94.9-100.0).

Similar to other tests, the differences in specificity between C-TST and TST were higher than those between C-TST and IGRA. The differences in one study were 39.9% for TST$^{5\ mm}$ (95%CI: 33.8-45.6) 24.5% (95%CI 18.6-30.2) for TST$^{10\ mm}$ and 3.5% (95%CI: -1.4-8.4%) for TSPOT.TB. The proportion of negative C-TST results in IGRA-negative participants was 95% (95%CI: 93-97%) in one study in China..

When combining all TBST, the pooled difference in specificity between TBST and IGRA was 2.29% (95%CI: -1.60-6.18%, 6 studies on Diaskintest, C-Tb and C-TST) and that between TBST and TST was 33.47% (95%CI: 18.16-48.78%, 14 studies including Diaskintest, C-Tb, and C-TST). The pooled proportion negative TBST results among IGRA-negative healthy individuals in 3 studies on Diaskintest and C-TB was 95% (95%CI: 93-97%).

Agreement

In a mixed TB and non-TB cohort of two studies allowing a 3-way head-to-head comparison, Diaskintest pooled agreement with IGRA was 88% (95%CI:80-93%) vs TST-5mm cut-off (TST^{5mm}) 52% (95%CI:42-61%). C-Tb agreement with IGRA in active TB in 3 studies was 80% (95%CI:76-84%) vs TST$^{5mm/15mm}$ cut-off 76% (95%CI:69-82%). Considering all studies with at least two-way test comparisons, pooled agreement of Diaskintest with IGRA was 94.62% (95% CI 90.49–97.02; I^2 = 56.2%) in five studies in participants with any tuberculosis status. By contrast, the agreement between Diaskintest and TST$^{5\ mm}$ showed considerable heterogeneity; the pooled agreement was estimated in children with active tuberculosis (97.39% [96.39–98.12]) and children without TB (17.62% [6.79-38.60]).

Considering all studies without restricting to 3-way head-to-head studies, the pooled agreement of C-Tb with TST was similar, 81% (95% CI, 76-85%) at TST^{5mm} in HIV-infected and 76% (95% CI, 71-81%) at TST^{15mm} in HIV-uninfected TB patients. Test agreement among individuals without TB was reported in two studies. In one study, C-Tb and IGRA agreement ranged from 92% to 97% across sub-populations with different levels of TB exposure, while it was 78% and 81% in HIV-infected and uninfected individuals, respectively, in the second study. The agreement between C-Tb and the TST^{5mm} in these two studies was 83% and 87%, respectively.

In two studies including a mix of healthy individuals, TB patients, and patients with other pulmonary diseases, the pooled agreement between C-TST and IGRA was higher (85.96% [78.82-90.97%]) than that between C-TST and TST$^{5\ mm}$ (68.23% [55.48-78.74]) and TST $^{10\ mm}$ (71.28% [67.12-75.11]).

When combining all TBST, the pooled agreement with IGRA was 89% (95%CI: 83-93%, 8 studies) in people without TB and 86% (95%CI: 80-90%, 8 studies) in people with TB. The agreement with TST was 59% (95%CI: 45-72%, 16 studies) in people without TB and 88% (95%CI: 82-93%, 13 studies) in people with TB.

Correlation with exposure gradient

A gradient of test positivity was found according to the proximity of contacts to a confirmed TB case for Diaskintest and C-Tb studies.

Interpretation

Due to the lack of longitudinal cohorts among studies included in this review, outcomes pertaining to prediction for disease progression and efficacy of preventive therapy based on test results could not be evaluated.

Although the literature search revealed a larger number of studies evaluating Diaskintest than C-Tb and C-TST performance, a considerable proportion of Diaskintest studies were not primarily designed to evaluate test performance. As a result, there are a number of concerns that affect the quality of the studies. Furthermore, for all studies, not only those for Diaskintest, potential conflicts of interest are possible with many of the included studies given many were industry-led and/or funded studies.

Despite the limitations, the performance of novel skin tests appears similar to IGRA. These tests may enable precise and accessible TBI testing that does not require expensive laboratory facilities or venepuncture.

For all index tests, there were limited studies included in the review that evaluated test performance in different populations, including children under 5 years of age, HIV-infected individuals and contacts with well-defined exposure to TB-infected individuals.

This systematic review performed an extensive literature search in 3 languages to maximise the number of studies that could be included in the analysis. Although further information was requested from study authors where studies did not meet strict inclusion criteria, a limited number of responses were received, rendering a large number of studies ineligible for inclusion.

Conclusion

An overview of currently available data on the performance of novel skin tests for TBI diagnosis is presented. Test performance does not differ significantly from that reported for IGRA. Significant variability was seen in Diaskintest performance probably because of the observational nature of the studies affecting the quality of studies and the impact of BCG vaccination. Data quality was higher for C-Tb and C-TST studies. Although the review suggests TBST may enable precise and accessible TBI testing, more research is needed to address fully diagnostic accuracy and predictive performance in different at-risk populations in post-licensure studies.

2. Background

Approximately 25-27% of the world's population is estimated to have TB infection[1 2] with a lifetime risk of progression to active disease of 5-10%, which is higher in those with predisposing factors or the first 18 months after acquisition of infection.[3 4] These are important populations to target for testing and treatment of TB infection to prevent reactivation and subsequent transmission. Currently, available tests for TB infection are imperfect, as they cannot accurately distinguish between active TB disease and infection, nor are they useful predictors of progression to active disease.[5] Given the recognition of the identification and management of TB infection as an essential element of the End TB Strategy, research into more accurate diagnostic tests is critical to achieving these milestones.[6]

The diagnostic tests in current use are the tuberculin skin test (TST) and interferon-gamma release assay (IGRA). TST has relatively low specificity (false positives in those with previous recent BCG vaccination),[7] lacks sensitivity in immunosuppressed individuals (e.g. HIV infected), requires two clinic visits (one to administer the test and one to read the result), and failure to attend the clinic for evaluation of reaction within 48-72 hours renders the results invalid. Despite its limitations, due to its low cost and wide availability, it remains the most commonly used test for TB infection.

The IGRA measures T-cell release of Interferon-gamma (IFNɣ) following stimulation by ESAT-6 and CFP-10 antigens that are specific to the *Mycobacterium tuberculosis* (MTB) complex.[8] There are two types of IGRA: the enzyme-linked immunosorbent assay (ELISA)-based whole-blood method, and the enzyme-linked immunosorbent spot (ELISPOT) assay. Unlike the TST, IGRAs are not affected by prior BCG vaccination as the RD1 locus is specific to the MTB genome. Therefore these antigens are not present in *Mycobacterium bovis* BCG strain, used in BCG vaccines, or other non-tuberculous mycobacteria (NTM).[9] Moreover, compared to the TST, some IGRAs remain relatively unimpaired in HIV and other immunosuppressive conditions.[10] Thus, these are useful for evaluation of TB infection in BCG-vaccinated individuals and with high specificity, particularly in countries where BCG vaccination is administered after infancy and/or repeated vaccinations are given. However, the IGRA platforms are more expensive to run, requiring specialised kits, a qualified technician and an accredited laboratory in order to ensure test results are reproducible, as well as a phlebotomist to obtain blood samples.[5] Furthermore, large variability has been observed even if pre-analytical steps were performed within the recommendations of the manufacturer, limiting the reproducibility of the tests.[11]

Over the last decade, novel *Mycobacterium tuberculosis* specific skin-based tests for TB infection (TBST) have been developed that aim to maximise the advantages of the currently available implementation platforms. Examples of these are the C-Tb (Staten Serum Institut), Diaskintest (Generium) and C-TST (formerly known as ESAT6-CFP10 test [Anhui Zhifei Longcom]), all of which contain recombinant ESAT-6 (dimer) and CFP10 (monomer) antigens derived from MTB that may provide diagnostic performance improvements over the standard TST (particularly in respect to specificity). Another new test is DPPD skin test which contains a recombinant protein rv0061, named DPPD. The gene coding DPPD is present only in the MTB complex (including *Mycobacterium bovis*-BCG) and is absent in NTMs.[12] All tests use an intradermal injection of antigen and, like TST, are read after 48-72 hours as induration in mm.[13 14] Emerging evidence suggests that compared to IGRAs, these tests may have similar specificity[15] and provide more reliable results in children and HIV-infected cohorts, with the C-Tb, for example, using *Mycobacterium*

tuberculosis antigen-based skin test showing similar sensitivity in HIV-infected and uninfected individuals (although lower sensitivity was found among HIV+ individuals with CD4 counts below 100).[16]

We previously conducted a systematic review to synthesise current evidence on the diagnostic performance of TBST compared to that of currently available *in vitro* IGRA tests and TST.[17] The review suggested that those tests perform similarly to TST or IGRA. We have updated the review to inform the development of WHO guidelines.

Hierarchy of reference standards

The study of the diagnostic performance of tests for TB infection is hampered by a lack of an adequate reference standard. Existing tests for TB infection measure the cell-mediated immune response (memory T cell response) to exposure to TB antigens and are thus proxies for infection. As the diagnostic accuracy for LTBI cannot be directly assessed, we utilised a hierarchy of *a priori* agreed reference standards that also reflect diagnostic accuracy study designs previously used in the evaluation of IGRA (Figure 1).[18 19]

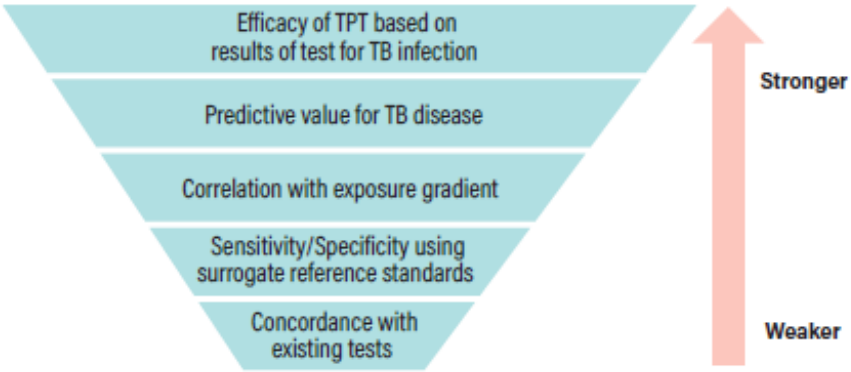


Figure 1: Hierarchy of reference standards[18]

3. Aims and objectives

Aim

To evaluate the performance of *Mycobacterium tuberculosis* specific TBST in at-risk populations compared to currently available *in vitro* IGRA tests or the TST.

PICO

1. Diagnostic performance (PICO question): Do TBST have similar or better diagnostic performance to TST or IGRA to detect infection with *M. tuberculosis?*

Table 1 presents detailed information about population, intervention, comparator and outcome.

Table 1. Population, intervention comparison and outcome for the study question.

Population	Intervention	Comparator	Outcome
– PLHIV; – <5 years – Household and other close contacts; – Other at-risk groups; ○ Immune compromised (individuals receiving anti-TNF-α treatment, dialysis, under preparation for an organ or haematological transplant, patients with silicosis, pregnant women, malnourished, diabetes mellitus, steroid use, smoker); ○ High risk of prior TB exposure (prisoners, health workers, immigrants from high TB burden countries, individuals with CXR abnormalities; homeless people and people who use drugs, inhabitants of high TB burden settings)[1] – BCG vaccinated vs non-vaccinated (in identified groups at risk of TB infection-stratified or in combination, as appropriate)	*Novel M. tuberculosis* antigen-based skin tests: – Diaskintest – C-Tb – C-TST – DPPD – Others	TST or IGRA	1. Efficacy of TB preventive treatment (TPT) based on diagnostic tests results; 2. Predictive value for progression to TB disease; 3. Correlation with exposure gradient; – 4.Sensitivity/Specificity for TB infecton[2]; – 5. Concordance with TST[3]; – 6.Concordance with IGRA; – 7.Proportion started on TPT.

[1] > 100/100,000 population
[2] For estimation of specificity in the primary analysis, the ideal population is the one with a very low likelihood of prior exposure to *M. tuberculosis. Further, sensitivity analyses(see below) were conducted regardless of the background TB incidence*
[3] TB disease is used as a proxy diagnosis for TB infection

4. Methods

Inclusion criteria

All cross-sectional, case-control (using authors' definitions of case and control, which were further characterised at analyses) and longitudinal (prospective or retrospective) original research studies evaluating the index tests alone or with recognised comparator tests (QFT, T-SPOT, TST) in humans were reviewed, with no date or language restrictions. Only peer-reviewed journals were included. Detailed inclusion criteria by outcome, are presented in Table 2.

Index tests:

- C-TB (Serum Institute of India)
- Diaskin Test (Generium)
- C-TST (formally called ESAT-6 CFP-10 test, Anhui Zhifei Longcom)
- DPPD
- Others

Comparator tests:

- QFT-gold or plus (Qiagen)
- T-SPOT TB test (Oxford Immunotec)
- TST

Exclusions criteria

Exclusion criteria: Publication types excluded will be: 1) letters without original data; 2) case reports; 3) review articles; 4) abstracts 5) studies reporting insufficient data to determine diagnostic accuracy measures; 6) studies evaluating non-commercial TST or IGRA as comparator; 7) mathematical modelling or case-base studies; 8) animal studies.

Table 2: Inclusion criteria according to objective

Outcome	Study design	Inclusion criteria
1	Longitudinal studies that report index test result in the population eligible for LTBI testing, preventive therapy given and cases of incident TB during the study period	Must be free of active disease at baseline; Must report method of TB diagnosis (microbiological or clinical)
2	Longitudinal studies reporting development of incident TB in the population tested with index test during the study period	Must be free of active TB at baseline; Must report method of TB diagnosis (microbiological or clinical)
3	Studies reporting index test result in contacts of active TB cases	Must stratify contacts according to proximity to TB cases
4	Studies that report index test result in participants with confirmed active TB (sensitivity) or populations at low risk for TB (specificity)	Bacteriologically-confirmed TB. (see case definitions) The study must be performed in a low TB incidence setting to calculate specificity for the diagnosis of TB infection. Studies must include either healthy individuals or people with diseases other than TB.
5	Studies reporting results of comparator test alongside index test in any population	Must report comparator test and index test result and cut-off measurement used
6	Studies reporting the number of people who test positive and start TB preventive treatment	Must report the number of people who test positive and start TB preventive treatment in both an index test and a comparator test

Search strategy:

We updated the previous search that identified papers published until 20 Oct 2020. The systematic review protocol and search strategy were registered on (CRD42021274437) and followed PRISMA guidelines. The initial search was carried out in Medline and Embase for all studies published until 30 July 2021 with no language restrictions. In order to include as many studies as possible, the test manufacturers were contacted for additional studies. As Generium is a Russian company and most studies evaluating Diaskintest performance have been carried out in the ex-Soviet bloc, we searched e-library (www.e-library.ru) to look for additional Russian language studies. We looked for additional Chinese language studies on skin tests manufactured by Chinese manufacturers such as C-TST in the Chinese Biomedical Literature Database and the China National Knowledge Infrastructure database. Bibliographies of studies included in the review were hand-searched to identify additional relevant studies. We also reviewed studies that were identified through a public call for data by WHO (https://www.who.int/news-room/articles-detail/public-call-for-data-on-diagnostic-accuracy-of-newer-skin-based-tests-based-on-specific-m.-tuberculosis-antigens). The detailed search strategy and search terms are provided in Appendix 1.

Study Screening and data collection process:

Since we developed a broad search strategy for English papers encompassing multiple systematic reviews addressing other objectives, the initial list of English titles and abstracts were reviewed by two independent reviewers (YH and LEZ) to identify studies reporting any new skin tests regardless of the outcomes of interest. This was followed by a screening of titles and abstracts by two independent reviewers (YH and ES) and then a screening of full-text articles. Two Russian speakers (ES and IK) independently screened titles and abstracts identified from the e-library and then full-text articles as well as Ukrainian papers identified through the public call. Chinese abstracts and titles were screened by two reviewers independently using google translation to identify relevant studies. Full-text articles were reviewed by two Chinese speaking reviewers independently.

Discrepancies in inclusion/exclusion between the 2 reviewers were resolved by discussion between the 2 reviewers or if needed with additional reviewers. We used the systematic review management platform Rayyan[20] for study screening and tracking of exclusion reasons. Data extraction was carried out using specific data extraction sheet in Microsoft Excel.

Case definitions:

Incident TB disease: any new case of TB (new or relapse) diagnosed subsequent to initial symptoms and signs screening

Prevalent TB: any case of known TB disease at the time of the diagnostic test

Active TB: Hierarchy of reference standards:

1. Bacteriologically confirmed TB as per the WHO definition.
2. Clinical diagnosis based on presenting symptoms, radiology and / or response to TB treatment without microbiological confirmation

Data variables:

Table 3 details the principle variables of interest. Data will be mapped to a data extraction sheet. Although not all studies included all of these data, the minimum data for inclusion are stated in the inclusion criteria. Data extraction was done by two reviewers independently.

Table 3: Variables of interest

Category	Variables
Study design	Study design, country, setting, period of recruitment, sample size
Population summary measures	Age, gender, history of immunosuppression, HIV status, BCG vaccination history, TB contact history (method of diagnosis and Drug-susceptible test of case, proximity to case), migration history, homelessness, imprisonment, health working experience, CXR abnormalities history, smoking
Index test	TBST used, cut-off point used, cost
Comparator	IGRA assay and cut-off used, TST dose and cut-off used
Outcome	Intervention test results, comparator test results, preventive therapy given, numbers progressing to active TB and method of diagnosis

Quality assessment (risk of bias):

The quality of each included study was formally evaluated using a quality assessment tool appropriate to the study design. Studies were stratified by study design to explore the bias. For all diagnostic accuracy studies, study quality was assessed using a modified version of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[21] This assessed risk of bias and concerns regarding applicability in four domains: patient selection; index test; reference standard; and flow & timing. An additional domain pertaining to the involvement of commercial test manufacturers in study design, conduct or analysis and related risk of bias was added to the QUADAS-2 tool to assess the impact of possible conflicts of interest. For studies that were included in the review by Krutikov et al,[17] we adopted the results of the assessment made by the authors.

The GRADE framework[22] was used to systematically assess the quality of evidence and strength of recommendations regarding the use of novel TBST.

Data analysis:

Where possible, outcome measures were stratified by: type of test and population. Outcome and effect measures of interest were evaluated separately as described in Table 4.

Sensitivity in those with microbiologically-confirmed active TB and specificity in those at low risk of TB infection (restricted to studies from low TB burden countries) was calculated where possible. For specificity, we also applied less restrictive criteria and provided the results as a sensitivity analysis of the specificity estimates as explained in detail below.

Test agreement between the index test and each comparator test was calculated as the agreement proportion (total for negatives and positives), with 95% confidence intervals (CIs) (Clopper-Pearson exact CIs, ensuring valid values at proportions close to 1).

For outcomes with two or more studies with available data, meta-analyses were performed where appropriate. Examples are if studies used the same reference test, e.g. culture-confirmed TB, and/or in the same subpopulation, e.g. HIV+, and/or used the same test cut-off for positivity, e.g. TST15mm or 10mm. Univariate random-effects models were used for meta-analyses of agreement, sensitivity and specificity estimates. While bivariate models are usually recommended for pooling sensitivity and specificity, we did not find studies evaluating sensitivity and specificity in the same cohorts, precluding such analysis. Random effects models were used (as opposed to fixed effects) to account for heterogeneity of study populations. In addition to pooling agreement for each comparison of a new skin test vs TST or IGRA (e.g. two-way head-to-head), we performed three-way head-to-head comparisons by restricting to studies that compared a new skin test vs TST and IGRA.

Meta-analysis of sensitivity and specificity was explored in two ways: (1) including all studies available for each test; and (2) in head-to-head comparisons. Three-way head-to-head analyses permit simultaneous comparison of all three tests in the same population under the same study conditions tests and were prioritised in the report over indirect comparisons.

To assess heterogeneity, we analysed data stratified by TB status (microbiologically-confirmed TB, under investigation for TB, no TB), age (children [< 5 years or 18 years where available] vs. adults), HIV status, previous BCG vaccination, and other sub-groups as defined already. Where feasible, results were pooled within these strata, and statistical heterogeneity was assessed using the I^2 statistic.

To assess 'dose-response' association along a gradient of exposure, we compared the proportion of positive index tests (with 95%CIs) in each contact group according to proximity from a source case.

We also conducted the following sensitivity analyses:

1) Including IGRA indeterminate results in positive and negative groups
2) Combining HIV-positive and HIV-negative groups
3) Combining bacteriologically confirmed and clinically diagnosed TB.

Furthermore, as sensitivity analyses to assess the specificity estimates using less restrictive criteria, we examined the proportion of negative results as a proxy for specificity regardless of background TB incidence by estimating the following measures:

1) Differences in proportion of negative results (i.e specificity) between index tests and comparator tests
2) Proportion of negative results in participants with negative IGRA results.

For these analyses, we restricted the analyses to the following populations:

- Presumed "healthy" or described as asymptomatic and no reported history of contacts and not presumed to have active TB and not suspected of active TB (this was assumed unless specified otherwise)
- Other diseases without TB symptoms and not suspected of having active TB and no reported history of contacts (this was assumed unless specified otherwise)

The term specificity (proxy) is used when describing the estimates above.

We did not conduct a test for publication bias because none of the quantitative syntheses included sufficient numbers of studies (≥ 10 studies).

Our primary analysis did not pool different TBST because of substantial heterogeneity in study designs. However, based on a request from WHO as the purpose of the review is for a class-based recommendation, as posthoc analysis, we combined data across different TBST. It required us to ignore heterogeneity in various aspects. As noted in the text, some of our analyses did not pool data even within the same test because of the heterogeneity in study designs (e.g. populations) and this was ignored. We also ignored the differences in TST cut-off values. Most Diaskintest studies were not primarily designed to evaluate diagnostic accuracy subject to high risk of bias, but we pooled regardless.

Table 4: Effect measures according to objective

Outcome	Effect measure
1. Efficacy of preventive therapy based on the test result	**Incidence rates** for disease progression stratified by test result **Incidence Rate Ratios** **Negative and Positive Predictive Values** for disease progression
2. Predictive value of novel recombinant skin tests for incident TB among risk-stratified populations	**Incidence rates** for disease progression in risk-stratified populations **Incidence Rate Ratios** **Negative and Positive Predictive Values** for progression with confidence intervals
3. The association between the test result and proximity of exposure among TB case contacts	**Odds Ratio** according to contact proximity for each study **Concordance** and **discordance** between index and comparator test result according to proximity among contacts
4. Sensitivity and specificity	**Sensitivity** = proportion of people with positive skin test among those with microbiologically confirmed TB (groups 1 and 2 in case definitions) **Specificity** = proportion of people with negative skin test among those at low risk for TB infection (primary analysis). We also examined proportion of people with negative skin test among those at risk for TB infection (alternate criteria).
5. Concordance and discordance of index test with comparators when using crude and BCG-stratified TST measurements	**% Concordance/Discordance,** total and by test pairs. **Concordance:** will be defined by summary comparison in proportion test positivity between index and comparator **Discordance:** will be measured by difference in proportions of test negativity for index and comparator test.
6. Proportion of participants who test positive by new skin tests and start TB preventive treatment	**Number of participants who test positive and start TB preventive treatment /Number of participants who test.** **Difference in proportions between an index test vs a comparator test**

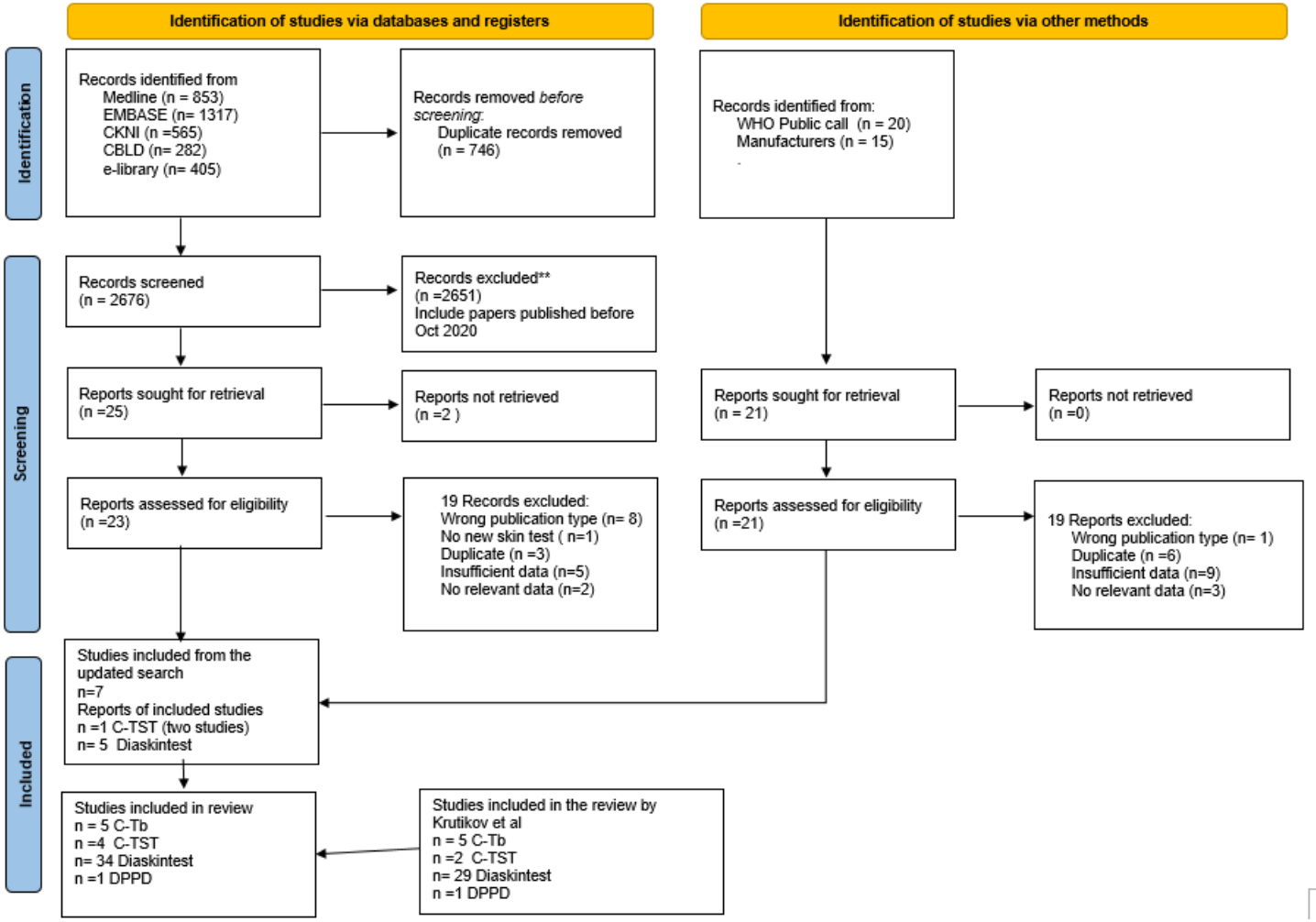
5. Results

5.1 Systematic literature review

Figure 2 presents the selection process. Our previous review identified 37 studies reporting TBST: 29 Diaskintest, five C-Tb, two C-TST, one DPPD, that were published from inception until 20 October 2020.[17]

Our updated search covering the period until 30 July 2021 identified four reports (three in Russian[23-25] on Diaskintest and one in English on C-TST[26]) via the database search. The report on C-TST reported the results of two studies. Additional two studies (one Russian[27] and one Ukrainian[28]) on Diaskintest were identified via other methods. By combining these with studies included in the previous review,[17] the total numbers of studies for each test were: five for C-Tb,[13 15 16 29 30] 34 for Diaskintest,[23-25 27 28 31-60] four for C-TST,[26 61 62] and one for DPPD.[63] The DPPD test is still undergoing evaluation and is not ready for commercialisation. DPPD test data is not included in the pooled analysis but is presented for completeness.

Figure 2 Study selection



5.2 Characteristics of included studies

All but one Diaskintest study was conducted in Russia (Table 5). All were cross-sectional assessments performed as a part of a routine care provision, and cohorts were recruited prospectively or constructed retrospectively. None of the studies randomised different tests into different groups or arms. Four studies[43 44 57 60] were head-to-head comparisons of Diaskintest with TST 5 mm cut-off (TST^{5mm}) and IGRA in the same study. Four studies[34 51 53 55] enrolled a total of 346 (4·9%) adults and 23 (0·3%) children with HIV, and 17 studies recruited children younger than 18 years (Table 5).[25 27 31-33 35-39 43 44 47 52 54-56] Approximately half of the individuals with HIV had a CD4 count lower than 200 cells per µL.[51 53] The proportion of participants who had received a BCG vaccination was reported in five studies;[25 34 38 43 44] the proportion ranged from 93–100%. Diaskintest threshold for positivity varied and included any skin induration ($Diaskintest^{AI}$) according to national guidance[45] or 5 mm ($Diaskintest^{5\ mm}$) or 7 mm ($Diaskintest^{7\ mm}$) as chosen by investigators. Diaskintest studies used PPD-L, a purified protein derivative developed in Russia that has previously been shown as bioequivalent to PPD-RT23 used in the non-Diaskintest studies.[64] A TST reading that was larger than 5 mm was considered positive, a reading between 1 mm and 4 mm was classed as indeterminate.[45] The categorisation of thresholds used for TST in these studies is different from how other studies have handled TST results where a binary classification of positive/negative was applied without consideration of indeterminate results. IGRA used included the T.SPOT.TB and QuantiFERON-TB Gold (QFT) tests.

All five studies evaluating C-Tb were designed as prospective clinical trials to evaluate diagnostic accuracy; 3 in South Africa,[16 29 30] 1 in Spain[15] and 1 in the United Kingdom.[13] Populations tested were predominantly adults; active TB, HIV infected individuals and children (Table 6). Two studies included individuals at low risk for TB infection: university students and staff without a history of TB exposure or TB signs and symptoms in Spain[15] and healthy adult volunteers in UK without a history of TB exposure who had negative QFT results.[13] All five conducted three-test head-to-head comparisons within the same tested cohort. C-Tb and TST were administered randomly to different arms, and the allocation was blinded. In all five studies, the threshold for positivity was stratified depending on the sub-population tested; TST^{5mm} for HIV+ and TST^{15mm} for BCG vaccinated populations, reported aggregated (shown as $TST^{5mm/15mm}$ cut-off) or disaggregated. By contrast, the manufacturer-recommended 5mm threshold for C-Tb positivity was consistently used. All included QFT IGRA as comparators.

Four studies conducted in China provided data for assessment of C-TST.[26 61 62] All studies were conducted as clinical trials performed in China to evaluate diagnostic accuracy. Two studies included individuals with active TB,[61 62] one with active TB and other pulmonary diseases,[26] and one healthy individual study (44% with BCG scar) (Table 6).[61] No studies included people living with HIV or children. One study evaluated the agreement between C-TST and TST, and both were given to the same individuals without randomisation or blinding of the allocation. C-TST was applied with different cut-offs by study, including induration ≥5mm, induration or redness ≥5mm, and erythema ≥5mm with induration or redness ≥5mm adopted in the package insert. In one study providing data on the comparison between C-TST and TST, a TST cut-off of both 5mm and 10mm were used.[26]

A study in Brazil[63] (n=173) assessed DPPD performance (5mm for HIV+ and 10mm for HIV-) *vs* the TST^{5mm} in HIV+ and TST^{10mm} in HIV-uninfected individuals with microbiologically-confirmed TB, and in healthy individuals (Table 6), all of whom were BCG-vaccinated.

We did not identify studies that followed up participants for risk of incident TB, evaluated the effectiveness of preventive treatment, or the proportion of participants with positive results starting TB preventive treatment. Table S3 in supplement summarises studies available for assessment of each review objective

Table 7 shows the cut-off defined for each test by test manufacturers. It should be noted, however, that they were not always followed, and the cut-off used in each study is indicated in the text and presented in supplementary tables.

Table 5 Characteristics of Diaskintest studies

Study	Country	Index Test	Comparators	Age, years (a)	Sample size (in review)	Study population (b)	Review Objective Addressed			
							Test concordance	Sensitivity	Specificity (c)	Dose-response association (d)
Aksenova 2011	Russia	Diaskintest^{AI}	TST^{5mm}	NS	1551(63)	Children; TB Screening	X	X		
Baryshnikova 2017	Russia	Diaskintest^{AI}	TST^{5mm}	NS	811(163)	Children; active PTB	X			
Baryshnikova 2021	Russia	Diaskintest^{7mm}	IGRA (TSPOT.TB)	18-65	4756 (645)	Children under routine surveillance in TB service	X			
Borodulina 2012	Russia	Diaskintest^{AI}	TST^{5mm}	28	274 (100)	HIV- adults, active Tb	X	X		
Borodulina 2014	Russia	Diaskintest^{AI}	None	NS	185 (12)	Children with and without TB		X		
Dotsenko 2015	Ukraine	Diaskintest$^{AI/5mm}$	TST^{5mm}	≥20	25 (25)	TB care workers	X			
Dovgalyuk 2013	Russia	Diaskintest^{AI}	TST^{5mm}	4.2	570 (570)	Children; TB Screening	X			
Fedorovykh 2014	Russia	Diaskintest^{5mm}	None	NS	551(83)	Children, household TB contacts				X
Kabanets 2016	Russia	Diaskintest^{AI}	TST^{5mm}	NS	1204 (1204)	Children, TB Screening	X			
Kibrik 2015	Russia	Diaskintest^{AI}	None	NS	2373 (1060)	Medical students; TB contacts; active TB; non-TB disease				X
Koretskaya 2012	Russia	Diaskintest^{5mm}	TST^{5mm}	23	109 (109)	Medical students	X			
Laushkina 2017	Russia	Diaskintest^{5mm}	None	42.9	70 (20)	Adults; TB investigation		X		

Losovskaya 2014*	Russia	Diaskintest[AI]	TST[5mm], IGRA (QFT-TB GIT)	0.5-15	50 (46)	Children; TB investigations	X			
Losovskaya 2016*	Russia	Diaskintest[AI]	TST[5mm], IGRA (QFT-TB GIT)	3-6	63 (63)	Children; TB investigation	X			
Mishin 2016	Russia	Diaskintest[AI]	None	NS	529 (103)	HIV- Adults; PTB; Healthy control;		X		
Nakonechnaya 2020	Russia	Diaskintest[AI]	IGRA (QFT-TB GIT)	1-17	62 (62)	Children with TB and other pulmonary conditions	X			
Nikitina 2019	Russia	Diaskintest[AI]	IGRA (QFT-TB GIT)	Adults: 42, 18-84 Children: 10, 3-16	181 (68)	Adults and children; TB investigation		X		
Salina 2011	Russia	Diaskintest[5mm]	TST[5mm]	17-80	142 (33)	Adult; TB investigation	X	X		
Salina 2019	Russia	Diaskintest[AI]	None	18-68	69 (69)	Active PTB		X		
Samorodov 2019	Russia	Diaskintest[AI]	None	37.1	336 (336)	Adults; respiratory illness (undetermined)		X		
Senin 2016	Russia	Diaskintest[5mm]	None	30-39	207 (124)	HIV+ adults (CD4 < 200 cells/mm^3 in 45%); active TB		X		
Shovkun 2014	Russia	Diaskintest[AI]	TST[5mm]	NS	220 (220)	Children; TB investigation	X			
Slogotskaya 2011 a	Russia	Diaskintest[AI]	None	31.5	88 (88)	HIV+ adults (CD4 < 200 cells/mm^3 in 46.6%); active TB	X	X		
Slogotskaya 2011b	Russia	Diaskintest[AI]	TST[5mm]	NS	1677 (23)	Children and adults; TB investigation	X			
Slogotskaya 2012	Russia	Diaskintest[AI]	IGRA (QFT-TB GIT)	12	122 (122)	Children, active PTB; Children, PPD-TST+/Diaskintest+	X			
Slogotskaya 2013	Russia	Diaskintest[AI]	TST[5mm]	7-14	521 (511)	Children, active TB	X			
Slogotskaya 2018	Russia	Diaskintest[5mm]	TST[5mm]	8.8	441(408)	Children; active TB		X		

Starshinova 2018*	Russia	Diaskintest^{5mm}	TST^{5mm}, IGRA (T.SPOT-TB, QFT-TB GIT)	Children 8.1, Adults: 37	860 (860)	HIV- ; BCG-vaccinated; TB Screening; Children; Adults		X		
Starshinova 2019a*	Russia	Diaskintest^{5mm}	TST^{5mm} , IGRA (QFT-TB GIT, T-SPOT.TB)	18-65	187 (135)	Adults, Culture+ TB; TB unexposed; IGRA+/Diaskintest+		X		
Starshinova 2019b	Russia	Diaskintest^{AI}	None	TB hospital: 42 (0.23), General hospital: 43 (0.27)	154 (154)	Healthcare professionals in TB hospitals and general hospitals				X
Stogova 2020a	Russia	Diaskintest^{AI}	TST^{5mm}	48.3	328 (328)	Adults with suspected TB and other pulmonary conditions	X	X		
Stogova 2020b	Russia	Diaskintest^{AI}	TST^{5mm}	NS	453 (296)	Adults with suspected TB, other pulmonary conditions and healthy subjects	X	X		
Vaganova 2015	Russia	Diaskintest^{AI}	None	NS	321 (321)	Medical doctors and nurses working in TB dispensaries				X
Yablonskiy 2013	Russia	Diaskintest^{5mm}	TST^{5mm}	Age 3-6: 4.5, Age 7-14: 12.3	120 (43)	Children; TB investigation	X			

* Studies included in three-way head-to-head analysis (index test compared with both IGRA and TST)

(a) Age: Average, either mean age (standard deviation) or median and/or range;

(b) Study population: Where HIV status not indicated=not specifified/unknown (explored in sensitivity analysis).

(c) Specificity could not be estimated in diaskintest studies (TB not ruled out; studies conducted in a high-burden country).

(d) Dose-response association: Studies evaluating index test performance amongst TB contacts of varying degrees of exposure.

AI (any skin induration); TST: Tuberculin skin test; IGRA: Interferon gamma release assay; QFT-TB GIT: QuantiFERON TB Gold In-Tube. PTB: pulmonary TB; TB Screening: Individuals undergoing routine TB screening; TB Investigation: Individuals with suspected TB undergoing investigation.

NS: not specified

Table 6 Characteristics of C-Tb, C-TST, and DPPD studies

Study	Country	Index Test	Comparators	Age, years (a)	Sample size (in review)	Study population (b)	Review Objective Addressed			
							Test concordance	Sensitivity	Specificity	Dose-response association (c)
Aggerbeck 2013	United Kingdom	C-Tb	TST (multiple thresholds), QFT	Cases: 33, 18-60, Controls:34, 18-65	189 (189)	Active Tb (3 participants selected on the basis of positive IGRA); TB unexposed adults	X		X	
Aggerbeck 2018	South Africa	C-Tb	$TST^{5mm/15mm,\ QFT}$	17, 0-65	1190 (1190)	Child case-contacts under 5 years and healthy controls; HIV+ (median CD4+ 314 cells/microlitre (IQR 164-502) and HIV- adults suspected of TB; Active TB	X			
Aggerbeck 2019	South Africa	C-Tb	$TST^{5mm/15mm}$, QFT	35; 18-64	456 (154)	Adults, active TB	X	X		
Hoff 2016	South Africa	C-Tb	TST (multiple thresholds), QFT	34; 18-64	253 (241)	HIV+ and HIV- adults with active TB	X	X		
Ruhwald 2017	Spain	C-Tb	$TST^{5mm/15mm,\ QFT}$	Controls: 24.1, Cases: 37.3, Close contacts: 32.9, Occasional: 31.5	979 (970)	Close TB contacts; occasional TB contacts; Active TB; TB-unexposed	X	X	X	X
Li 2016	China	C-TST	TST^{5mm}, T-SPOT.TB	Controls: 45, Cases: 41.3	144 (144)	TB unexposed; Active TB		X		
Xu 2021a	China	C-TST	TST (multiple thresholds), T-SPOT	TB:38.8 Non-TB: 51	192 (95)	Active TB and patients with other pulmonary diseases	X	X		
Xu 2021b	China	C-TST	TST (multiple thresholds), T-SPOT	46.3	777 (396)	Healthy adults with normal chest X-ray results and no tuberculosis history			X	
Zhang 2020	China	C-TST	None	18.77 (13.11); 18-65	2257 (743)	Active TB		X		
Badaro 2020, Brazil	Brazil	$DPPD^{5mm/10mm}$	$TST^{5mm/10mm}$	HIV+: 31.2; 18-54, HIV-: 39.9; 19-64 Healthy: 29.8; 18-47	173 (173)	Active TB; HIV+ adults (6/38 (15.8%) had CD4 < 200 cells/mm^3); HIV – adults; healthy volunteers	X	X		

(a) Age: Average, either mean age (standard deviation) or median and/or range;

(b) Study population: Where HIV status not indicated=not specified/unknown (explored in sensitivity analysis).

(c) Dose-response association: Studies evaluating index test performance amongst TB contacts of varying degrees of exposure.

AI (any skin induration); TST: Tuberculin skin test; IGRA: Interferon-gamma release assay; QFT-TB GIT: QuantiFERON TB Gold In-Tube. PTB: pulmonary TB; TB Screening: Individuals undergoing routine TB screening; TB Investigation: Individuals with suspected TB undergoing investigation.

NS: not specified

Table 7 Cut off of TBST defined by their manufacturers

Test	Source	Cut off
Diaskintest	Package insert	Negative response: The absence of infiltration and hyperaemia or the presence of 'prick response' up to 2 mm
		Ambiguous response: The presence of hyperaemia without infiltrate.
		Positive response: The presence of infiltrate (papule) of any size
C-Tb	The most recent study by Aggereck et al. (The test is not yet commercialized. No package insert is available)	Induration ≥ 5mm
C-TST	Package insert	A positive result is interpreted by an average diameter of redness or induration (sum of transverse and longitudinal diameters divided by 2) no less than 5mm.
DPPD	The most recent study by Badaro et al (The test is not yet commercialized. No package insert is available)	Induration ≥ 10 mm for healthy individuals and ≥ 5mm for people living with HIV

5.3 Risk of bias in individual studies

The quality of studies evaluating Diaskintest performance was difficult to assess due to inconsistencies and incomplete reporting of study methods and sample recruitment. This resulted in several "unknown" assessments against the quality criteria. Of the 16 studies evaluating the sensitivity of Diaskintest, risk of bias was high in 5 (31.3%) studies where test assessors were not blinded to TB culture results,[34 46 49 51 53] and unclear in at least one of the four risk of bias criteria in 14 (87.5%) studies as information on patient selection or blinding was not presented.[23 24 31 34 35 42 46 48-51 53 57 65] Of those evaluating Diaskintest concordance, 14/18 (84.6%) had high risk of bias in the reference standard criterion because assessors of reference standard (TST) were not blinded to index test results and/or the use of TST as a reference test.[23 24 28 32 34 36 38 39 41 43 44 52 55 56] For the index test criterion, two had high risk of bias as index test assessors were not blinded to reference standard results[28 34] and 15 (88.9%) were classed as unclear as this information was not provided.[23-25 27 31 32 36 38 39 41 43 44 52 54-56] Of all Diaskintest studies, patient selection bias was unclear for 22 out of 34 (73.5%) studies as reporting of patient selection was incomplete[25 27 31 34 36-38 40 41 43 44 46 48 49 51-55 57-59] and one had high risk of bias.[28]

One C-Tb study scored high on the risk of bias criterion because not all participants received the same reference standard (IGRA or TST).[15] Four out of five (80.0%) C-Tb studies[13 16 29 30] and three C-TST-skintest studies[26 62] had a conflict of interest concerns, as studies either did not report disclosures or were directly affiliated with the test manufacturer. In addition, for two C-TST studies,[61 62], it was unclear whether the patient selection was random or consecutive. Applicability concerns and risk of bias were low for the DPPD study.[63] (See Table S45 in supplement for QUADAS-2 results).

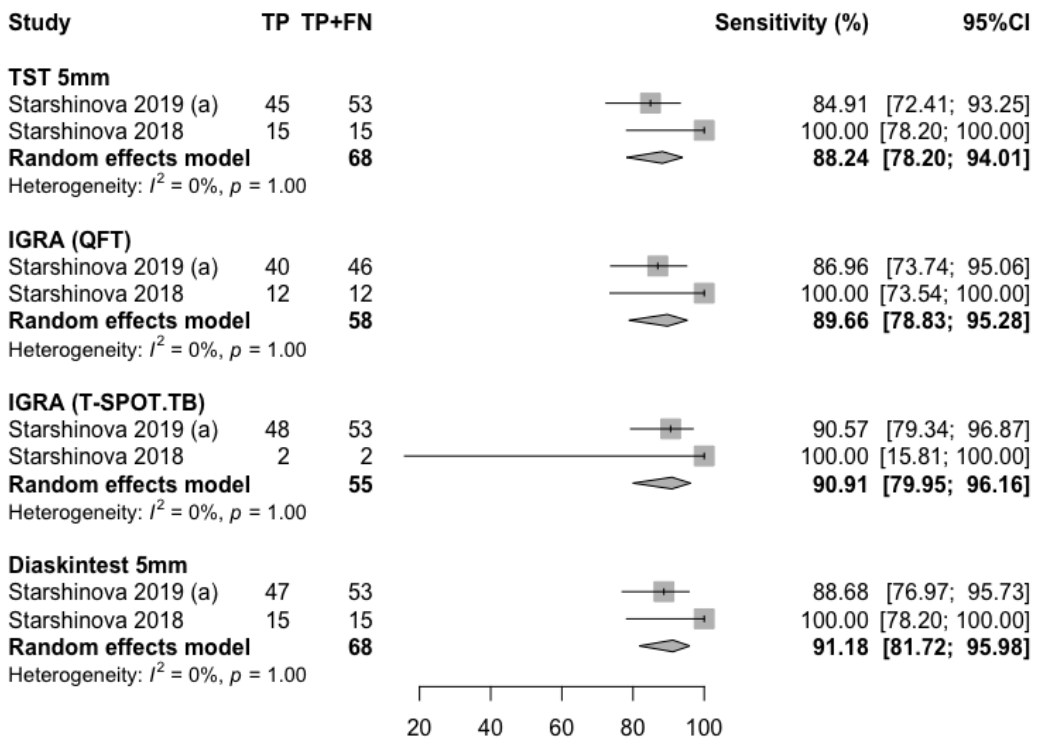
5.4 Diaskintest

Based on two studies, pooled sensitivity for Diaskintest$_{5\,mm}$ was 91.18% (95% CI 81.72–95.98), 88.24% (78.20–94.01) for TST$_{5\,mm}$, 89.66% (78.83–95.28) for QFT, and 90.91% (79.95–96.16) for TSPOT.TB (Figure 3).[57 60]

Two studies provided data on head-to-head comparisons of Diaksintest with IGRA and TST in agreement.[43 44] Both included children who did not have HIV but were under investigation for tuberculosis or with clinically diagnosed tuberculosis. The pooled test agreement of Diaskintest$_{AI}$ with IGRA was 87.16% (95% CI 79.47–92.24), considerably higher than the agreement between TST$_{5\,mm}$ and IGRA (51.38% [42.05–60.60]) and the agreement between Diaskintest$_{AI}$ and TST$_{5\,mm}$ (55.45% [46.08–64.45]; figure 4).[43 44]

Considering all studies with at least two-way test comparisons, pooled agreement of Diaskintest$_{AI}$ with IGRA was 94.62% (95% CI 90.49–97.02; I^2 = 56.2%) in five studies in participants with any tuberculosis status (Table S3).[25 43 44 54 60] By contrast, the agreement between Diaskintest$_{AI}$ and TST$_{5\,mm}$ showed considerable heterogeneity; the pooled agreement was estimated in children with active tuberculosis (97.39% [96.39–98.12])[32 34 35 44 55 56 60] and in children without TB (17.62% [6.79-38.60] (Table S6).[25 44] The heterogeneity may be due to the impact of repeated BCG vaccination.[66] Agreement between Diaskintest$_{5\,mm}$ and TST$_{5\,mm}$ is reported in Table S7. For people who did not have HIV, pooled estimates of Diaskintest sensitivity was 66% (95%CI 59%-73%) for Diaskintest$_{AI}$ and 88% (78–94) for Diaskintest$_{5\,mm}$.[23 24 34 42 46-50 57]

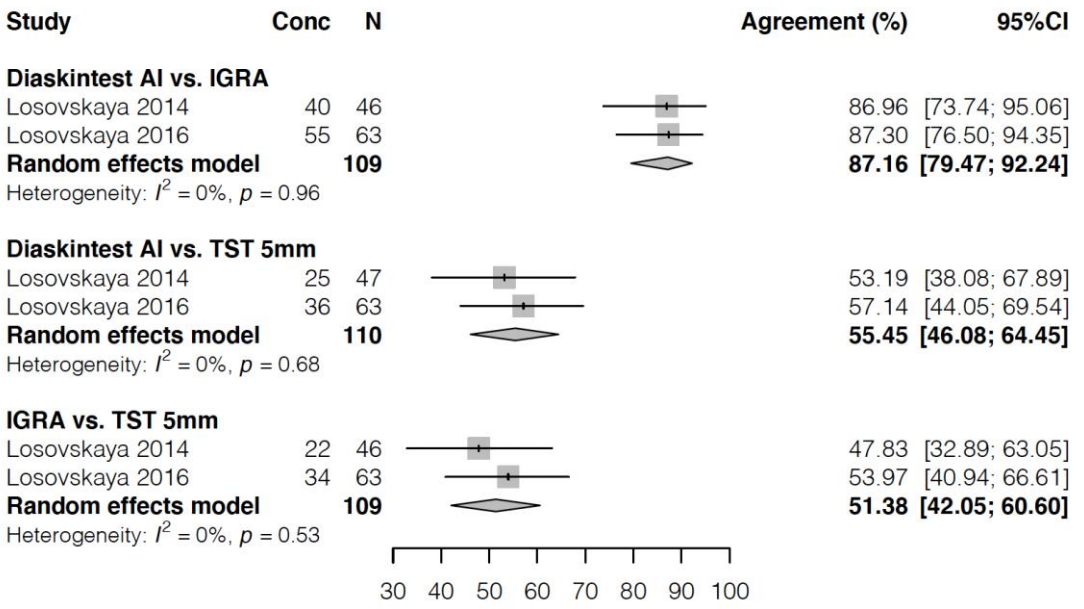
Figure 3. Test sensitivity in three-way head-to-head studies comparing Diaskintest, IGRA and TST



TP = True positive FN = False negative

Includes HIV-uninfected adults with microbiologically-confirmed active TB.

Figure 4 Test agreement in head-to-head Diaskintest studies comparing all three tests



Includes HIV-uninfected children under investigation for TB and those with active TB (clinical and confirmed). AI = Any induration ; Conc = Concordant; N = Total, concordant + discordant; % Agreement: represents agreement with IGRA as the comparator

Highly variable methods and sub-populations precluded meaningful meta-analysis for most risk groups; sensitivity estimates from individual studies ranged from 40%-71% in HIV-infected adults[51 53] and from 92% to 100% in HIV-uninfected children,[31 33 35] (Supplement Table S4-S12).

In the study that stratified results by age,[36] in children under 5 years of age (N=570), the proportion Diaskintest^{AI} positive in a cohort of children undergoing TB screening was 2.5% and test agreement with TST^{5}mm was 76.4% (72.7–80.1%). The study did not compare Diaskintest against IGRA.

Specificity was not estimated for Diaskintest as TB infection had not been excluded in enrolled populations, and studies were conducted in a high-burden setting. Proportion test positive appeared to vary by exposure gradient and was higher in contacts proximal to a source case (Table S13).[37 40 58 59]

In the sensitivity analysis using less restrictive criteria for specificity, the differences in specificity between Diaskintest^{5mm} and QFT ranged from -1.9 to 10.6% with the pooled difference of 4.5% (95%CI -13.1- 22.1) (Table S14).[60 67] The differences were substantially larger between Diaskintest^{5mm} and TST^{5mm} (Table S15). The specificity of Diaskintest results among QFT-negative individuals in one study was 99.1% (95%CI 94.9-100.0) (Table S16).[60]

5.5 C-Tb

In four head-to-head studies[15 16 29 30] (Figure 5), pooled sensitivity for C-Tb was 74.52% (95% CI 70.39–78.25), similar to that for $TST_{15\,mm}$ (77.18% [66.44–85.25]) and for the aggregated $TST_{5\,mm/15\,mm}$ threshold (78.18% [67.75–85.94]). In the same four studies, sensitivity for $TST_{5\,mm}$ was 82.68% (95% CI 74.91–88.42) and 71.67% (63.44–78.68) for IGRA; however, the 95% CIs overlapped. The sensitivity of C-Tb was lowest at 61% in a study in Spain[15] while it ranged from 73% to 85% in the other three studies in South Africa.[16 29 30] Evaluation of specificity was possible in two studies that evaluated all three tests in low-burden settings (Figure 7).[13 16] Pooled specificity estimates for C-Tb (98%, 95% CI 94-99%) and IGRA (99%, 95% CI 80-100%) were similarly high, but slightly lower for TST^{15mm} (93%, 95% CI 90-95%); the analysis was not possible for TST^{5mm} due to insufficient data.

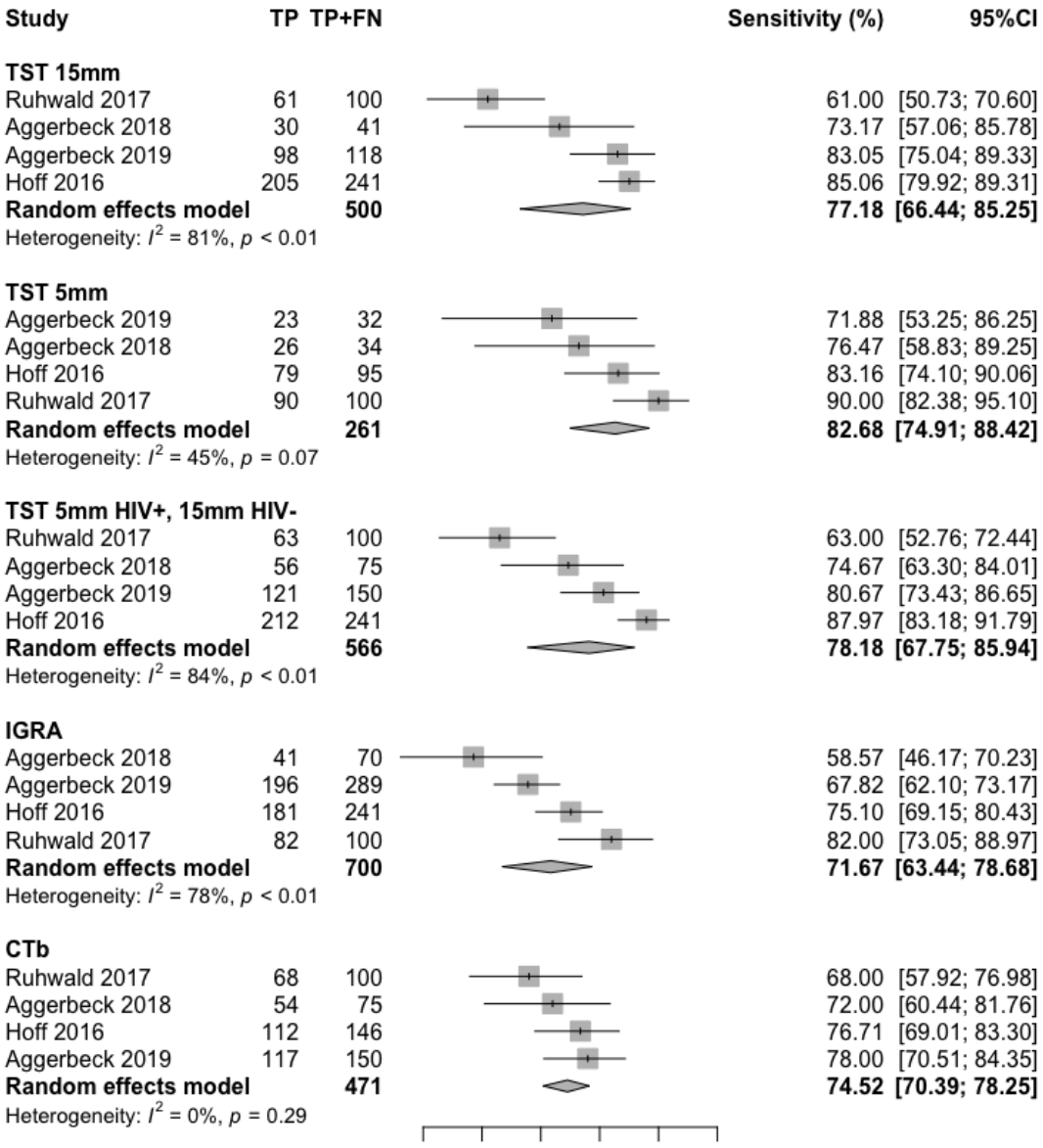
Three studies provided suitable head-to-head data for agreement comparisons between C-Tb, IGRA, and TST. Pooled test agreement between C-Tb and IGRA was 79.80% (95% CI 76.10–83.07), similar to that between IGRA and $TST_{5\,mm/15\,mm}$ (74.67% [64.01–83.01]) and C-Tb and $TST_{5\,mm/15\,mm}$ (78·92% [74·65–82·63]; figure 7).[16 29 30]

C-Tb results from studies that only compared two tests are shown in supplement section 4 (Tables S21-S26, Figure S2, Figure S3). These showed a pooled agreement of C-Tb with TST to be similar, 81% (95% CI, 76-85%) at TST^{5mm} in HIV-infected and 76% (95% CI, 71-81%) at TST^{15mm} in HIV-uninfected (Table S15).[15 16 29 30] Test agreement among individuals without TB was reported in two studies. In one study,[15] C-Tb and IGRA agreement ranged from 92% to 97% across sub-populations with different levels of TB exposure, while it was 78% and 81% in HIV-infected and uninfected individuals, respectively, in the second study.[13] Agreement between C-Tb and the TST^{5mm} in these two studies was 83% and 87%, respectively (Table S21). A dose-response association between C-Tb test positivity and proximity to a source-case was demonstrated. (Figure S3).[15]

In the sensitivity analysis, the differences in specificity between C-Tb and IGRA ranged from -0.7% to 7.7%, which was highest in a study in South Africa (Table S28).[29] The pooled difference was 0.7% (95%CI -7.0, 8.4). The differences in specificity between C-Tb and TST were generally higher than those between C-Tb and IGRA (Table S29).

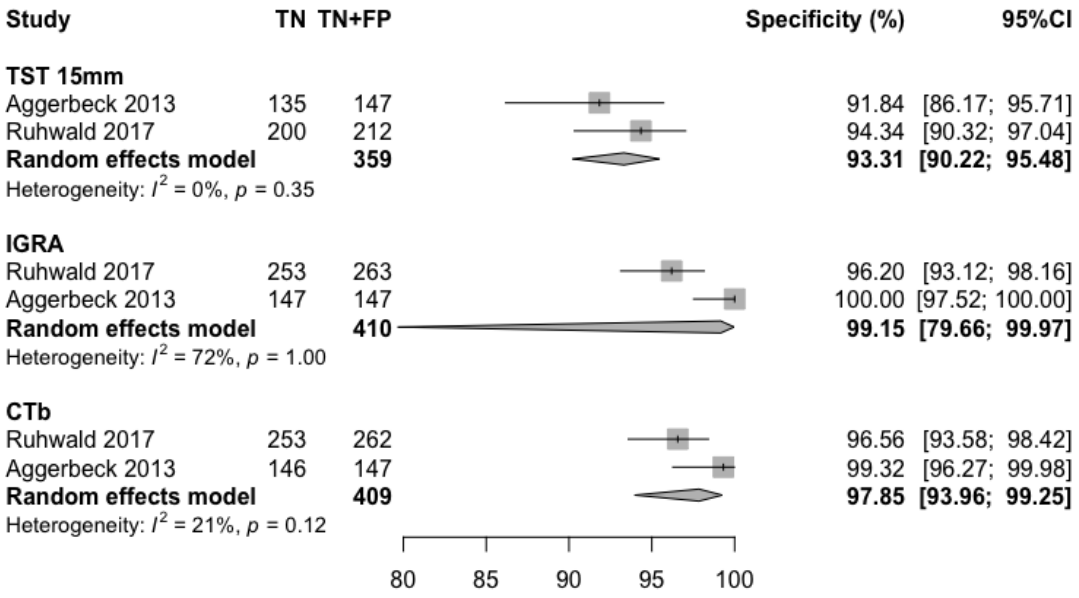
The specificity of negative C –Tb results in IGRA-negative participants was 91% in one study and 99% in two studies, respectively (Table S30).[13 15 29]

Figure 5. Test sensitivity in head-to-head studies comparing C-Tb, IGRA and TST



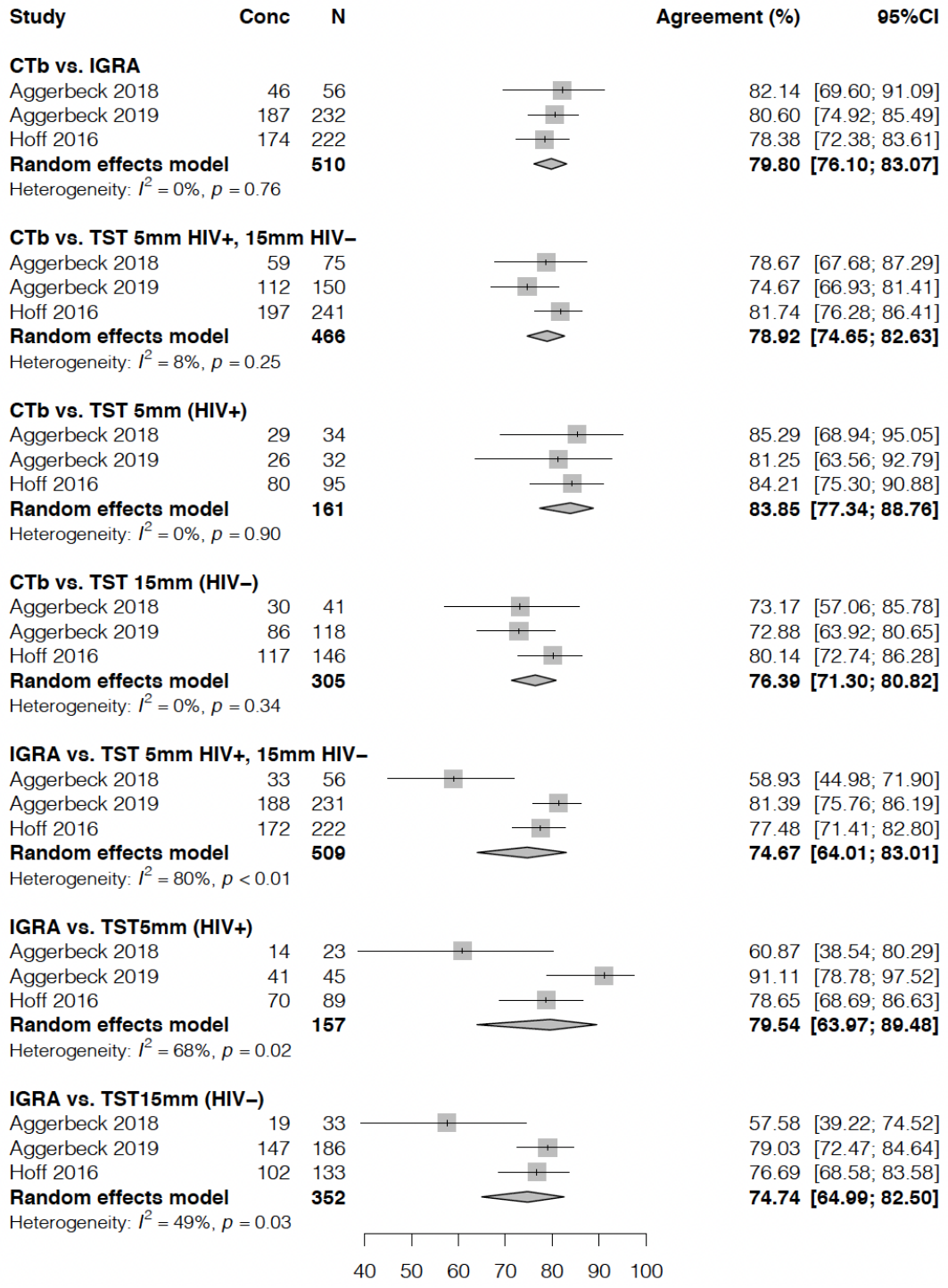
TP = True positive; FN = False negative. Results include individuals with microbiologically-confirmed active TB.

Figure 6. Test specificity in head-to-head studies comparing C-Tb, IGRA and TST



Individuals without active TB in studies conducted in TB low-incidence settings. TN = True negative FP = False positive

Figure 7 Test agreement in head-to-head C-Tb studies comparing all three tests



Includes individuals with bacteriologically-confirmed active TB. Ruhwald 2017 and Aggerbeck 2013, although did three- test comparisons, did not report data suitable for estimation of %TST-IGRA agreement. Conc = Concordant; N = Total, concordant + discordant; % Agreement: represents agreement with IGRA as the comparator

5.6 C-TST

Three studies evaluated sensitivity of the C-TST.[26 61 62] Sensitivity at the ≥5mm induration threshold ranged from 77% to 90%, with a pooled estimate of 86% (95%CI: 83-89%) (Table S34). In one study, the sensitivity of C-TST (90% [73-98%]) was similar to that of TPOT.TB (89% [78-95%] and TST $^{10\,mm}$ (87% [76-94%] and slightly higher than TST $^{10\,mm}$ (82% [69-90%].[26]

In two studies including healthy individuals as well as TB patients and patients with other pulmonary diseases,[26] the pooled agreement between C-TST and IGRA was higher (85.96% [78.82-90.97%]) than that between C-TST and TST $^{5\,mm}$ (68.23% [55.48-78.74]) and TST $10^{\,mm}$ (71.28% [67.12-75.11]). Specificity was not estimated.

Xu et al.[26] compared the results of C-TST, TST, and TSPOT.TB in healthy adults 12 weeks after BCG vaccination. The agreement between C-TST and TSPOT.TB was 97%, while the agreement between C-TST and TST was substantially lower (7% for TST^{5mm} and 27% for TST^{10mm}) because of the impact of BCG vaccination.

Similar to other tests, the differences in specificity between C-TST and TST were higher than those between C-TST and IGRA.[26] The differences were 39.9% for TST $^{5\,mm}$ (95%CI 33.8-45.6) 24.5% (95%CI 18.6-30.2) for TST $^{10\,mm}$ and 3.5% (-1.4-8.4%) for TSPOT.TB (Table S36-S37). The specificity of C-TST results in IGRA-negative participants was 95% (93-97%) in one study in China (Table S38).[26]

5.7 DPPD

Sensitivity was 89% in HIV-infected and 100% in HIV-uninfected compared to 50% and 100%, respectively for TST $^{5\ mm}$.[63] Test specificity was not estimated. Full results are presented in supplementary (Table S40-S41). For DPPD, agreement with the TST in active TB was 60% in HIV-infected individuals.[63] In HIV-uninfected individuals, agreement was 100% in active TB and 56% in healthy BCG-vaccinated controls. The difference in specificity for DPPD was only available against TST (44.2 [33.3-53.6]) (Table S42).

5.8 Sensitivity analyses

We conducted sensitivity analyses which included: (1) classification of indeterminate Diaskintest results first into the positive results group and then into the negative results group for test agreement and test sensitivity objectives; (2) inclusion of clinical diagnosis of TB instead of only microbiologically-confirmed cases (from studies already included in data synthesis that report test performance in microbiologically-confirmed as well as clinically-diagnosed cases (3) inclusion of groups with 'unknown' HIV status in the HIV- and HIV+ groups separately, to create composite groups for test agreement and sensitivity objectives for C-Tb. Results did not vary considerably and did not alter conclusions (Tables S17-S20, S36-S39, S43-S44). Sensitivity analyses regarding specificity are already presented in the text above.

6. Post-hoc meta-analysis of different TBST

When combining all studies on Diaskintest, C-Tb, C-TST, and DPPD, the pooled sensitivity was 76% (95%CI: 70-81%, 17 studies) in individuals with HIV-negative or unknown status and 63% (95%CI: 53-73%, 5 studies) in HIV-positive individuals.

While this may be seen as the lower sensitivity in HIV-positive individuals, these estimates are based on different studies, and thus not conclusive.

When combining all TBST, the pooled difference in specificity between TBST and IGRA was 2.29% (95%CI: -1.60-6.18%, 6 studies on Diaskintest, C-Tb and C-TST) and that between TBST and TST was 33.47% (95%CI: 18.16-48.78%, 14 studies including Diaskintest, C-Tb, and C-TST). The difference in agreement between TBST vs IGRA and TBST vs TST is consistent with 3-way head-to-head analysis of agreement (i.e. lower agreement of TBST with TST than with IGRA). Again, the indirect comparison might have affected the difference, while the difference is most likely explained by the impact of BCG. Similarly, the pooled agreement of TBST with IGRA was 89% (95%CI: 83-93%, 8 studies) in people without TB and 86% (95%CI: 80-90%, 8 studies) in people with TB. The agreement with TST was 59% (95%CI: 45-72%, 16 studies) in people without TB and 88% (95%CI: 82-93%, 13 studies) in people with TB.

Given the caveats explained in the method section, these estimates need to be interpreted with caution since meta-analysis is not recommended when heterogeneity can be explained. As the Cochrane handbook states, “The confidence interval from a random-effects meta-analysis describes uncertainty in the location of the mean of systematically different effects in the different studies”, and thus confidence intervals do not describe the degree of heterogeneity among studies”. “When there are many studies in a meta-analysis we may obtain a very tight confidence interval around the random-effects estimate of the mean effect even when there is a large amount of heterogeneity.”[68] This applies to the present analysis.

7. Interpretation

Our review showed that the sensitivity of TBST is similar to existing tests, including TST and IGRA. Based on specificity estimates in C-Tb studies conducted in populations with a low risk for TB infection and estimates derived using alternate less-stringent criteria, the specificity of TBST also appears similar to IGRA. Two studies on C-Tb also showed that the sensitivity of TBST is similar to TST^{15mm}.

On the other hand, the specificity of TBST was substantially higher than TST in studies from China and Russia. It is most likely due to the impact of BCG vaccination, especially because Russia implements booster BCG vaccination and China used to recommend it in the past (http://www.bcgatlas.org/about.php).

Likewise, the agreement of TBST was substantially higher when compared with IGRA than with TST in those countries.

Limited data were available in sub-groups. Still, studies in C-Tb showed the robustness of results by HIV status. Furthermore, Diaskintest studies in children primarily including those aged < 18 years and C-Tb studies (one of which included children < 5 years old) suggested they perform similarly in children.

A considerable proportion of Diaskintest studies were not primarily designed to evaluate test performance. In these studies, Diaskintest was performed in TB dispensaries (facilities responsible for all TB care at a regional level) for indications outlined in the national recommendations which include; annual TB screening of schoolchildren to determine those in need of vaccination; initial screening to determine those who require investigation for active disease; for TB diagnosis; or to monitor treatment response. As a result, there are a number of concerns that affect the quality of the studies. Notably, clinical and test procedures across settings are inconsistent, and reporting is often insufficient. Ascertainment of TB was inadequate; the diagnosis often pragmatically made on clinical and/or radiological findings rather than microbiologically-confirmed. Although Russian national TB guidelines define Diaskintest positivity as induration of any size, more than a third of studies used the 5mm cut-off, making comparison between studies and products difficult. Incorporation bias is a risk in studies that selected study participants based on TST-positivity or had followed Russian national TB recommendations and used Diaskintest for TB diagnosis. There are also concerns that are common across the index test studies. Potential conflicts of interest are possible with many of the included studies, given many were industry-led and/or funded studies. Studies often did not stratify TST cut-off according to the history of BCG vaccination, HIV infection or other immunosuppression, which may influence test agreement, especially with the TST.

Although Russian national tuberculosis guidelines and the test manufacturer defines Diaskintest positivity as induration of any size, we identified studies that used the 5mm cut-off. Interestingly, the pooled sensitivity was lower in studies using induration of any size as cut-off than 5 mm (66% vs 88%), which is counterintuitive. However, this needs to be interpreted with caution as the comparison is indirectly based on different sets of studies.

In the primary analysis, specificity could be estimated only for C-Tb because studies on other tests were not done in low TB incidence settings. However, the levels of concordance between C-TST vs IGRA and Diaskintest vs IGRA were similar to that between C-Tb vs IGRA. Therefore, it would be reasonable to expect that they have similar specificity.

Despite the limitations, the performance of novel skin tests appears similar to IGRA. While there is no direct evidence on the predictive performance of these tests, the concordance between IGRA and the novel skin tests was high and was higher than the concordance between IGRA and TST. Given the similar predictive performance of IGRA and TST despite the level of discrepancy in results, it would be reasonable to consider that the predictive performance of the novel skin tests to be similar to be the existing tests.

Overall these tests may enable precise and accessible TBI screening that does not require expensive laboratory facilities or venepuncture.

References

1. Hay SI, Abajobir AA, Abate KH, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* 2017;390(10100):1260-344. doi: 10.1016/S0140-6736(17)32130-X
2. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLOS Medicine* 2016;13(10):e1002152. doi: 10.1371/journal.pmed.1002152
3. Comstock GW, Livesay VT, Woolpert SF. The prognosis of a positive tuberculin reaction in childhood and adolescence. *American journal of epidemiology* 1974;99(2):131-8. doi: 10.1093/oxfordjournals.aje.a121593 [published Online First: 1974/02/01]
4. Small PM, Fujiwara PI. Management of Tuberculosis in the United States. *New England Journal of Medicine* 2001;345(3):189-200. doi: 10.1056/nejm200107193450307
5. Pai M, Behr M. Latent Mycobacterium tuberculosis Infection and Interferon-Gamma Release Assays. *Microbiology spectrum* 2016;4(5) doi: 10.1128/microbiolspec.TBTB2-0023-2016 [published Online First: 2016/10/21]
6. World Health Organization.The End TB Strategy. Geneva, 2015.
7. Farhat M, Greenaway C, Pai M, et al. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2006;10(11):1192-204. [published Online First: 2006/11/30]
8. Sørensen AL, Nagai S, Houen G, et al. Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. *Infection and immunity* 1995;63(5):1710-17. doi: 10.1128/iai.63.5.1710-1717.1995
9. Andersen P, Munk ME, Pollock JM, et al. Specific immune-based diagnosis of tuberculosis. *Lancet* 2000;356(9235):1099-104. doi: 10.1016/s0140-6736(00)02742-2 [published Online First: 2000/09/29]
10. Rangaka MX, Diwakar L, Seldon R, et al. Clinical, Immunological, and Epidemiological Importance of Antituberculosis T Cell Responses in HIV-Infected Africans. *Clinical Infectious Diseases* 2007;44(12):1639-46. doi: 10.1086/518234
11. Tagmouti S, Slater M, Benedetti A, et al. Reproducibility of interferon gamma (IFN-γ) release Assays. A systematic review. *Annals of the American Thoracic Society* 2014;11(8):1267-76. doi: 10.1513/AnnalsATS.201405-188OC [published Online First: 2014/09/05]
12. Campos-Neto A, Rodrigues-Junior V, Pedral-Sampaio DB, et al. Evaluation of DPPD, a single recombinant Mycobacterium tuberculosis protein as an alternative antigen for the Mantoux test. *Tuberculosis* 2001;81(5-6):353-8.
13. Aggerbeck H, Giemza R, Joshi P, et al. Randomised clinical trial investigating the specificity of a novel skin test (C-Tb) for diagnosis of M. tuberculosis infection. *PLoS ONE [Electronic Resource]* 2013;8(5):e64215.

14. Kiselev VI, Baranovskii PM, Rudykh IV, et al. Clinical trials of the new skin test Diaskintest for the diagnosis of tuberculosis. [Russian]. *Problemy tuberkuleza i boleznei legkikh* 2009(2):11-16.
15. Ruhwald M, Aggerbeck H, Gallardo RV, et al. Safety and efficacy of the C-Tb skin test to diagnose Mycobacterium tuberculosis infection, compared with an interferon gamma release assay and the tuberculin skin test: a phase 3, double-blind, randomised, controlled trial. *The Lancet Respiratory Medicine* 2017;5(4):259-68.
16. Hoff ST, Peter JG, Theron G, et al. Sensitivity of C-Tb: a novel RD-1-specific skin test for the diagnosis of tuberculosis infection. *European Respiratory Journal* 2016;47(3):919-28.
17. Krutikov M, Faust L, Nikolayevskyy V, et al. The diagnostic performance of novel skin-based in-vivo tests for tuberculosis infection compared with purified protein derivative tuberculin skin tests and blood-based in vitro interferon-γ release assays: a systematic review and meta-analysis. *The Lancet Infectious diseases* 2021 doi: 10.1016/s1473-3099(21)00261-9 [published Online First: 2021/10/05]
18. Hamada Y, den Boon S, Maria Cirillo D, et al. Framework for the evaluation of new tests for tuberculosis infection. *European Respiratory Journal* 2021:2004078. doi: 10.1183/13993003.04078-2020
19. World Health Organization. Use of tuberculosis release assays (IGRAs) in low and middle-income countries. World Heal Organ Policy Statement. Published online 2011. doi:WHO/HTM/TB/2011.18.
20. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews* 2016;5(1):210. doi: 10.1186/s13643-016-0384-4
21. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009 [published Online First: 2011/10/19]
22. Moberg J, Oxman AD, Rosenbaum S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Research Policy and Systems* 2018;16(1):45. doi: 10.1186/s12961-018-0320-2
23. Stogova NA. Use of a sample with recombinant tuberculosis antigen (Diaskintest®) in the differential diagnosis of lung diseases of lower lobe localization. *Tuberculosis and socially infectious diseases* 2020
24. Stogova NA Immunological skin tests in differential diagnostics disseminated processes in the lungs. PULMONOLOGIYA . 2020; 30 (4): 446-452. (In Russ.) Https://doi.org/10.18093/0869-0189-2020-30-4-446-452.
25. S.L. Nakonechnaya VAA, Yu.L. Mizernitsky. COMPARATIVE EVALUATION OF THE RESULTS OF THE MANTU AND DIASKIN TEST WITH THE LABORATORY TEST QuantiFERON® - Gold In-Tube IN CHILDREN WITH TUBERCULOSIS AND CHILDREN WITH CHRONIC NON-SPECIFIC LUNG DISEASES. *Tuberculosis and socially infectious diseases* 2020(3):20-25.
26. Xu M, Lu W, Li T, et al. Sensitivity, specificity, and safety of a novel ESAT6-CFP10 skin test for tuberculosis infection in China: two randomized, self-controlled, parallel-group phase 2b trials. *Clin Infect Dis* 2021;22:22. doi: 10.1093/cid/ciab472
27. L.A. Baryshnikova RHF, D.A. Kudlay, A.P. Alekseev, M.N. Kabaeva. The technological concept of in vitro gamma-interferon tests to identify effector T cells that respond to

stimulation with the Mycobacterium tuberculosis antigen and their use among children. *Pacific Medical Journal* 2021(2)
28. Dotsenko Ya. I. SMA, Bilogortseva O. I., Pobedyonna G. P. POSSIBILITIES OF MODERN DIAGNOSTICS AND PROPHYLAXIS OF TUBERCULOSIS IN MEDICAL WORKERS OF ANTI-TUBERCULOSIS MEDICAL CENTERS (Ukrainian). *Ukrainian Journal of Occupational Health* 2015;1(42)
29. Aggerbeck H, Ruhwald M, Hoff ST, et al. C-Tb skin test to diagnose Mycobacterium tuberculosis infection in children and HIV-infected adults: A phase 3 trial. *PLoS ONE [Electronic Resource]* 2018;13(9):e0204554.
30. Aggerbeck H, Ruhwald M, Hoff ST, et al. Interaction between C-Tb and PPD given concomitantly in a split-body randomised controlled trial. *International Journal of Tuberculosis & Lung Disease* 2019;23(1):38-44.
31. Aksenova V, Barishnikova L, Klevno N, et al. New screening and diagnostic options of various traits of TB infection among children and adolescents in Russia. Curr Pediatr 2011; 10: 16–22.
32. Baryshnikova LA, Aksenova VA, Klevno NI. Detection and differential diagnostics of tuberculosis in children and adolescents. Tuberk i Bolezn lëgkih 2017; 95: 34–9.
33. Borodulina EA, Akhmerova TE, Yeremenko ES, Amsova EA, Titugina A. Features of Diagnostics of Tuberculosis in children using Diaskintest. Èpidemiologiâ i vakcinoprofilaktika 2014; 6: 51–5.
34. Borodulina EA, Lebedeva NO, Povalyaeva L V, Tsygancov IL, Borodulin BE. Tuberculin diagnostics in pulmonary Tuberculosis combined with HIV-infection. Kazan Med J 2012; 93: 576–9.
35. Slogotskaya L, Bogorodskaya E, Ivanova D, et al. Comparative sensitivity of the test with tuberculosis recombinant allergen, containing ESAT6-CFP10 protein, and Mantoux test with 2 TU PPD-L in newly diagnosed tuberculosis children and adolescents in Moscow. *PLoS ONE [Electronic Resource]* 2018;13(12):e0208705.
36. Dovgalyuk IF, Starshinova AA, Zubareva S V. Possibilities of application of Diaskintest in the practical work of the paediatrician. Pediatr J named after GNSperansky 2013; 92: 65–8.
37. Fedorovykh VS, Markelov YM, Karbauskene SI, Bakunovich T V. Results on Tuberculosis detection in children and adolescents in Karelian Republic and evaluation tests with preparation 'Diaskintest' in risk groups for Tuberculosis. Proc Petrozavodsk State Univ 2014; 6: 62–6.
38. Iablonskiy PK, Dovgaljuk IF, Starshinova AA, Iakunova O. Role of modern immunological tests in diagnostics of Tuberculosis in children. Med Immunol 2013; 15: 37–44.
39. Kabanets NN, Filinyuk OV, Urazova OI, Morozova KS, Kolokolova OV. Comparative evaluation of immunological tests for children at risk in the Tomsk Region. Bull Sib Med 2016; 15: 30–8.
40. Kibrik BS, Melnikov VP. The results of allergen Tuberculosis recombinant usage among healthy population, people with social contacts and TB patients. Tuberc Soc Signif Dis 2015; 1: 36–9.
41. Koretskaya NM, Bolshakova IA, Chushkina AA. Comparative evaluation of the results of Mantoux test with 2 TE PPD-L and Diaskintest in medical university students. Sib Med J 2012; 112: 97–9.
42. Laushkina ZA, Krasnov VA, Petrenko TI. Immunological tests in differential diagnostics of Pulmonary Tuberculosis. Tuberk i Bolezn Legk 2017; 95: 26–30.

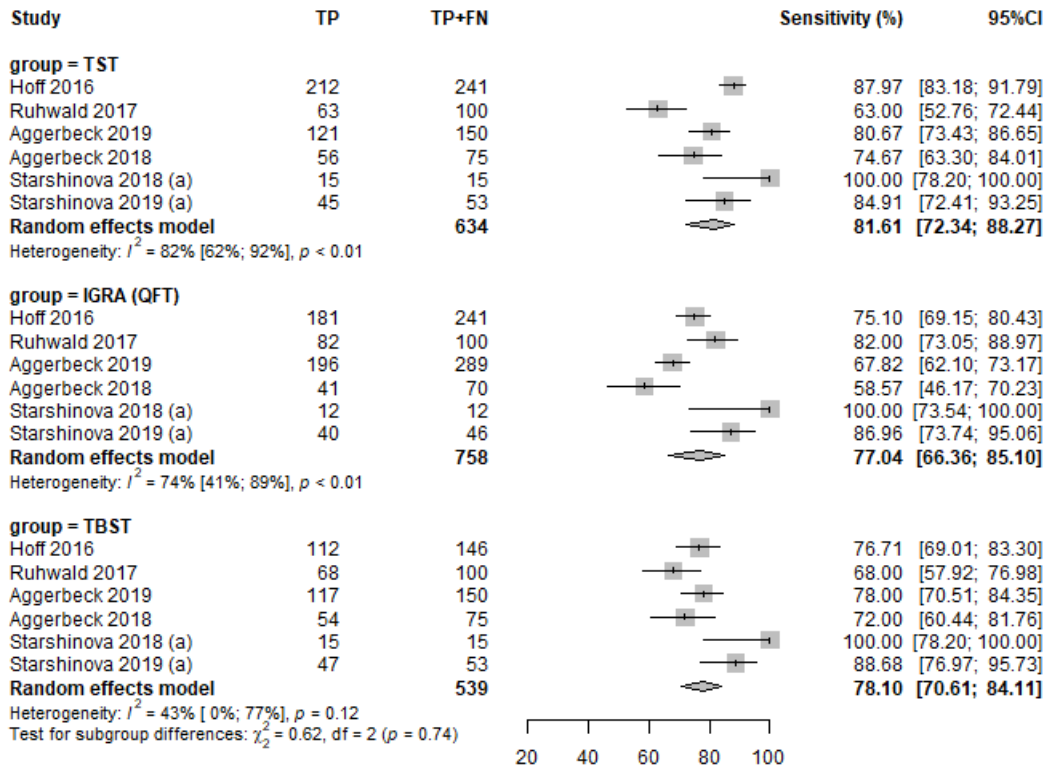
43. Losovskaya ME, Belushkov VV, Gurina OP, et al. The comparison of laboratory tests Quantiferon, Tubinferon and Diaskintest in children with tuberculosis infection. Russ Clin Lab Diagnostics 2016; 61: 838–42.
44. Lozovskaya ME, Belushkov VV, Gurina OP, Vasilyeva YB, Klochkova LV. Comparative evaluation of innovative diagnostic tests for latent and active TB infection in children. Pediatr 2014; 5: 46–50.
45. Ministry of Health, and Social Development of the Russian Federation. Order of the Ministry of Health and Social Development of the Russian Federation of October 29, 2009 N 855 'On Amendments to Appendix N 4 to the Order of the Ministry of Health of Russia of March 21, 2003 N 109'. 2009. https://www.garant.ru/products/ipo/prime/doc/4089228/ (accessed Feb 7, 2020).
46. Mishin V, Morozov I, Grigor'yev Y, et al. A diagnostic and clinical significance of the intradermal Mantoux test and Diaskintest in young healthy individuals and people with Pulmonary Tuberculosis. Russ Infect Dis 2016; 14: 56–62.
47. Nikitina IY, Karpina NL, Kasimceva O V., Gergert VY, Ergeshov A, Lyadova I V. Comparative performance of QuantiFERON-TB Gold versus skin test with tuberculosis recombinant allergen (Diaskintest) among patients with suspected pulmonary tuberculosis in Russia. Int J Infect Dis 2019; 86: 18–24.
48. Salina T, Morozova TI. Immunological methods in differential diagnostics. Tuberk i Bolezn lëgkih 2011; 88: 50–3.
49. Salina TY, Morozova TI, Kudlay DA. Results of the skin test of diaskintest in pulmonary tuberculosis patients caused by different genotypes of M. tuberculosis. Tuberc Lung Dis 2019; 97: 66–7.
50. Samorodov NA, Kibishev VM, Tilova LA, et al. Analysis of the results of the use of recombinant TB allergen 'Diaskintest' in patients with respiratory diseases. J New Med Technol eEdition 2019; 13: 114–9.
51. Senin AM, Eismont N V, Golubev DN. 'Diaskintest' as evaluation of activity of the disease in patients with Tuberculosis and HIV co-infection. Vestn Sovrem Klin Med 2016; 9: 101–7.
52. Shovkun LA, Romantseva NE, D KE. Diagnostics of the active and latent Tubercular infection among children and teenagers using Diaskintest preparation. Med Vestn Ûga Ross 2014; 1: 124–9.
53. Slogotskaya LI, Litvinov VI, Seltsovsky PP, et al. A skin test with recombinant allergen of Mycobacterium Tuberculosis (Diaskintest) to detect Tuberculosis in HIV patients. Pulmonologiya 2011; 1: 60–4.
54. Slogotskaya LV, Ivanova DA, Kochetkov YA, Kulikovskaya NV, Vaneeva TV, Filippov AV. Comparative results of the skin test with the preparation containing the recombinant protein CFP-10-ESAT-6 and the laboratory test Quantiferon-GIT. Tuberk i Bolezn lëgkih 2012; 89: 27–32.
55. Slogotskaya LV, Kochetkov YA, Filippova AV. Diaskintest is a new method to detect Tuberculosis. Tuberk i Bolezn lëgkih 2011; 88: 17–22.
56. Slogotskaya LV, Sentchichina O, Bogorodskaya EM. Sensitivity of the test with the Tuberculosis allergen, containing recombinant protein ESAT6-CFP10, in new cases of Tuberculosis in children and adolescents in Moscow. Tuberc Soc Signif Dis 2013; 1: 37–44.

57. Starshinova AA, Istomina EV, Zinchenko YS, et al. Diagnostic value of specific immune complexes in detection of active Tuberculosis infection. Med Immunol 2019; 21: 269–78.
58. Starshinova AA, Zinchenko YS, Istomina E V., et al. Diagnosis of Latent Tuberculosis Infection in Personnel of Various Institutions and Determination of the Risk Group for Tuberculosis. Bioprep Prev Diagnosis, Treat 2019; 19: 178–84.
59. Vaganova US, Russkih OE. Ways of rising detectability of Tuberculosis among health workers. J Pharm Qual Assur 2015; 2: 61–7.
60. A S. A comparison of intradermal test with recombinant tuberculosis allergen (diaskintest) with other immunologic tests in the diagnosis of tuberculosis infection. *Int J Mycobacteriology* 2018;7:32-9.
61. Zhang H, Wang L, Li F, et al. Induration or erythema diameter not less than 5 mm as results of recombinant fusion protein ESAT6-CFP10 skin test for detecting M. tuberculosis infection. *BMC Infect Dis* 2020;20(1):685. doi: 10.1186/s12879-020-05413-9 [published Online First: 2020/09/20]
62. Li F, Xu M, Qin C, et al. Recombinant fusion ESAT6-CFP10 immunogen as a skin test reagent for tuberculosis diagnosis: an open-label, randomized, two-centre phase 2a clinical trial. *Clinical Microbiology & Infection* 2016;22(10):889.e9-89.e16.
63. Badaro R, Machado BAS, Duthie MS, et al. The single recombinant M. tuberculosis protein DPPD provides enhanced performance of skin testing among HIV-infected tuberculosis patients. *AMB Express* 2020;10(1):133. doi: 10.1186/s13568-020-01068-6 [published Online First: 2020/08/02]
64. Levi DT. Modern tuberculin preparation — methods of receiving, control, standardization and use. Moscow, Russia, 1987.
65. Starshinova A, Zhuravlev V, Dovgaluk I, et al. A comparison of intradermal test with recombinant tuberculosis allergen (diaskintest) with other immunologic tests in the diagnosis of tuberculosis infection. *International Journal of Mycobacteriology* 2018;7(1):32-39.
66. Zwerling A, Behr MA, Verma A, et al. The BCG World Atlas: A Database of Global BCG Vaccination Policies and Practices. *PLOS Medicine* 2011;8(3):e1001012. doi: 10.1371/journal.pmed.1001012
67. Starshinova AA, Istomina EV, Belyaeva EN, Yablonsky PK. Detection of latent tuberculosis infection in various units of an anti-TB facility. Med Alliance 2018;4:11-8.
68. Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Annex 1: Pooled results of the systematic review.

Figure 1. Sensitivity of TBST in head-to-head studies.

{Ruhwald, 2017 #2778}{Ruhwald, 2017 #2778}

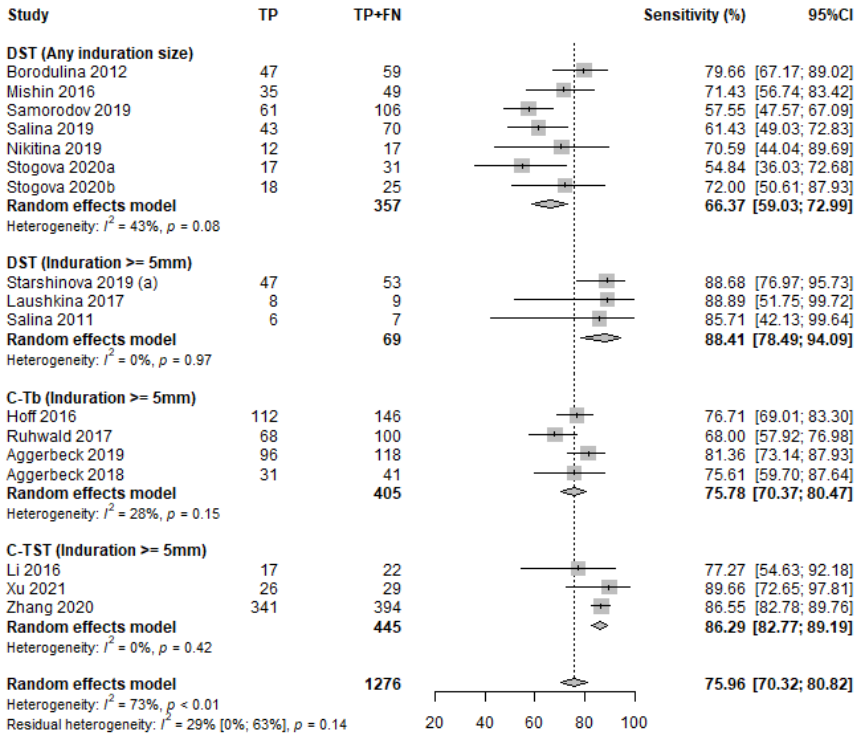


Starshinova 2018 (a) and Starshinova 2019 (a) studied DST^{5mm}. The rest studied C-Tb.

TST cut-off was 5mm for HIV+ and 15 mm for HIV- in Aggerbeck, 2018, Aggerbeck 2019, Hoff 2016, and Ruhwald 2017. TST cut-off was 5mm in Starshinova 2018 (a) and Starshinova 2019 (a).

TBST: novel skin tests for TB infection; TST: tuberculin skin test: QFT: quantiferon; CT: confidence interval; TP: true positive; FN: false negative.

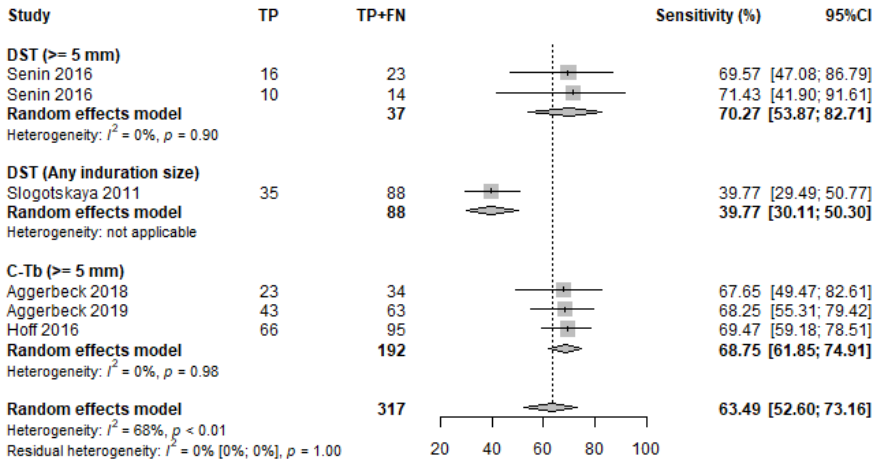
Figure 2. Sensitivity of TBST in all studies in individuals with HIV-negative or unknown status



The pooled sensitivity when DST^{5mm} was excluded: 74.21% [95%CI 68.30; 79.36]

TBST: novel skin tests for TB infection; DST: Diaskintest; TST: tuberculin skin test: QFT: quantiferon; CI: confidence interval; TP: true positive; FN: false negative.

Figure 3. Sensitivity of TBST in HIV-positive individuals

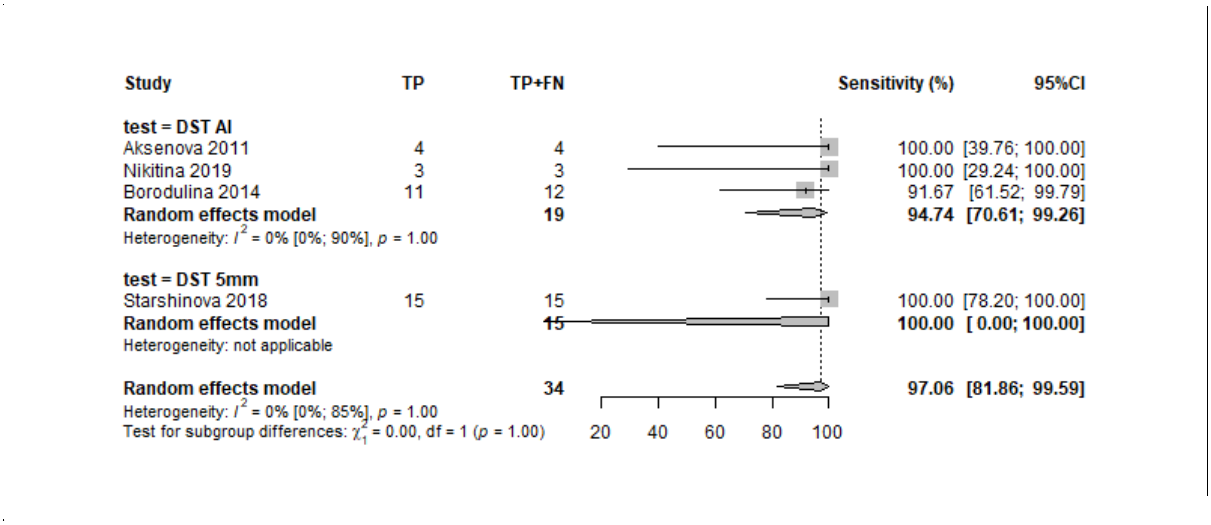


–

Pooled sensitivity when DST^{5mm} was excluded: 61.31% [47.90; 73.19]

TBST: novel skin tests for TB infection; DST: Diaskintest; TST: tuberculin skin test: QFT: quantiferon; CI: confidence interval; TP: true positive; FN: false negative.

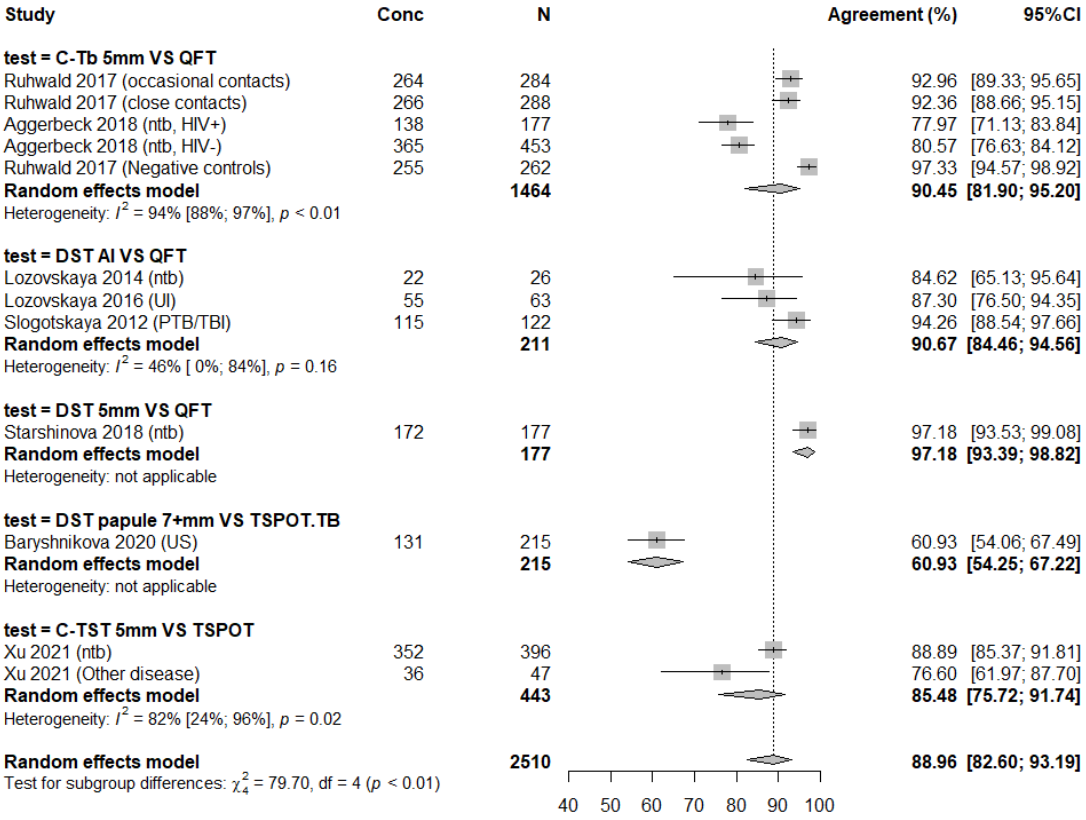
Figure 4 Sensitivity of TBST in children



TBST: novel skin tests for TB infection; DST: Diaskintest; AI: Any induration size; CI: confidence interval; TP: true positive; FN: false negative.

Aggerbeck 2018 estimated the sensitivity of C-Tb in 12 children with TB but only two of them were bacteriologically confirmed. This was not included in the plot.

Figure 5 Agreement of TBST vs IGRA in all studies including participants without active TB

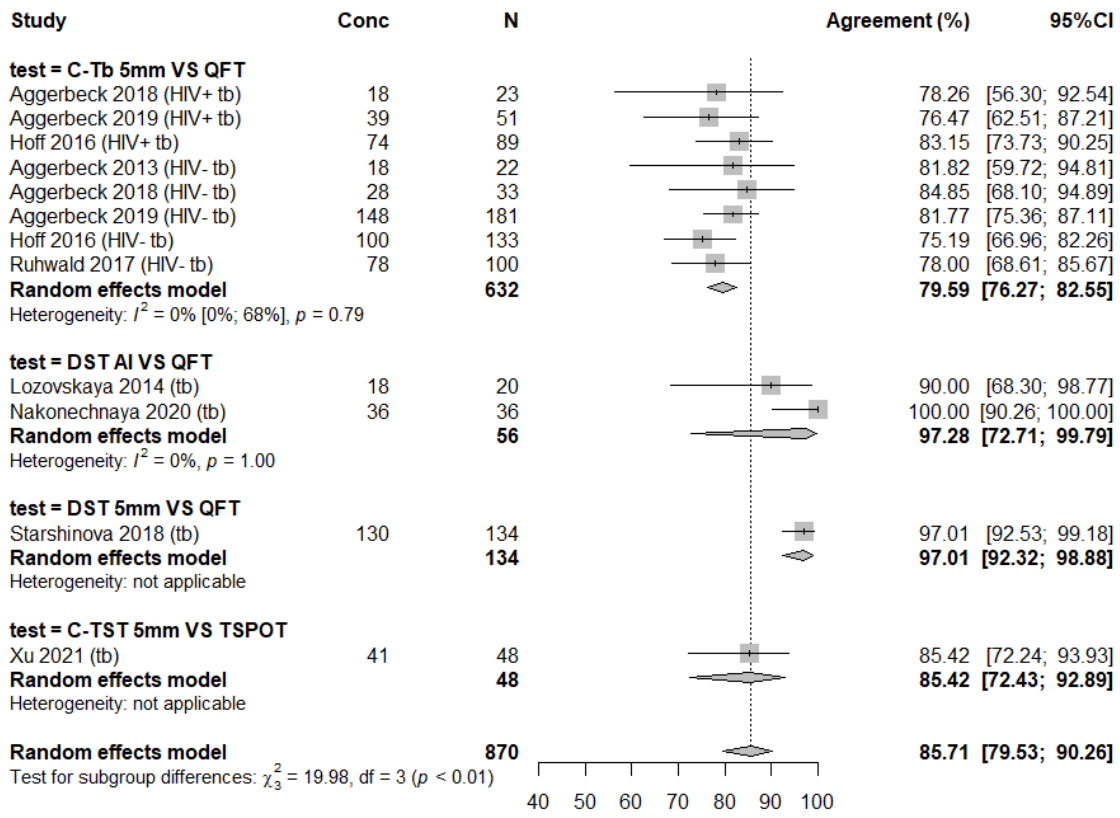


Pooled agreement when only DST^{AI} was used for DST: 89.28 [95%CI 84.13; 92.90]

TBST	novel skin tests for TB infection	**QFT**	quantiferon
DST	Diaskintest	**CI**	confidence interval
AI	any induration size	**ntb**	non-TB
US	Under surveillance for TB		

PTB/TBI study sample includes individuals with pulmonary TB or latent TB infection

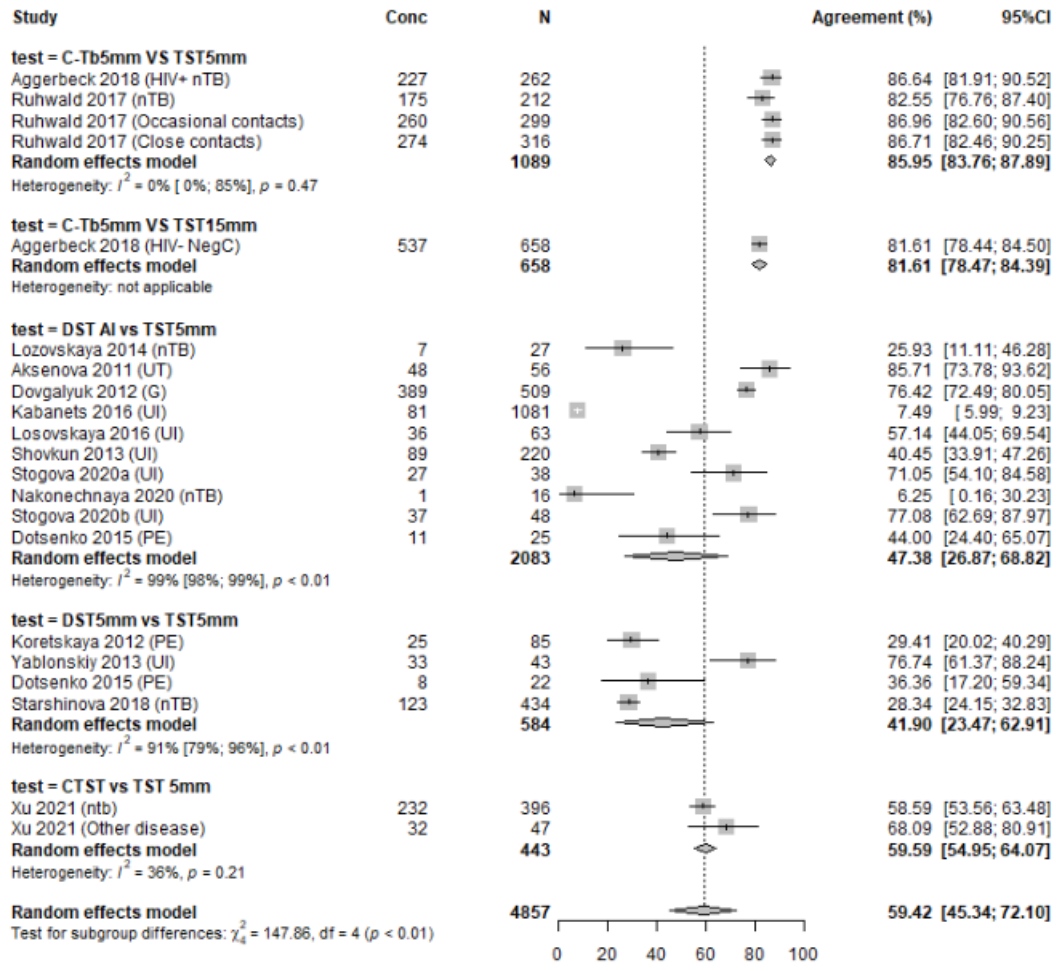
Figure 6. Agreement of TBST vs IGRA in all studies including people with active TB



Pooled agreement when DST ^{5mm} was excluded: 81.83 [95%CI 77.56; 85.45]

TBST	novel skin tests for TB infection	**QFT**	quantiferon
DST	Diaskintest	**CI**	confidence interval
AI	any induration size		

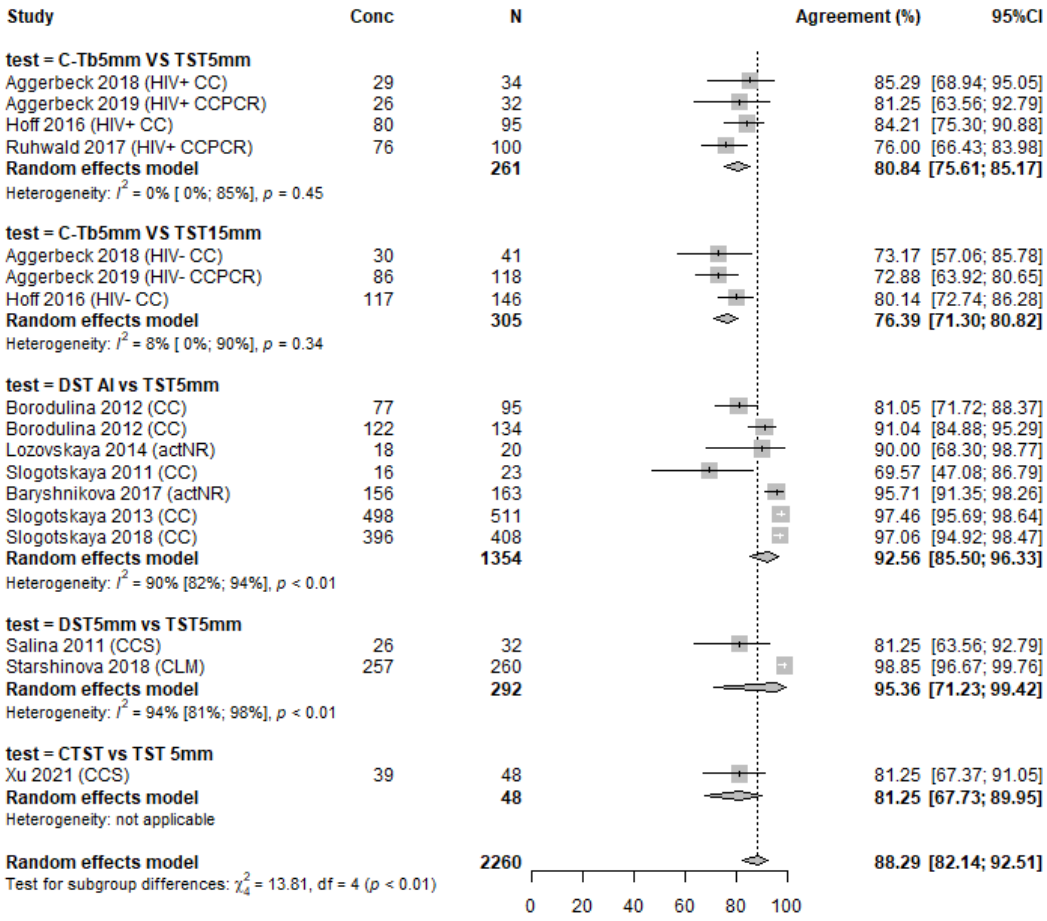
Figure 7 Agreement of TBST vs TST in all studies including participants without active TB



Pooled agreement when DST^{5mm} was excluded: 63.20% [95%CI: 47.38; 76.62]

UI under investigation for TB
nTB non-TB
G general population
UT active TB or under investigation for TB, results not reported separately
DST Diaskintest
PE potentially exposed (health care workers)
CI confidence interval
TST tuberculin skin test

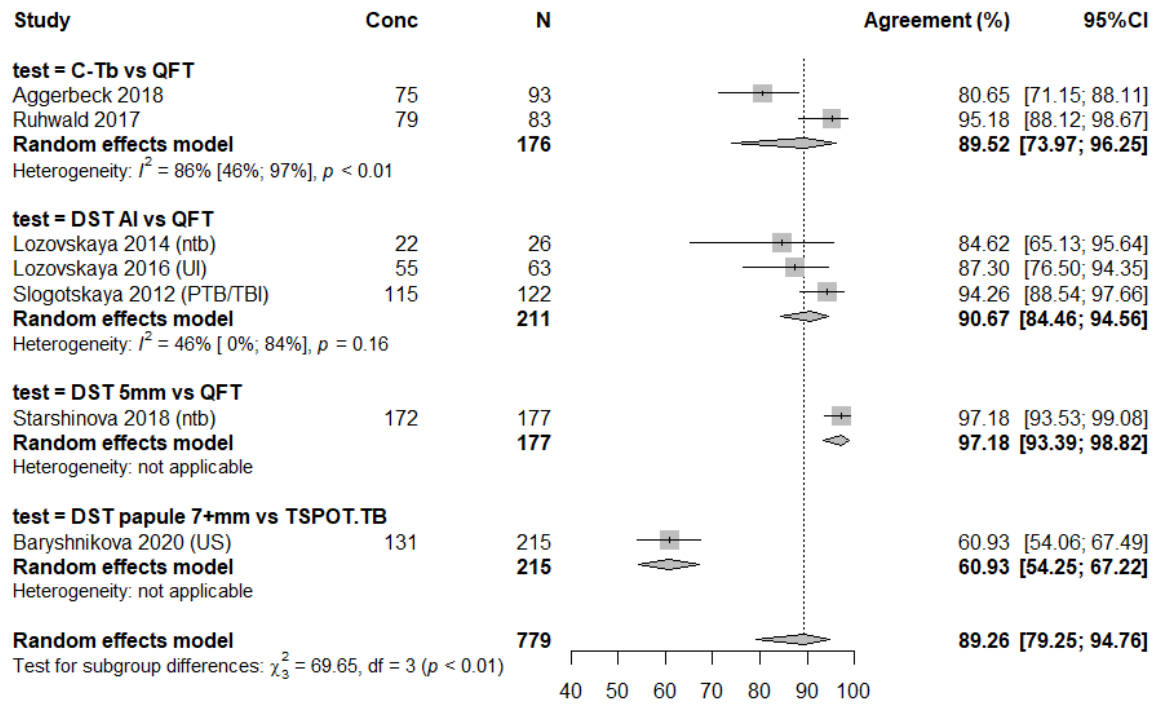
Figure 8 Agreement of TBST vs TST in all studies including people with active TB



Pooled agreement when DST^{5mm} was excluded: 86.98% [80.76-91.41%]

actNR active TB, diagnostic method not reported **TST** tuberculin skin test
CI confidence interval **CC** culture-confirmed TB
CCPCR culture or PCR confirmed
CLM clinically or microbiologically confirmed TB
CCS culture or smear confirmed TB **DST** Diaskintest

Figure 9 Agreement of TBST vs IGRA in children without active TB

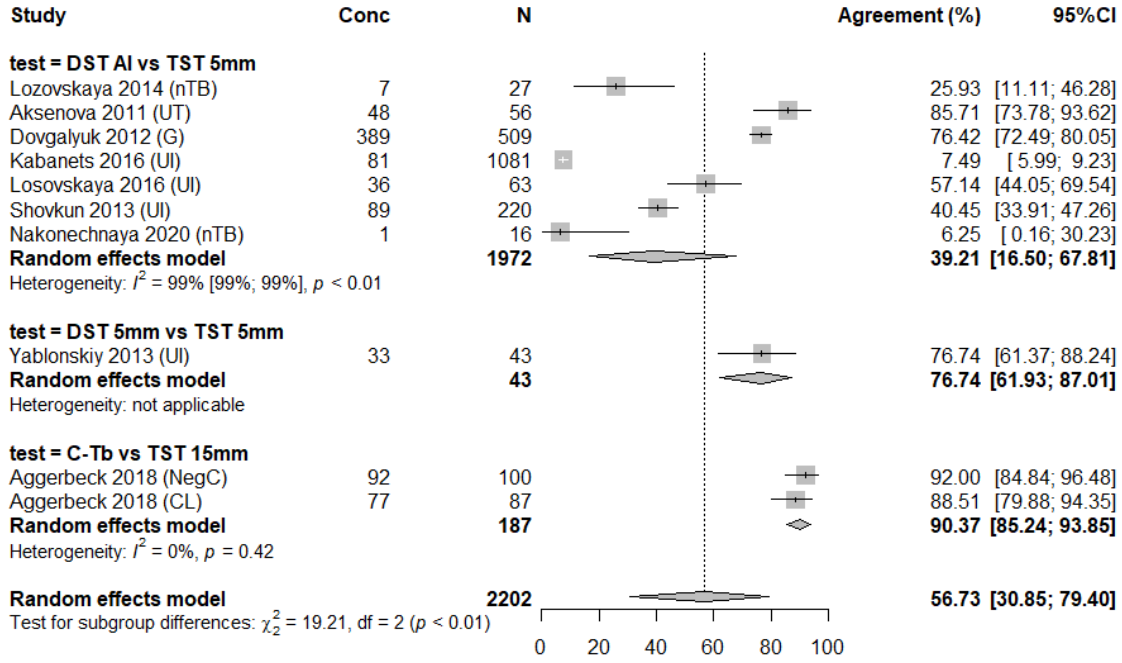


Pooled agreement when DST^{5mm} was excluded: 54.02 [26.46; 79.31]

TBST	novel skin tests for TB infection	**QFT**	quantiferon
DST	Diaskintest	**CI**	confidence interval
AI	any induration size	**ntb**	non-TB
US	Under surveillance for TB		

PTB/TBI study sample includes individuals with pulmonary TB or latent TB infection

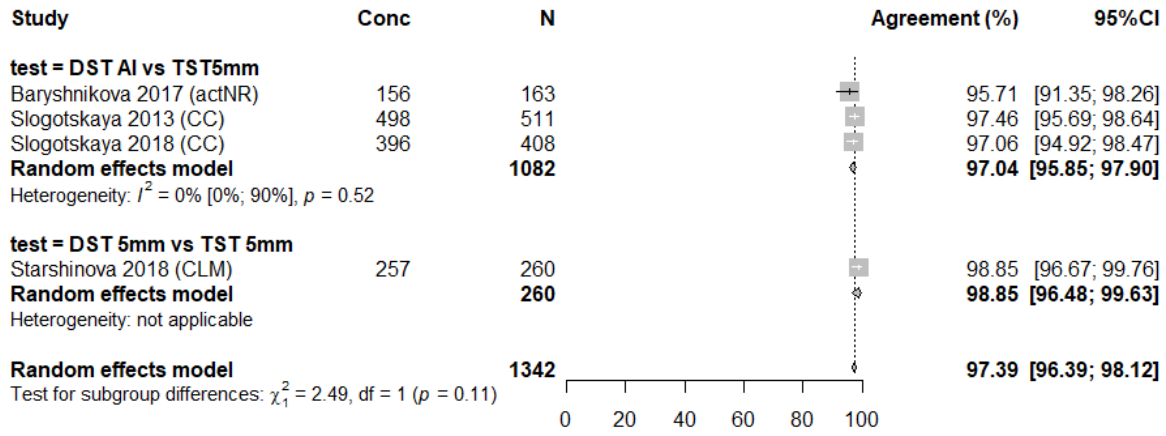
Figure 10 Agreement of TBST vs TST in children without active TB



Pooled agreement when DST^{5mm} was excluded 89.81 [83.30; 93.96]

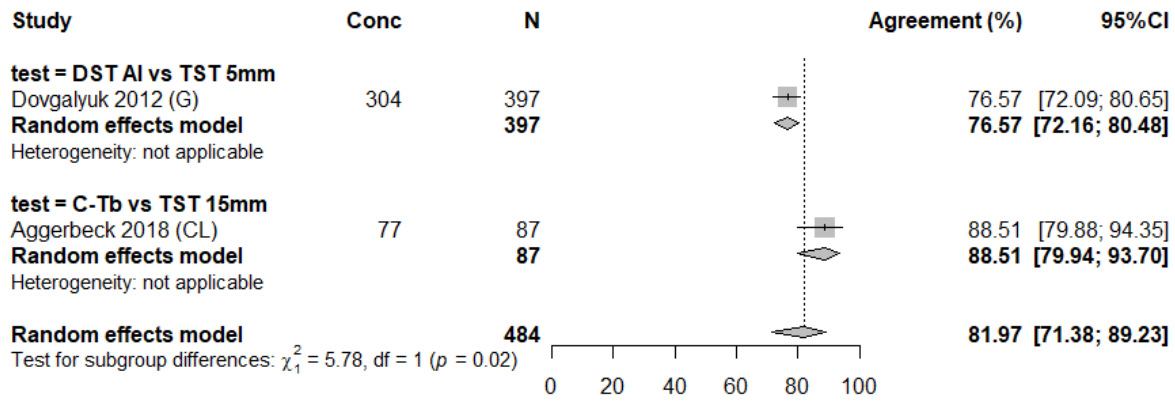
UI	under investigation for TB	**CI**	confidence interval
nTB	non-TB	**TST**	tuberculin skin test
G	general population		
UT	active TB or under investigation for TB, results not reported separately		
DST	Diaskintest	**NegC**	Negative control
CL	Close contacts		

Figure 11 Agreement of TBST vs TST in children with active TB



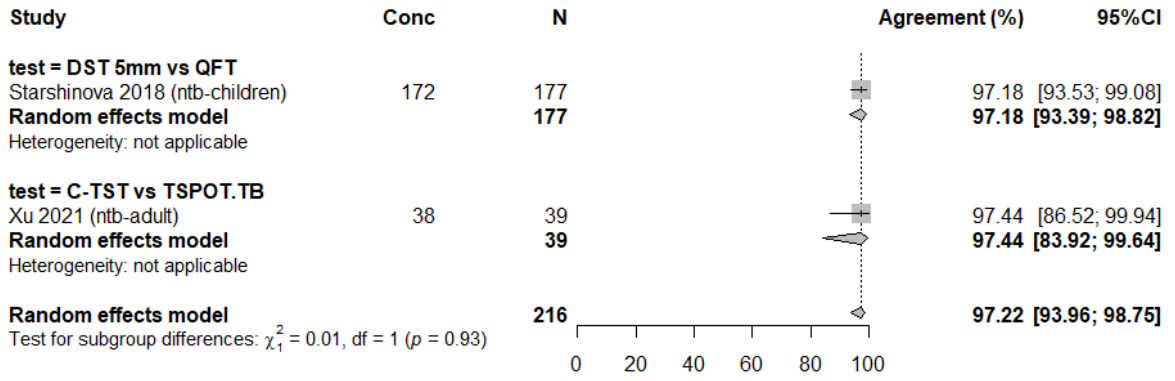
actNR active TB, diagnostic method not reported **TST** tuberculin skin test
CI confidence interval **CC** culture-confirmed TB
CLM clinically or microbiologically confirmed TB
CCS culture or smear confirmed TB **DST** Diaskintest

Figure 12 Agreement of TBST vs TST in children aged < 5 years without active TB



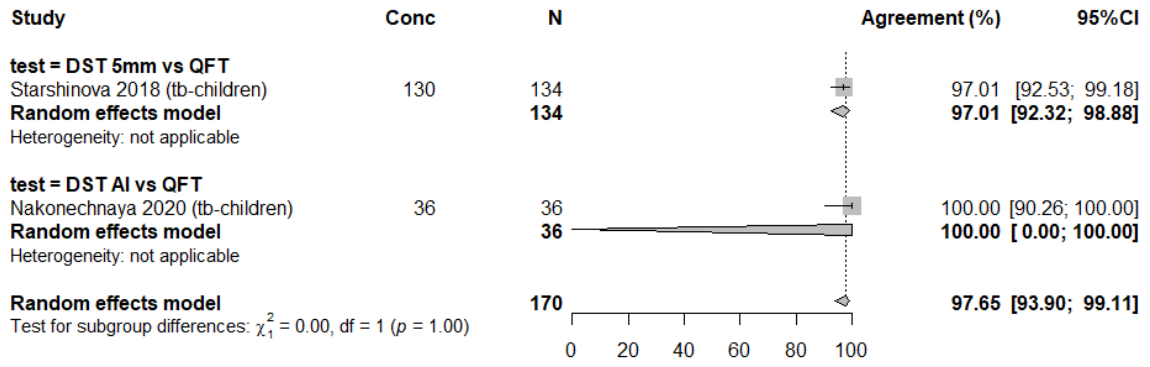
CI confidence interval
G general population
DST Diaskintest
TST tuberculin skin test
CL Close contacts

Figure 13 Agreement of TBST vs QFT in BCG-vaccinated individuals without active TB



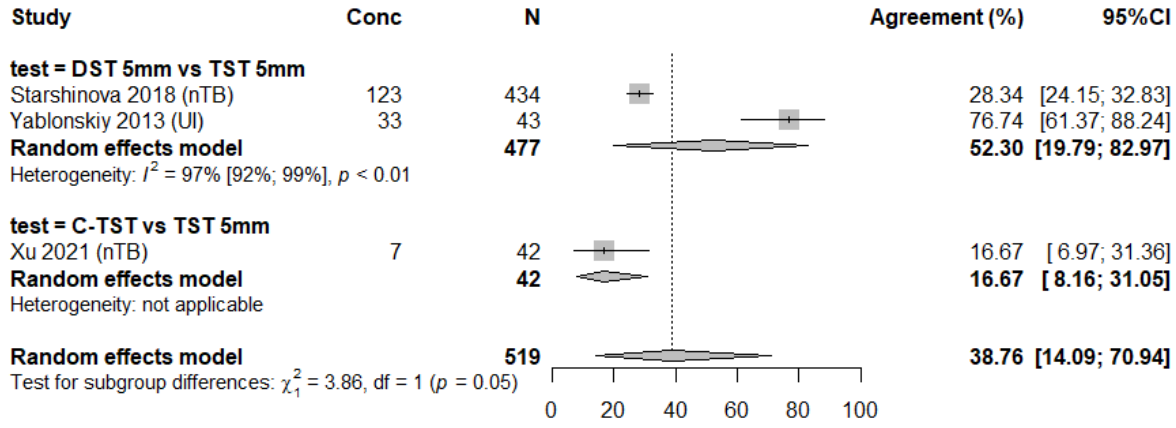
CI confidence interval
ntb non-TB
DST Diaskintest
QFT quantiferon

Figure 14 Agreement of TBST vs QFT in BCG-vaccinated individuals with active TB



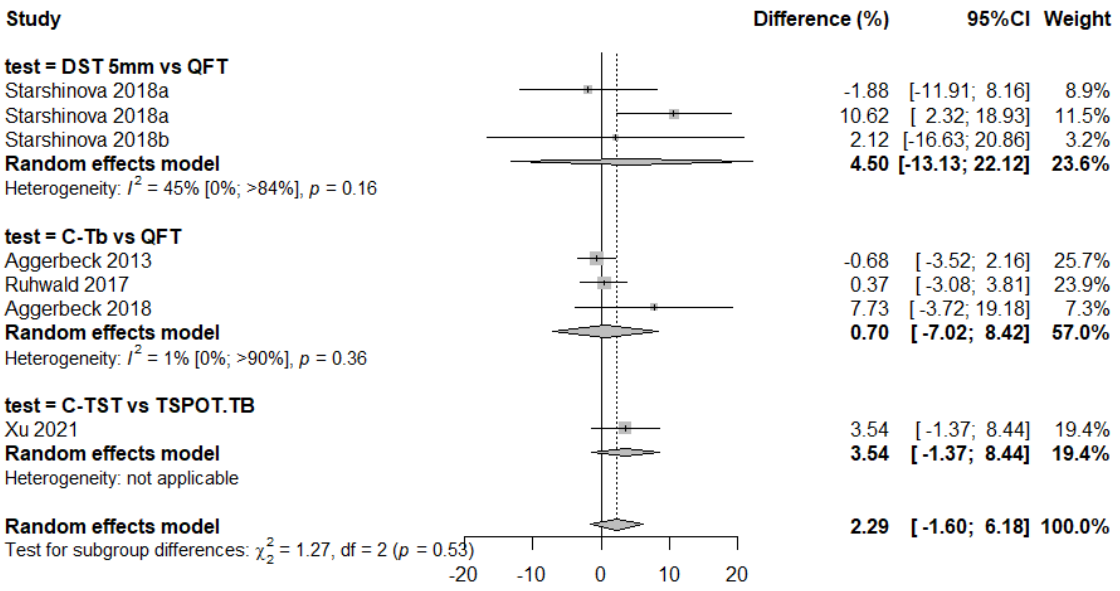
CI confidence interval **QFT** quantiferon
ntb **non-TB**
DST Diaskintest

Figure 15 Agreement of TBST vs TST in BCG-vaccinated individuals without active TB



CI confidence interval **TST** tuberculin skin test
ntb non-TB **UI** under investigation for TB
DST Diaskintest

Figure 16 Difference in specificity -TBST vs IGRA (using data on DST vs QFT from Starshinova 2018)

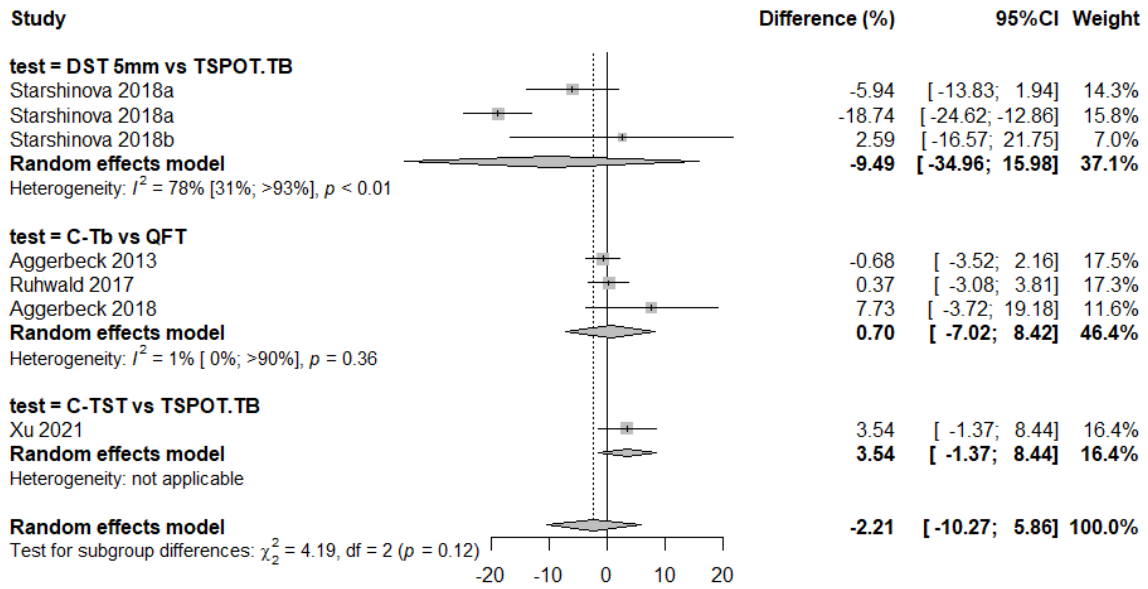


CI confidence interval **QFT** quantiferon
DST Diaskintest

Note: Starshinova 2018a and Starshinova 2018b presented data on DST 5mm vs QFT and DST 5mm and TSPOT.TB using the same cohort of participants. Because of the overlap in the cohort, both DST 5mm vs QFT and DST 5mm and TSPOT.TB cannot be pooled together. Hence, the graph includes only estimates for DST 5mm vs QFT. The next graph includes DST 5mm vs TSPOT.TB.

*Alternate specificity measure. Estimates difference in proportion negative in healthy populations (see the method section for the definition).

Figure 17 Difference in specificity - TBST vs IGRA (using data on DST vs TSPOT.TB from Starshinova 2018)

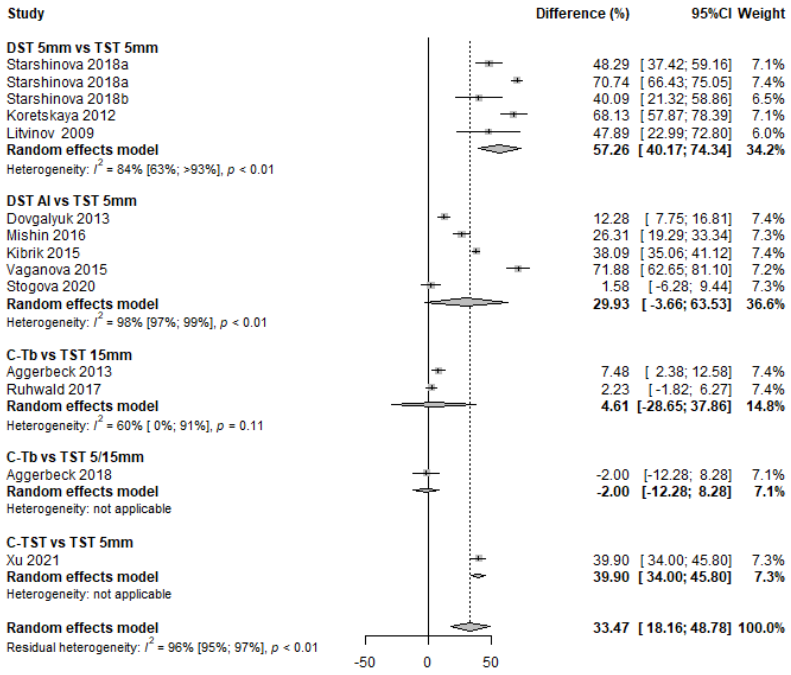


CI confidence interval **QFT** quantiferon
DST Diaskintest

Note: Starshinova 2018a and Starshinova 2018b presented data on DST 5mm vs QFT and DST 5mm and TSPOT.TB using the same cohort of participants. Because of the overlap in the cohort, both DST 5mm vs QFT and DST 5mm and TSPOT.TB cannot be pooled together. Hence, the graph includes only estimates for DST 5mm vs TSPOT.TB. The previous graph includes DST 5mm vs QFT.

*Alternate specificity measure. Estimates difference in proportion negative in healthy populations (see the method section for the definition).

Figure 18 Difference in specificity - TBST vs TST

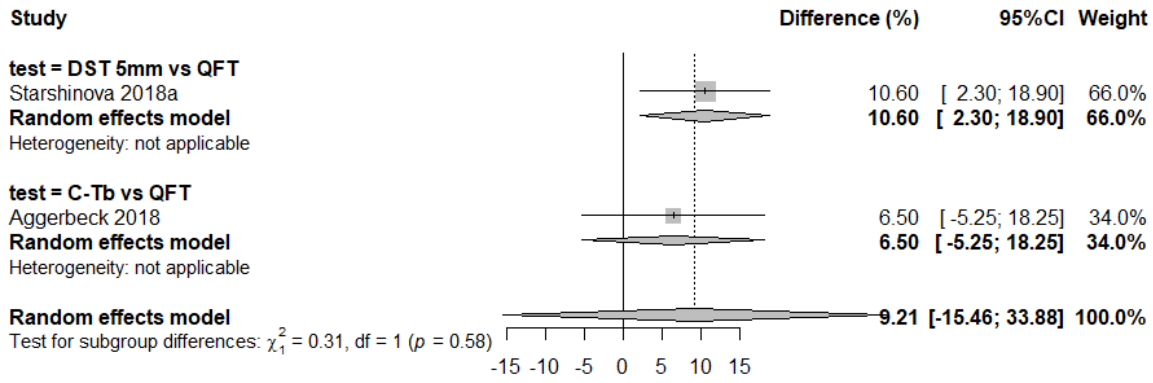


Pooled difference when DST^{5mm} was excluded: 21.95 [3.26; 40.63]

CI confidence interval **TST** tuberculin skin test
DST Diaskintest

*Alternate specificity measure. Estimates difference in proportion negative in healthy populations (see the method section for the definition).

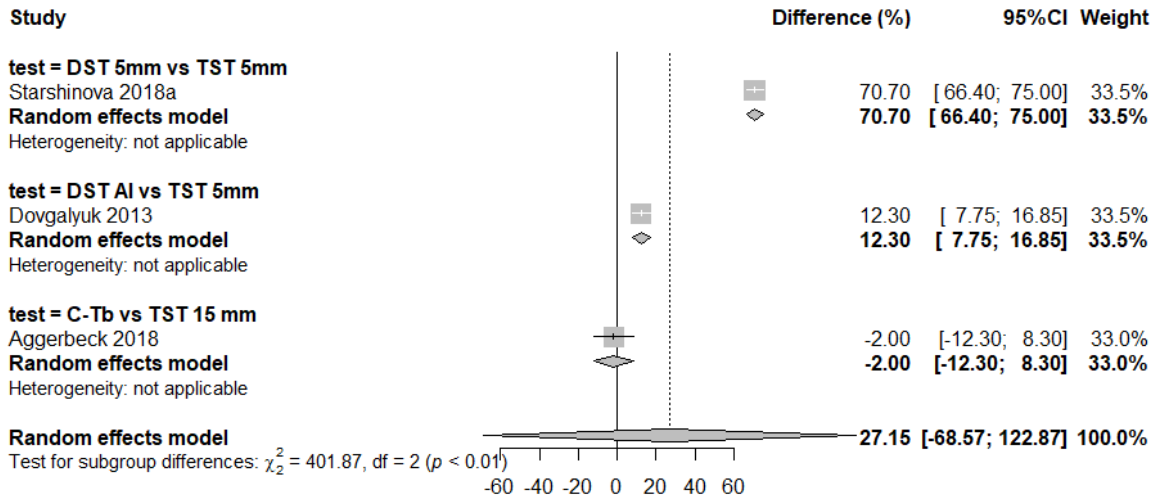
Figure 19 Difference in specificity - TBST vs IGRA in children



CI confidence interval **QFT** quantiferon
DST Diaskintest

*Alternate specificity measure. Estimates difference in proportion negative in healthy populations (see the method section for the definition).

Figure 20 Difference in specificity - TBST vs TST in children

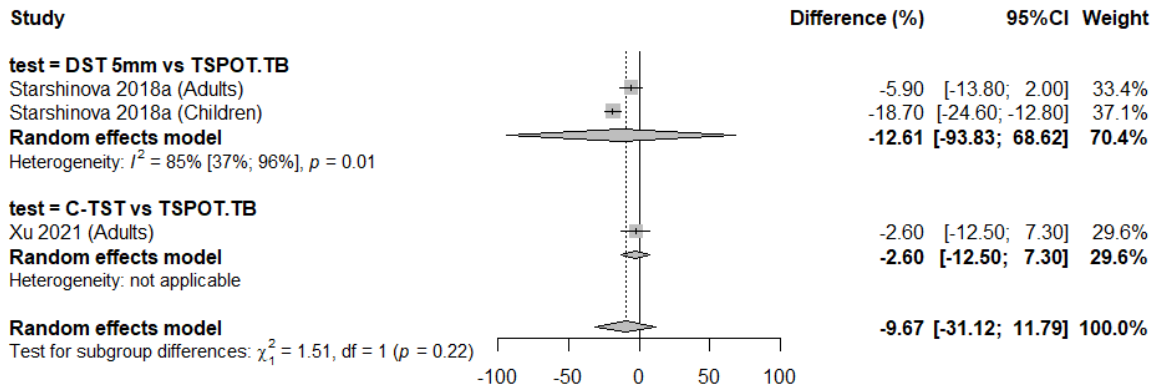


Pooled difference when DST^{5mm} was excluded: 20.17 [95%CI 4.93; 35.42]

CI confidence interval **TST** tuberculin skin test
DST Diaskintest

*Alternate specificity measure. Estimates difference in proportion negative in healthy populations (see the method section for the definition).

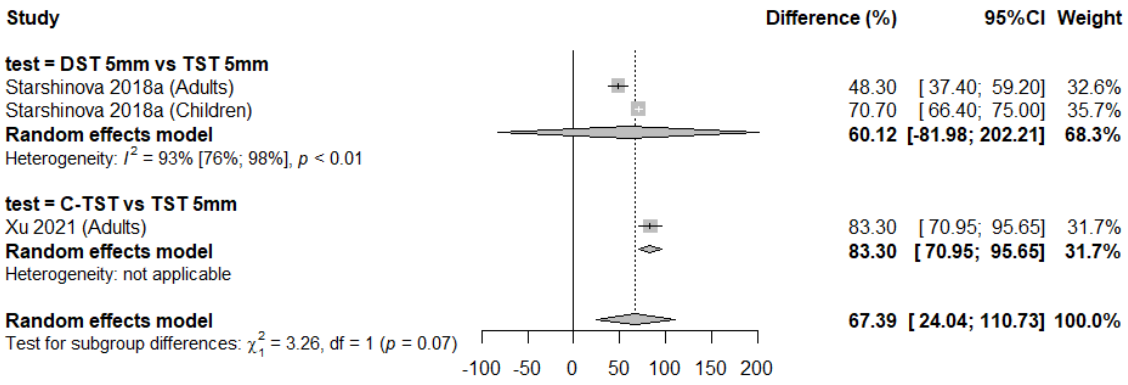
Figure 21 Difference in specificity - TBST vs IGRA in BCG-vaccinated individuals



CI confidence interval **QFT** quantiferon
DST Diaskintest

*Alternate specificity measure. Estimates difference in proportion negative in healthy populations (see the method section for the definition).

Figure 22 Difference in specificity- TBST vs TST in BCG-vaccinated individuals

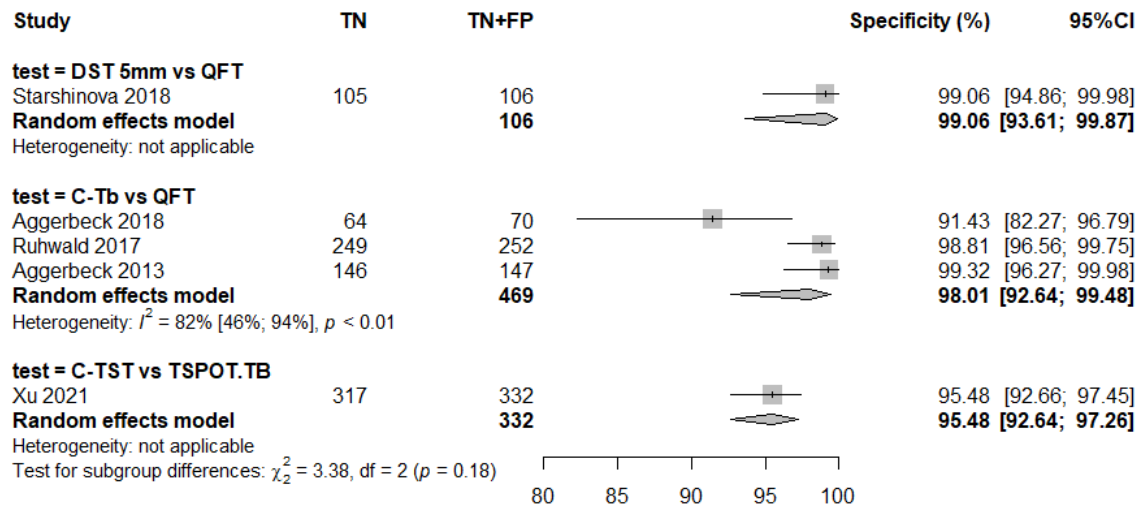


Pooled difference when DST^{5mm} was excluded: 20.17 [95%CI 4.93; 35.42]

CI confidence interval **TST** tuberculin skin test
DST Diaskintest

*Alternate specificity measure. Estimates difference in proportion negative in healthy populations (see the method section for the definition).

Figure 23 Specificity in healthy individuals with negative IGRA results

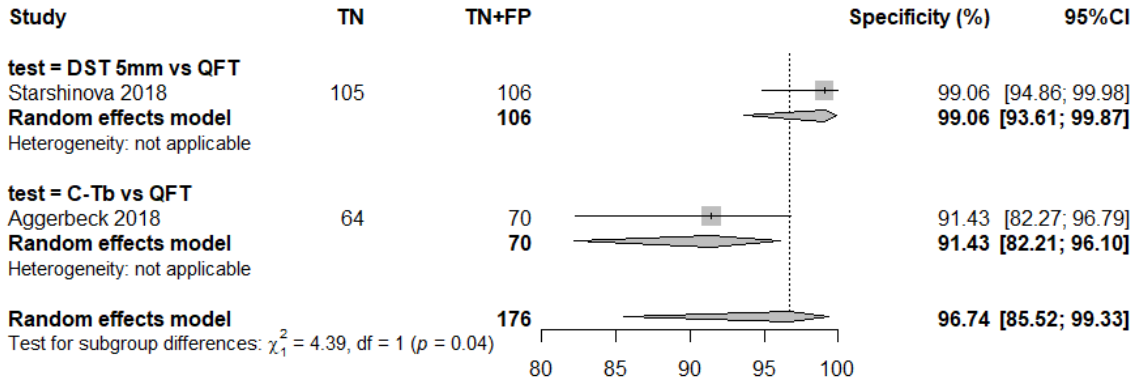


Pooled estimated when DST^{5mm} vs TSPOT.TB was used instead of DST^{5mm} vs QFT: 98.46 [95%CI 94.36; 99.60]

CI confidence interval **TST** tuberculin skin test

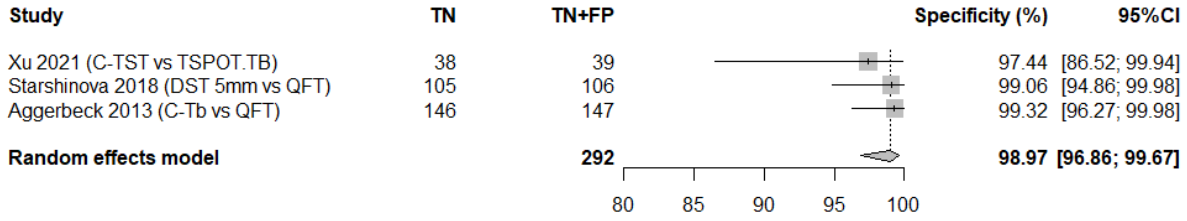
DST Diaskintest

Figure 24 Specificity in healthy children with negative IGRA results



CI confidence interval **QFT** quantiferon
DST Diaskintest

Figure 25 Specificity in BCG-vaccinated individuals with negative IGRA results



Pooled estimated when DST^{5mm} vs TSPOT.TB was used instead of DST^{5mm} vs QFT: 99.49 [97.97; 99.87]

CI confidence interval **QFT** quantiferon
DST Diaskintest

World Health Organization