WHO consolidated guidelines on tuberculosis

Module 3: diagnosis. Tests for TB infection

Web Annex D

Cost–effectiveness of *Mycobacterium tuberculosis* antigen-based skin tests: a systematic review

Review team: Lara Goscé[1], Finna Parkinson[1], Kasim Allel,[1,2] Elena Surkova[3], Irina Kontsevaya[4], Ting Ting Wang[1], Victoria Liu[2], Yohhei Hamada[1], Peter White[5], Molebogeng X Rangaka[1]

[1]Institute for Global Health, University College London, United Kingdom

[2]London School of Hygiene and Tropical Medicine, United Kingdom

[3]Royal Brompton Hospital. Part of Guy's and St Thomas' NHS Foundation Trust London, United Kingdom

[4]Research Center Borstel, Germany

[5]Imperial College, United Kingdom

Technical advisor: Peter White[1], Ibrahim Abubakar[2]

1. Faculty of Medicine, School of Public Health, Imperial College, United Kingdom
2. Faculty of Population Health Science, University College London, United Kingdom

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

Report version 1.0, Date 10 December 2021

World Health Organization

WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for TB infection. Web Annex D. Cost–effectiveness of *Mycobacterium tuberculosis* antigen-based skin tests: a systematic review/ Lara Goscé, Finna Parkinson, Kasim Allel, Elena Surkova, Irina Kontsevaya, Ting Ting Wang et al.

ISBN 978-92-4-005662-6 (electronic version)

Suggested citation. Goscé L, Parkinson F, Allel K, Surkova E, Kontsevaya I, Wang TT et al. Web Annex D. Cost–effectiveness of *Mycobacterium tuberculosis* antigen-based skin tests: a systematic review. 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

Abbreviations

BCG: Bacillus Calmette-Guérin

CBA: Cost-benefit analysis

CUA: Cost-utility analysis

CEA: Cost-effective analysis

DOT: Directly observed therapy

HICs: High-income countries

HIV: Human immunodeficiency virus

ICER: incremental cost-effectiveness ratios

IGRA: interferon-gamma release assays

LMICs: Low- and middle-income countries

PPD: purified protein derivative

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PSA: probabilistic sensitivity analysis

QALYS: Quality Adjusted Life Years

QFT: QuantiFERON-TB Gold

TB: Tuberculosis

TBI: Latent tuberculosis infection

TBST: *Mycobacterium tuberculosis* specific antigen-based skin tests

TPT: Tuberculosis preventive treatment

TST: Tuberculin skin tests

USD: United states dollars

WHO: World Health Organization

Background

With an estimated 1.7 billion people infected with tuberculosis infection (TBI), progression of infection to active TB disease poses a large public health risk[1]. Prominent risk factors of this transition include clinical risk factors such as human immunodeficiency virus (HIV), diabetes, undernutrition, and contextual or societal risk factors in the most vulnerable and disadvantaged populations including, TB case contacts especially children, immigrants[2]. WHO recommends TB preventive treatment (TPT) to disrupt disease progression. Testing for TB infection is recommended to guide treatment where possible. However, access and affordability of tests often present as barriers. Efficient and affordable tests for TBI are thus necessary in the effort to halt (re)activation and spread of TB.[2]

Different strategies and tests are currently used to identify TBI, they are the tuberculin skin tests (TST) based on non-specific mycobacterial antigen, purified protein derivative (PPD), and the RD1-specific interferon-gamma release assays (IGRA). Even though both tests are useful in the control of TB, they present implementation weaknesses. The PPD-TST requires two clinical visits within 2-3 days of testing which might be expensive and unfeasible for those individuals having limited access to healthcare, resulting in incomplete processes. Also, the PPD-TST has low specificity for people with previous Bacillus Calmette-Guérin (BCG) vaccination[3], which is universally delivered at birth in many countries that have high exposure to TB [4]. IGRA test, on the other hand, is rapidly read through the use of blood samples, requires one clinical visit for testing and results remain on record indefinitely. Moreover, IGRA has higher specificity in BCG vaccinated individuals [5,6]. However, IGRA can be an expensive platform to set-up and maintain and assays require trained laboratory personnel to execute. Notwithstanding, neither test can accurately distinguish between TB infection and active TB disease [7], and current guidance from WHO is that either can be used in TBI testing and treatment algorithms. Moreover, due to access and implementation challenges of these tests, WHO recommends TPT without testing in select high-risk groups in high burden settings.

Newer skin-based tests (TBST) based on specific TB antigens are now available, these include tests such as the C-Tb (Staten Serum Institut, Denmark), DiaskinTest (Generium, Russian Federation), and C-TST (*nee* ESAT6-CFP10 or ECskintest; Anhui Zhifei Longcom, China). These specific TBST have been developed to be more accurate than the PPD-TST based on PPD and possibly offer an affordable

alternative to IGRA tests. These new tests work by using a complex of recombinant proteins in a similar way to IGRA [2] and are thus expected to be as accurate in performance. A recent systematic literature review and meta-analysis by Krutikov et al on the diagnostic performance of TBST has shown that novel skin-tests may perform similarly to IGRA and PPD-TST in different populations and settings [8]. Clinical trials have also shown that novel skin-tests have higher specificity and sensitivity for TBI, especially in resource-constrained settings and contexts where BCG vaccination is implemented routinely. [8,9]. The diagnostic accuracy of the tests alone, however, is not sufficient evidence for the recommendation of their use in TBI testing guidelines. To provide sufficient evidence for the intervention, cost-effectiveness must also be considered to ensure that the tests are affordable and feasible to implement.

The present report provides a systematic review of the literature on the cost-effectiveness of the new tests compared to currently available tests for TBI. In anticipation of a paucity of studies on the new specific TBST, we additionally undertook a systematic review on cost-effectiveness of the current test for TBI, the PPD-TST and IGRA and decided *a priori* to undertake a primary study to evaluate the cost-effectiveness of TBST vs current TBI tests by employing a Markov-chain model calibrated to different countries/contexts data. The work and report were conducted to support the GDG deliberations on the use of novel TBST.

Research question

Are novel *Mycobacterium tuberculosis* specific antigen-based skin tests for the detection of tuberculosis infection (TBST) cost-effective compared to currently available in vitro IGRA or the PPD-TST tests?

Objectives

1) To perform a systematic review of the cost-effectiveness of novel TBST compared to PPD-TST or IGRA and summarise the resource considerations and costs of implementing TBST tests.

2) To do a systematic review of literature assessing costs and cost-effectiveness of comparison tests (in anticipation of few studies in (1)) and assess incremental costs on TBST compared to PPD-TST or IGRA.
3) To model the possible cost-effectiveness of index tests vs current tests by using unit costs of index tests and their accuracy (sensitivity and specificity), to inform a Markov-chain model simulating a cohort of individuals being offered TBI testing. Includes unit costs of index tests as supplied by the manufacturer.
4) To perform univariate and probabilistic sensitivity analysis to identify cost-effectiveness thresholds in relation to index tests accuracy if costs of index tests are unavailable or subject to high levels of uncertainty.

Outcomes

Objective 1 – 2: Costs and cost-effectiveness.

Objective 3 – 4: TB cases averted or Incremental Cost-effectiveness Ratio (ICER) or Incremental Net Benefit (INT) per Quality Adjusted Life Years (QALYs) gain for novel TBST, compared to PPD-TST or IGRA tests.

LAYOUT OF THE REPORT

The report is presented in two parts.

Section A: Abstract, methods and results for Objectives 1-2

Section B: Abstract, methods and result for Objectives 3-4.

This document contains Section A. Section B is submitted as a separate attachment.

A. Systematic literature reviews

Abstract

Background: The tuberculin skin test (TST) and interferon-gamma release assays (IGRA) are the currently used tests for identifying individuals with TB infection that should be offered TB preventive treatment, however challenges around access and implementation have limited their use. Novel *Mycobacterium tuberculosis* specific skin tests (TBST) such as the DiaskinTest, ESAT6-CFP10 (now called C-TST) C-Tb, and DPPD have been developed in recent years, these may provide accurate and scalable options. We conducted a systematic review of the economic evidence of these novel tests to support WHO guideline development on tests for TB infection.

Methods: Two reviews following PRISMA guidelines were carried out to look at costs and cost-effectiveness of (1) novel TBST, the DiaskinTest, C-TST, C-Tb, DPPD (*primary review*), and (2) TST and IGRA tests (*secondary review*). We searched those articles presenting economic evaluations of the diagnostic tests (costs and cost-effectiveness) using a health provider perspective, and related to TB infection in humans. We reviewed papers written in the English, Chinese and Russian languages published in Medline, OVID, Chinese biomedical literature, China National knowledge Infrastructure, and Russian e-library databases. Quality of studies was assessed using Drummond's checklist.

Results: Papers on the economic evidence for novel TBST were limited. 8 studies were found; one in Brazil assessed cost-effectiveness of C-TST and DiaskinTest and 7 in Russian Federation assessed the DiaskinTest; none evaluated C-Tb or DPPD. In the 8 studies that assessed DiaskinTest kit, most estimated a cost of $1.6 . One study evaluated the unit costs considering staff time, consumable and laboratory costs, resulting in a cost of $5.07. This study, using the same costing factors, also evaluated C-TST unit cost estimated as $9.96. Based on Drummond's scores, the quality of the studies in this review is concerning; only one high quality study found.

We found 29 studies on the IGRA and/or the TST, which presented an average TST cost of $37.84, and $89.33 for IGRA (accounting for different ingredients). Most studies were based in high-income and

low-TB burden settings, and the cost-effectiveness of the tests varied between and within risk groups without clear economic consensus around cost-effectiveness of comparison tests. Based on Drummond's scores, the quality of these studies is generally high.

Interpretation: There is insufficient evidence regarding both the cost and cost-effectiveness of novel TBST. The quality of the studies is concerning according to the Drummond's checklist for economic evaluations. More high-quality studies are needed considering different health settings and risk-populations to estimate cost-effectiveness and understand likely economic impact.

Objectives:

1. (a) To perform a systematic review of the cost-effectiveness of novel TBST compared to PPD-TST or IGRA. (b)To summarise the resource considerations and costs of implementing TBST tests as replacement test to TST or IGRA (includes unit costs of index tests as supplied by the manufacturer). (Systematic Review 1)
2. To do a systematic review of literature assessing costs and cost-effectiveness of comparison tests (in anticipation of few studies in (1)) and assess incremental costs on TBST compared to PPD-TST or IGRA. (Systematic Review 2)

Methods

We performed two different literature reviews.

(a) a primary review looking at the costs and cost-effectiveness of the novel skin tests (registered on PROSPERO: CRD42021275585)

(b) a secondary review for the cost-effectiveness of PPD-TST and IGRAs tests (registered on PROSPERO: CRD42021275684)

Firstly, we reviewed the literature on the cost and cost-effectiveness of the novel tests. Due to the anticipated lack of evidence, a secondary review of the literature surrounding the cost and cost-effectiveness of the standard PPD-TST and IGRA was also conducted to supplement the primary review given that operational and logistic requirements for TBST and PPD-TST are same. This secondary review provides a wider scope to understand the TBI field and includes a breakdown of the unit costs of these currently used tests. We present both systematic reviews using a combined approach that is structured into four sections: **search strategy and data sources, study selection and data extraction, results, and interpretation of the findings**.

Search strategy and data sources

The search for the primary review was conducted on 1 August 2021 and for the second review on 20 July 2021. The databases Medline (OVID), Embase (OVID), Chinese biomedical literature, China National knowledge Infrastructure, and Russian e-library were used to carry out the literature searches. The search strategy for the primary and secondary reviews was to split the keywords into three key

concepts: (1) “Tuberculosis”, (2) “Diagnostic Test" and (3) “Cost-effectiveness”. The search strategies used per systematic review are shown in Table A1.1 and A1.2 (supplementary material). We also reviewed papers shared by test manufacturers and those 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).

We did not have an exclusion criterion for specific populations, and all articles had to be published in English, Chinese or Russian and relate to TBI in humans. Papers had to be full original economic evaluations, meaning having clear measures of costs, outcomes, and sufficient incremental analysis using the healthcare perspective. No time restriction was imposed for the primary review on novel skin tests, whereas we searched those articles published between 2011 to July 2021 for the secondary review on PPD-TST or IGRA tests.

Study selection & data extraction

Studies were included if they provided any economic evidence directly related to test or test implementation costs for the following products: novel TBST (Diaskintest, C-Tb, EC skin test, DPPD), and PPD-TST or IGRA (QuantiFERON®-TB Gold In-Tube /Gold Plus/ T-SPOT®.TB).

The study selection followed PRISMA guidelines [10] and the flow diagrams are presented within the results section (Figure 1-2). Conference abstracts, reviews, letters, or opinion pieces were removed. Full articles were reviewed by two reviewers after applying the inclusion/exclusion criteria. All disagreements were resolved by discussion. YH, IK and ES carried out the initial search for the primary review and removed duplicates and papers not reporting on novel skin tests, LG and FP screened abstracts and full text of papers in the English language, while IK and ES screened abstracts and full text of Russian papers. Chinese abstracts and titles were screened by two reviewers independently, relying on web-based Google translation to identify relevant studies. Full-text articles were independently reviewed by two Chinese speaking individuals (VL and TW). FP carried out the secondary review and LG and FP screened abstracts and full texts. We performed double data extraction.

Data extracted from the economic evaluations include title, author, and year of study, study population and interventions evaluated. Information about methods such as the analytical model used, time

horizon, discount rate, measure of effectiveness was also extracted. The results of the base-case analysis (incremental cost-effectiveness ratios (ICER)) alongside the "baseline" intervention used, and sensitivity analysis were recorded.

Cost components and unit test costs were extracted from the studies along with key costing input parameters. The studies mostly include test kits/drug, staff time (nurse or laboratory), consumables (syringes, gloves, etc.), and equipment (fridge storage, laboratory). Costs are presented in 2021 USD (United States Dollars).

Furthermore, all articles were assessed using Drummond's checklist for healthcare economic evaluations to assess study quality [11].

Results

Identification, screening, eligibility, and inclusion

For the primary search (systematic review 1) on cost-effectiveness of novel skin tests, 367 records were identified for full text screening (103 written in English/Chinese, and 264 in Russian) of which only 8 were relevant to the research question (only one of those written in English and 7 in Russian) (See PRISMA flow in Figure 1). For the secondary search on cost-effectiveness of PPD-TST or IGRA tests, 56 out of 407 records were chosen for full-text screening and 29 papers were selected for analysis (see PRISMA flow in Figure 2).

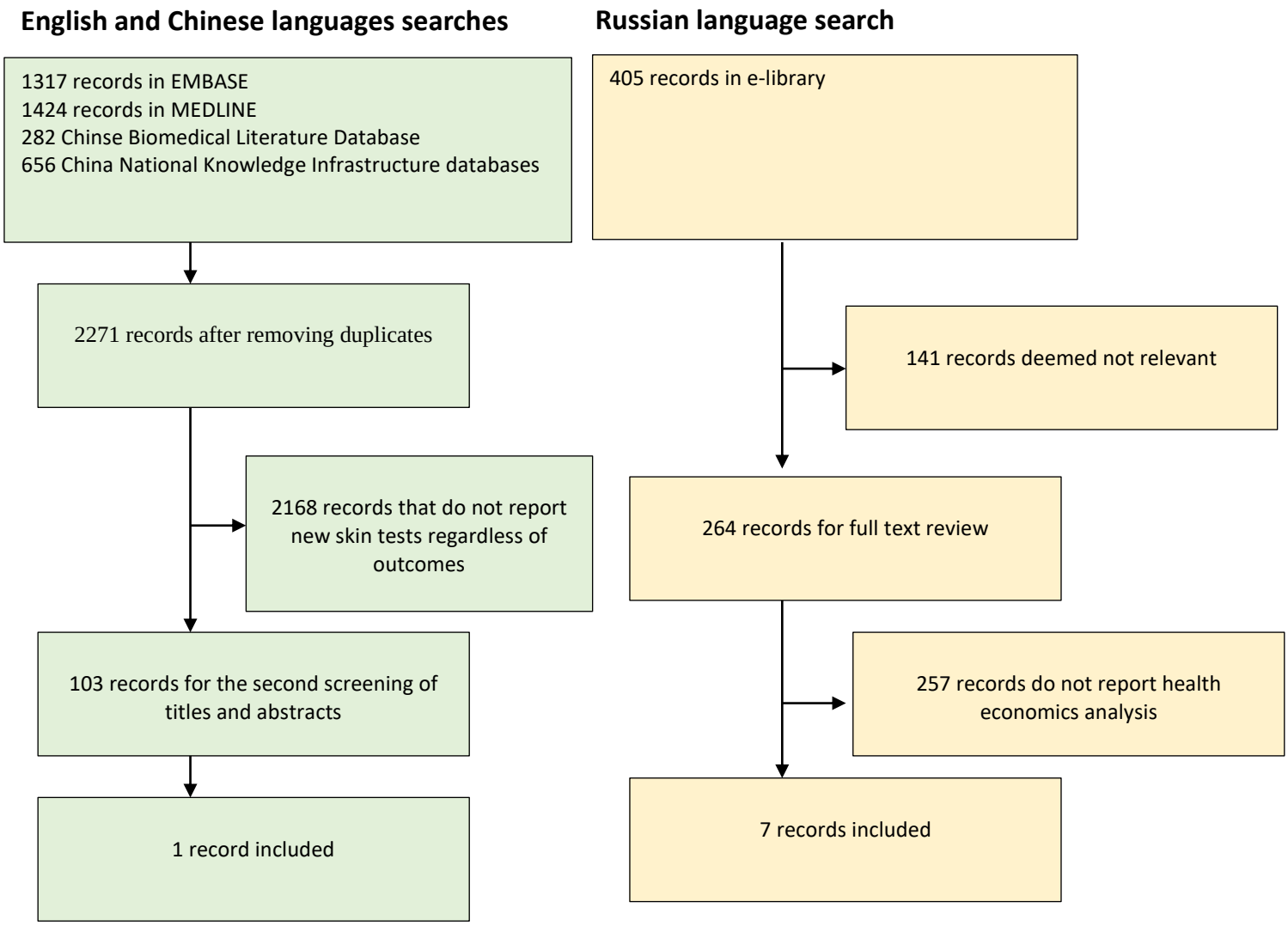

Figure 1. PRISMA flowchart for primary systematic review (Novel tests)

Note: no eligible study was identified among papers shared by test manufacturers and those identified though a WHO public call for data.

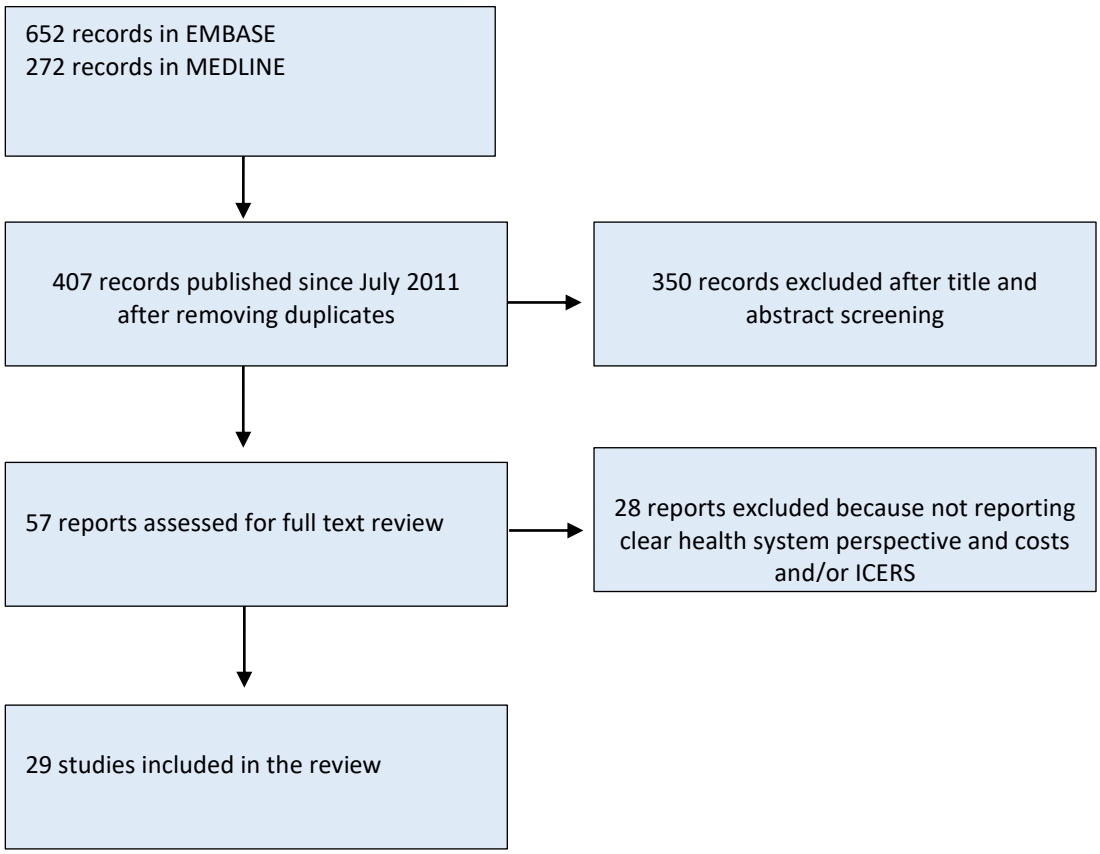

Figure 2. PRISMA flowchart for the secondary systematic review (PPD-TST and IGRA)

Description of the articles found in the primary review (specific TBST)

Eight economic evaluations were conducted from middle-income countries (Russian Federation and Brazil), these mostly compared DiaskinTest against PPD-TST (Table A2, supplementary). One study met most of the Drummond's checklist factors and reported all the required information. Steffen et al (2020) [12] studied the cost-effectiveness of two novel TBST (Diaskintest and C-TST) for people living with HIV (PLHIV) in Brazil compared to PPD-TST and QFT-GIT. A Markov model was used to compare single screening strategies of the respective tests. The primary outcome used was incremental cost-effectiveness ratio (ICER) per incremental gain in quality-adjusted life years (QALY).

The rest of the articles came from Russian Federation and were primarily focused on children using Diaskintest as an alternative strategy to PPD-TST. Most of these studies included cost-effectiveness analyses using the ICER (per case averted or diagnosed) as measure of effectiveness. The type of model used, time horizon, discounting rate, and sensitivity analyses were either uncertain or not stated in the majority of these articles which might bias the results of the studies [13-17].

We evaluated study quality using the Drummond's checklist for health economic evaluations (supplementary material, Table A7). On average, we identified that most studies did not present any sensitivity analysis, and that the information on relevant parameters such as time horizon, type of model used, year of the evaluation, and discounting rate were unclear, or not stated. Overall, study quality was concerning.

Description of the articles in the secondary review (PPD-TST and IGRA)

Of the 29 papers included in the review, 8 were based in the United Kingdom, 6 in the United States of America, 5 in Canada, 2 in Brazil, 2 in South Africa, and one in each of China, Hong Kong SAR, Japan, Norway, Oman, Singapore, and Spain as shown in Table A3 (supplementary material). The interventions studied are wide-ranging (including single and dual testing). The two primary outcome measures used in the analyses are Quality Adjusted Life Years (QALYs) and number of cases of active TB averted. Most Cost-utility analysis (CUA) and cost-effective analysis (CEA) provided incremental analysis using ICERs (i.e. cost/QALY gained). The methods ranged from discrete event simulation, Markov models, decision trees and a combination of decision tree and Markov model. Most articles included either one-way or two-way deterministic sensitivity analysis or probabilistic sensitivity analysis (PSA), with a number providing both deterministic and probabilistic analysis.

We evaluated study quality using the Drummond's checklist for health economic evaluations (supplementary material, Table A8-A9). Overall, study quality was high as most of them included and accounted for differential in time and uncertainty, more than one alternative strategy, and discussed issues of concern given the specific targeted groups explored. However, some issues are encountered for the provision of sensitivity analyses, specification of time horizons, discounting rates used, specifically for the novel skin test-related studies.

Summary of cost and cost-effectiveness findings

Systematic review on novel TBST (primary review)

One paper on the DiaskinTest and C-TST [12] was found in the English and Chinese language searches, while seven papers on DiaskinTest resulted from the Russian language searches. No papers on C-Tb test or DPPD were identified. All papers reported strategies involving DiaskinTest (and one for C-TST) as cost-effective and/or cost-saving. For unit costs, these were mostly comprised of test kits/drugs, staff time, consumables, disposables, equipment used and, less commonly, overheads. However, unit costs vary by economic evaluation and some of them provide no information on the composition. The unit cost of DiaskinTest was estimated in $5.07, whereas C-TST was $9.96 as per calculated by Steffen *et al* 2020 [12] (Table 1). DiaskinTest was preferred to QFT-GIT and PPD-TST (cost saving estimate per QALY was US $1,375). No ICER was shown for C-TST compared to DiaskinTest due to having equal effectiveness [12]. In probabilistic sensitivity analysis (PSA), Steffen compared strategies to DiaskinTest only. The dominance of DiaskinTest was very sensitive to unit costs of DiaskinTest which is highly uncertain due to using market value and hence varies widely by health system and country. The rest of the articles found that the cost of the DiaskinTest kit ranged from $1.29 to $3.49 and that it was not very sensitive to unit costs after employing univariate sensitivity analyses (if measured). All these studies, apart from Steffen *et al*, found that DiaskinTest was cost-effective using a wide-range of methods, ranging from a cost-effectiveness ratio of 2.28 times the local currency compared to 3.42 for PPD-TST to total costs saving of $757.7 for DiaskinTest compared to TST in children populations between 2009 and 2020 (Table A2). For instance, Yagudina *et al* 2013[15] found an ICER=$1,666, whereas Solodun *et al* 2017[16] found that it was $10,586.6 for DiaskinTest, being highly cost-effective, compared to TST (ICER=$49,523.9, ICER=$40,641, respectively) (Table A2 for studies details). The difference in ICERs between these two DiaskinTest studies is that Solodun *et al* 2017[16] included costs for chest radiography and additional tests in the costing scheme.

Table 1: Unit costs of novel test DiaskinTest resulted from the primary review (2021 USD)

			NOVEL TESTS COSTS						
Study ID	**Country**	**Test**	Test Kit	Staff Time	Consumable	Overheads	Laboratory	Disposable	**Unit Costs**
Aksenova (2021) [13]	Russian Federation	Diaskintest	$1.7						$1.7
Kulikov (2009) [14]	Russian Federation	Diaskintest	$1.5						$1.5
Solodun (2017) [16]	Russian Federation	Diaskintest	$1.6						$1.6
Steffen (2020) [12]	Brazil	Diaskintest: C-TST:	$1.5 $6.3	$2.24	$1.38		$0.04		$5.07 $9.96
Yagudina (2013) [15]	Russian Federation	Diaskintest:							$3.5

Notes: We only presented the total unit cost for those articles without information on costs components due to lack of evidence provided. The costs for all screening strategies include the costs of the tests (disposables, administration, reading, laboratory technicians), two clinic visits and one chest radiograph.

Systematic review on PPD-TST and IGRA (secondary review)

Most studies evaluated the costs and cost-effectiveness of PPD-TST and IGRA with wide-ranging study populations including PLHIV, immunocompromised people other than PLHIV, immigrants/migrants, healthcare workers and different methods were used, especially time horizons. 22 studies were set in low TB burden countries (United Kingdom, United States of America, Canada, Norway, Oman, Spain), 4 in lower moderate (Brazil, Japan, Singapore), 1 in upper-moderate (China, Hong Kong SAR) and two in a high burden country (South Africa). Four studies were based in low-and-middle income countries, whereas 25 in high-income countries. Most studies that used models were decision analytic, used Markov-chain techniques, and all papers carried out sensitivity analysis on model parameters. Results from the articles suggested that testing of any form (PPD-TST or IGRA) was more likely to be cost-effective when done for high-risk populations or higher burden contexts, but no consensus exists

about whether to utilise PPD-TST or IGRAs. Of the 6 studies analysing the cost-effectiveness of TB infection screening in PLHIV, all found IGRA to be more cost-effective including one PPD-TST (Tasillo et al., 2017)[18], one a combined sequential strategy of QFT+ PPD-TST (Auguste et al. 2016)[19], and one found IGRA to be cost-saving over PPD-TST but Diaskintest to be the most cost-effective overall (Steffen et al., 2020)[12]. For the remaining three, Jo *et al* 2020[20] found that an ICER of $11,000/QALY gained (New York) and as low as only $5,000/QALY (Texas) for IGRA compared to TST, whereas Capocci *et al* 2015[21] found that QFT was the most cost-effective strategy with an ICER of £9,332/QALY gained compared to no testing. Finally, Linas *et al* 2011[22] found similar results for IGRA with an ICER of $23800/QALY compared to TST.

Three studies focused on groups of immunocompromised individuals other than PLHIV, these found no testing to be the most cost-effective strategy. However, all three studies are based in low-burden countries [18,22,23]. Among healthcare workers, the five studies reported either IGRA or PPD-TST to be the most cost-effective strategy. Similarly, the five studies focusing on the screening of contacts of active TB cases, showed no consensus about whether to utilise PPD-TST or IGRA, with similar numbers being marginally more cost-effective for either one or the other alternative, or a strategy combining the two tests. All 12 studies analysing cost-effectiveness of TB infection screening in migrants in high-income countries, showed that screening with either IGRA or PPD-TST is preferred to no screening strategies. 80% of the studies comparing the two tests reported IGRA more likely to be cost-effective, one study (Abubakar et al, 2018)[24] found the combined sequential PPD-TST+QFT strategy to be the most cost-effective, and one (August et al., 2016)[19] found PPD-TST (>5mm cut off) to be the most cost-effective strategy compared to QFT-GIT. Table 2-3 presents the unit costs extracted from the secondary review by type of test. Mean PPD-TST unit cost was $37.84 and IGRA mean cost was $89.33. High variability of staff costs, especially among high-income countries, represents the major driver of heterogeneity among unit costs from different sources; some studies included more consultations/visits (provided by medical staff rather than nurses), resulting in higher unit costs. Finally, we found greater costs for the IGRA test in South Africa due to the inclusion of chest radiography within the costing scheme (Table 3).

Table 2: Unit costs of PPD-TST (2021 USD)

		PPD-TST **COSTS**						
Study ID	**Country**	Test Kit	Staff Time	Consumable	Overheads	Laboratory	Disposables	**Unit Costs**
Linas (2011) [22]	USA	$2.88	$23.92					$26.81
del Campo, (2012) [25]	Spain	✓	✓	✓				$60.35
Eralp, (2012) [a 26]	United Kingdom							$31.12
Mandalakas, (2013) [27]	South Africa	✓	✓				✓	$21.92–99.13
Kim (2018) [28]	South Africa		$1.75	$1.70	$0.08			$4.40
Pareek (2013) [b 29]	United Kingdom							$68.34
Steffen (2013) [7]	USA	$5.83	$3.79	$2.84		$0.10		$12.56
Swaminath, (2013)	USA							$48.99
Verma (2013) [30]	Canada	$16.81	$26.77					$43.58
Capocci, (2015) [c 21]	United Kingdom							$29.60
Wingate (2015) [31]	USA							USA: $28.54 Kenya (Pre-arrival): $5.35
Auguste (2016) [19]	United Kingdom	✓	✓				✓	$27.61
Nijhawan, (2016) [32]	USA	$9.38	$12.73					$22.10
Campbell, (2017).	Canada	$9.44	$17.16					$26.61
Haukaas (2017) [33]	Norway							$34.24
Mullie (2017) [34]	Canada							$13.51
Tasillo (2017) [18]	USA							$9.06
Abubakar (2018) [d 24]	United Kingdom	$1.77	$179.87 (2 clinic visits)					$181.63

Sohn (2018) [35]	Japan	✓	✓	✓	✓			$32.61
Campbell (2019a) [36]	Canada	$9.44	$17.16					$26.61
Campbell, (2019b) [23]	Canada	$9.44	$17.16					$26.61
Loureiro (2019) [37]	Brazil	$4.71	$2.41	$1.49		$0.06		$8.67
Steffen (2020) [12]	Brazil	$3.99	$2.24	$1.38		$0.04		$7.66

Notes: A tick mark *(✓) s*tands for those articles that mention they included certain cost components but did not explicitly state the figures and only included the total costs per test diagnostic. There were no further details provided but only the overall costs of the test If no tick (*✓)* is observed*.*

[a] Calculated from the National Institute for Health and Clinical Excellence. Clinical diagnosis and management of Tuberculosis, and measures for its prevention and control. NICE Clinical Guidelines 2011

b Calculated from the National Collaborating Centre for Chronic Conditions. Tuberculosis: aclinical diagnosis and management of tuberculosis, and measures for its prevention and control. LoNdon: Royal College of Physicians, 2011.

[c] Calculated from NICE. Tuberculosis - clinical diagnosis and management of tuberculosis, and measures for its prevention and control. NICE Clinical Guideline 117 2011.

[d] Staff time money values provided by the Department of Health and Social Care (DHSC). NHS Tariffs Reference Costs. London: DHSC; 2014. URL: www.gov.uk/government/collections/nhs-reference-costs .

The costs for all screening strategies include the costs of the tests (disposables, administration, reading, laboratory technicians), two clinic visits and one chest radiograph.

Table 3: Unit costs of IGRA (2021 USD)

		IGRA COSTS						
Study ID	**Country**	Test Kit	Staff Time	Consumable	Overheads	Laboratory	Disposable	**Unit Costs**
Linas (2011) [22]	USA							$62.83
Pareek, (2011) [38]	United Kingdom	✓		✓				$87.05
Del Campo (2012) [25]	Spain	✓	✓	✓		✓		**$**65.27
Eralp, (2012) [a] [26]	United Kingdom							$87.08
Shah (2012) [39]	USA	$28.44	$4.61	$6.75	$0.63	$1.36	$8.44	$50.24
Mandalakas, (2013) [27]	South Africa	✓	✓			✓	✓	T-SPOT: $247.12 QFT: $220.02
Kim (2018) [28]	South Africa		$1.63	$13.7	$0.05	$49.97		QFT: $65.35
Pareek (2013) [b] [29]	United Kingdom	✓		✓				QFT: $103.84 T-SP0T: $163.76
Steffen (2013) [7]	USA	$51.07	$1.90	$2.78		$1.63		$57.38
Swaminath, (2013) [40]	USA							$60.67
Verma (2013) [30]	Canada	$33.63	$26.77					$60.39
Capocci, (2015) [c] [21]	United Kingdom							$109.99
Auguste (2016) [19]	United Kingdom	✓	✓	✓		✓		QFT-GIT: $76.98 T-SPOT: $55.29
Nijhawan, (2016) [32]	USA	$43.36	$3.20					$46.56
Campbell, (2017) [41]	Canada	$40.34	$6.01					$46.34
Haukaas (2017) [33]	Norway							$76.27
Mullie (2017) [34]	Canada							$45.04
Tasillo (2017) [18]	USA							$97.16
Abubakar (2018) [d] [24]	United Kingdom	T-SPOT: $102.99	$89.93 (1 clinic					T-SPOT: $192.91 QFT-GIT: $149.40

		QFT-GIT: $59.47	visit)					
Li (2018) [42]	China, Hong Kong SAR							$76.10
Sohn (2018) [35]	Japan	$76.25	✓	✓	✓			$97.44
Campbell (2019a) [36]	Canada	$40.34	$6.01					$46.34
Loureiro (2019) [37]	Brazil	$38.54	$2.55	$2.06		$1.22		$44.36
Png (2019) [43]	Singapore							$81.90
Jo (2020) [20]	USA							$81.54-$92.41
Steffen (2020) [12]	Brazil	$16.77	$2.36	$1.91		$1.13		$22.17

Notes A tick mark (✓) stands for those articles that mention they included certain cost components but did not explicitly state the figures and only included the total costs per diagnostic test. There were no further details provided but only the overall costs of the test If no tick (✓) *is observed.*

[a] Calculated from the National Institute for Health and Clinical Excellence. Clinical diagnosis and management of Tuberculosis, and measures for its prevention and control. NICE Clinical Guidelines 2011

b Calculated from the National Collaborating Centre for Chronic Conditions. Tuberculosis: clinical diagnosis and management of tuberculosis, and measures for its prevention and control. London: Royal College of Physicians, 2011.

[c] Calculated from NICE. Tuberculosis - clinical diagnosis and management of tuberculosis, and measures for its prevention and control. NICE Clinical Guideline 117 2011.

[d] Staff time money values provided by the Department of Health and Social Care (DHSC). NHS Tariffs Reference Costs. London: DHSC; 2014. URL: www.gov.uk/government/collections/nhs-reference-costs .

The costs for all screening strategies include the costs of the tests (disposables, administration, reading, laboratory technicians), two clinic visits and one chest radiograph. Screening strategies that include an IGRA also include the cost of one outpatient laboratory visit.

QFT: QuantiFERON-TB Gold.

Interpretation

Considering the same factors (ingredients) for economic costing, costs of novel skin tests (DiaskinTest=$5.07, C-TST=$9.99[12]) were substantially lower than those of IGRA (average cost was $89.33; ranging from $22.17[12] to $247.12[27]) and the TST (average cost=$37.84, ranging from $4.40[28] to $181.63[24]). However, in isolation, observed low costs of DiaskinTest and C-TST compared to TST and IGRA are not sufficient evidence for likely economic impact, particularly given the limited number of studies. More high-quality studies are needed (only one high-quality study found) considering different health settings and risk-populations to estimate cost-effectiveness and understand likely impact.

Vast literature is available on TST and IGRA tests' costs and cost-effectiveness, whereas it is very limited on novel TB skin tests. Available TBST studies were based on limited settings (either Brazil or Russian Federation), most studies focused only on DiaskinTest and reported its cost-effectiveness compared to TST alone. Based on Drummond's scores, the quality of these studies is concerning. Specifically, data did not arise from randomized controlled trials, discount rates were not always used (only in 60% of all the articles), and 34% of the studies were sensitive to change in values. Moreover, sources for all values were not clearly specified (only 25% of DiaskinTest studies provided explanations). This made interpretation of study results difficult.

This review provides a basis for future cost-effectiveness analyses of novel tests by providing cost and cost-effectiveness data for TBST and for the current testing strategies, the PPD-TST and IGRA. Based on the synthesised evidence a primary modelling study that considers different populations and contexts was undertaken; results are presented in Part B.

References

1. Houben RM, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS medicine* 2016; **13**(10): e1002152.
2. World Health Organization. Global tuberculosis report 2020: World Health Organization; 2020.
3. Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? *The International Journal of Tuberculosis and Lung Disease* 2006; **10**(11): 1192-204.
4. SAGE. Report on BCG vaccine use for protection against mycobacterial infections including tuberculosis, leprosy, and other nontuberculous mycobacteria (NTM) infections, 2017.
5. Diel R, Goletti D, Ferrara G, et al. Interferon-γ release assays for the diagnosis of latent Mycobacterium tuberculosis infection: a systematic review and meta-analysis. *European Respiratory Journal* 2011; **37**(1): 88-99.
6. Diel R, Loddenkemper R, Niemann S, Meywald-Walter K, Nienhaus A. Negative and positive predictive value of a whole-blood interferon-γ release assay for developing active tuberculosis: an update. *American journal of respiratory and critical care medicine* 2011; **183**(1): 88-95.
7. Steffen RE, Caetano R, Pinto M, et al. Cost-effectiveness of Quantiferon®-TB Gold-in-Tube versus tuberculin skin testing for contact screening and treatment of latent tuberculosis infection in Brazil. *PloS one* 2013; **8**(4): e59546.
8. 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.
9. 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.
10. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews* 2015; **4**(1): 1-9.
11. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes: Oxford university press; 2015.
12. Steffen RE, Pinto M, Kritski A, Trajman A. Cost-effectiveness of newer technologies for the diagnosis of Mycobacterium tuberculosis infection in Brazilian people living with HIV. *Scientific reports* 2020; **10**(1): 1-12.
13. Аксенова ВА, Барышникова Л, Клевно НИ, et al. Новые возможности скрининга и диагностики различных проявлений туберкулезной инфекции у детей и подростков в России. *Вопросы современной педиатрии* 2011; **10**(4).
14. Куликов А, Зинчук И, Проценко М, Крысанов И. Диаскинтест для скрининга детей и подростков на туберкулезную инфекцию: подходы к ценообразованию и анализ затраты–эффективность. *Туберкулез и болезни легких* 2009; **9**: 41-6.
15. Ягудина Р, Зинчук И. Фармакоэкономическое исследование лекарственных средств для диагностики туберкулезной инфекции. *Фармакоэкономика Современная фармакоэкономика и фармакоэпидемиология* 2013; **6**(1).
16. Солодун ИЮ, Эва ХМ, Башлакова ЕЕ, Ермолаева ТН, Давыдовская МВ, Евдошенко ЕП. Клинико-экономический анализ применения метода диагностики туберкулезной инфекции у детей и подростков с использованием аллергена туберкулезного рекомбинантного. *Проблемы стандартизации в здравоохранении* 2017; (3-4).
17. Чугаев Ю, Камаева Н, Цветков А, Кудлай Д, Черняев И. ИННОВАЦИОННЫЕ РЕКОМБИНАНТНЫЕ ТЕХНОЛОГИИ ВЫЯВЛЕНИЯ И ДИАГНОСТИКИ ТУБЕРКУЛЕЗА У ДЕТЕЙ И ПОДРОСТКОВ: ДОСТИЖЕНИЯ И ПРОБЛЕМЫ. *Pediatriya named after GN Speransky* 2020; **96**(6).
18. Tasillo A, Salomon JA, Trikalinos TA, Horsburgh CR, Marks SM, Linas BP. Cost-effectiveness of testing and treatment for latent tuberculosis infection in residents born outside the United States with and without medical comorbidities in a simulation model. *JAMA internal medicine* 2017; **177**(12): 1755-64.
19. Auguste P, Tsertsvadze A, Pink J, et al. Accurate diagnosis of latent tuberculosis in children, people who are immunocompromised or at risk from immunosuppression and recent arrivals from countries with a high

incidence of tuberculosis: systematic review and economic evaluation. *Health Technology Assessment* 2016; **20**(38): 1-678.
20. Jo Y, Shrestha S, Gomes I, et al. Model-based cost-effectiveness of state-level latent tuberculosis interventions in California, Florida, New York, and Texas. *Clinical Infectious Diseases* 2021; **73**(9): e3476-e82.
21. Capocci S, Smith C, Morris S, et al. Decreasing cost effectiveness of testing for latent TB in HIV in a low TB incidence area. *European Respiratory Journal* 2015; **46**(1): 165-74.
22. Linas BP, Wong AY, Freedberg KA, Horsburgh Jr CR. Priorities for screening and treatment of latent tuberculosis infection in the United States. *American journal of respiratory and critical care medicine* 2011; **184**(5): 590-601.
23. Campbell JR, Johnston JC, Ronald LA, et al. Screening for latent tuberculosis infection in migrants with CKD: a cost-effectiveness analysis. *American Journal of Kidney Diseases* 2019; **73**(1): 39-50.
24. Abubakar I, Lalvani A, Southern J, et al. Two interferon gamma release assays for predicting active tuberculosis: the UK PREDICT TB prognostic test study. *Health technology assessment (Winchester, England)* 2018; **22**(56): 1.
25. del Campo MT, Fouad H, Solís-Bravo MM, Sánchez-Uriz MA, Mahíllo-Fernández I, Esteban J. Cost-effectiveness of different screening strategies (single or dual) for the diagnosis of tuberculosis infection in healthcare workers. *Infection Control & Hospital Epidemiology* 2012; **33**(12): 1226-34.
26. Eralp MN, Scholtes S, Martell G, Winter R, Exley AR. Screening of healthcare workers for tuberculosis: development and validation of a new health economic model to inform practice. *BMJ open* 2012; **2**(2): e000630.
27. Mandalakas AM, Hesseling AC, Gie RP, Schaaf H, Marais BJ, Sinanovic E. Modelling the cost-effectiveness of strategies to prevent tuberculosis in child contacts in a high-burden setting. *Thorax* 2013; **68**(3): 247-55.
28. Kim H, Hanrahan C, Martinson N, Golub J, Dowdy D. Cost-effectiveness of universal isoniazid preventive therapy among HIV-infected pregnant women in South Africa. *The International Journal of Tuberculosis and Lung Disease* 2018; **22**(12): 1435-42.
29. Pareek M, Bond M, Shorey J, et al. Community-based evaluation of immigrant tuberculosis screening using interferon γ release assays and tuberculin skin testing: observational study and economic analysis. *Thorax* 2013; **68**(3): 230-9.
30. Verma G, Chuck A, Jacobs P. Tuberculosis screening for long-term care: a cost-effectiveness analysis. *The International journal of tuberculosis and lung disease* 2013; **17**(9): 1170-7.
31. La'Marcus TW, Coleman MS, de la Motte Hurst C, et al. A cost-benefit analysis of a proposed overseas refugee latent tuberculosis infection screening and treatment program. *BMC public health* 2015; **15**(1): 1-14.
32. Nijhawan AE, Iroh PA, Brown LS, Winetsky D, Porsa E. Cost analysis of tuberculin skin test and the QuantiFERON-TB Gold In-tube test for tuberculosis screening in a correctional setting in Dallas, Texas, USA. *BMC infectious diseases* 2016; **16**(1): 1-11.
33. Haukaas FS, Arnesen TM, Winje BA, Aas E. Immigrant screening for latent tuberculosis in Norway: a cost-effectiveness analysis. *The European Journal of Health Economics* 2017: 405-15.
34. Mullie GA, Schwartzman K, Zwerling A, N'Diaye DS. Revisiting annual screening for latent tuberculosis infection in healthcare workers: a cost-effectiveness analysis. *BMC medicine* 2017; **15**(1): 1-15.
35. Sohn H, Kim H, Lee S. Cost-effectiveness of contact screening strategies for tuberculosis among high-school adolescents in South Korea. *The International Journal of Tuberculosis and Lung Disease* 2018; **22**(5): 496-503.
36. Campbell JR, Johnston JC, Cook VJ, Sadatsafavi M, Elwood RK, Marra F. Cost-effectiveness of latent tuberculosis infection screening before immigration to low-incidence countries. *Emerging infectious diseases* 2019; **25**(4): 661.
37. Loureiro RB, Maciel ELN, Caetano R, et al. Cost-effectiveness of QuantiFERON-TB Gold In-Tube versus tuberculin skin test for diagnosis and treatment of Latent Tuberculosis Infection in primary health care workers in Brazil. *PloS one* 2019; **14**(11): e0225197.
38. Pareek M, Watson JP, Ormerod LP, et al. Screening of immigrants in the UK for imported latent tuberculosis: a multicentre cohort study and cost-effectiveness analysis. *The Lancet infectious diseases* 2011; **11**(6): 435-44.
39. Shah M, Miele K, Choi H, et al. QuantiFERON-TB gold in-tube implementation for latent tuberculosis diagnosis in a public health clinic: a cost-effectiveness analysis. *BMC infectious diseases* 2012; **12**(1): 1-10.

40. Swaminath A, Bhadelia N, Wang YC. Cost-effectiveness of QuantiFERON testing before initiation of biological therapy in inflammatory bowel disease. *Inflammatory bowel diseases* 2013; **19**(11): 2444-9.
41. Campbell JR, Johnston JC, Sadatsafavi M, Cook VJ, Elwood RK, Marra F. Cost-effectiveness of post-landing latent tuberculosis infection control strategies in new migrants to Canada. *PloS one* 2017; **12**(10): e0186778.
42. Li J, Yip BH, Leung C, et al. Screening for latent and active tuberculosis infection in the elderly at admission to residential care homes: a cost-effectiveness analysis in an intermediate disease burden area. *PloS one* 2018; **13**(1): e0189531.
43. Png ME, Yoong J, Ong CWM, Fisher D, Bagdasarian N. A screening strategy for latent tuberculosis in healthcare workers: Cost-effectiveness and budget impact of universal versus targeted screening. *Infection Control & Hospital Epidemiology* 2019; **40**(3): 341-9.
44. Al Abri S, Kowada A, Yaqoubi F, Al Khalili S, Ndunda N, Petersen E. Cost-effectiveness of IGRA/QFT-Plus for TB screening of migrants in Oman. *International Journal of Infectious Diseases* 2020; **92**: S72-S7.

Supplementary documentation

Section A: Systematic literature reviews

I. Literature review

For the primary systematic literature review (new skin tests), the search included papers from inception until 30 July 2021. Since the previous review included papers published until 20 October 2020, our updated search excluded studies published before October 2020.

Table A1.1. Search strategy for the primary systematic literature review (nov

	Search Term
1	exp TUBERCULOSIS/ or tuberculosis.mp. or exp *MYCOBACTERIUM TUBERCULOSIS*/ or tb.mp.
2	exp Recombinant Proteins/ or (recombinant or novel or dppd or esat 6 or esat6 or cfp 10 or cfp10 or early secretory antigenic target* or culture filtrate protein* or rd* or region of difference or Rv0061 or recombinant tuberculosis allergen).mp.
3	skin test*.mp. or Skin Tests/
4	(c tb or diaskintest or c-tst or dppd).mp.
5	1 and 2 and 3
6	(recombinant and allergen).mp.
7	6 and 1
8	4 or 5 or 7

Other databases:
e-library (www.e-library.ru) for Russian literature
Search Terms: Диаскинтест* или «Аллерген* туберкулезн* рекомбинантн*»
The Chinese Biomedical Literature Database and the China National Knowledge Infrastructure databases
Search terms: “ESAT-6” and “CFP-10” or “ESAT6” and “CFP10”

For the secondary systematic review, we included all the articles looking at TST and IGRA cost-effectiveness since 2011.

Table A1.2 Search strategy for the secondary systematic literature review (TST and IGRA)

	Search Term
1	exp TUBERCULOSIS/ or tuberculosis.mp. or exp *MYCOBACTERIUM TUBERCULOSIS*/ or tb.mp.
2	(TST or Tuberculin Skin Test or Tuberculin Test* or IGRA* or Interferon gamma release assay* or Interferon gamma release test* or QFT or QFT-GIT or QuantiFERON-TB Gold In-tube or T-SPOT).mp.
3	(TST or Tuberculin Skin Test or Tuberculin Test* or IGRA* or Interferon gamma release assay* or Interferon gamma release test*

Table A2: Data extraction results for all the articles found in the primary systematic review (N=8)

Study Characteristics	Kulikov 2009 [14]	Aksenova 2011 [13]	Yagudina 2013 [15]
Country Setting	Russian Federation	Russian Federation	Russian Federation
Year of Cost Valuation	Not Stated	2010 (Assumption)	Not stated.
Currency	Rubles	Rubles	Rubles
Study Population	Children and Adolescents	Children	Children and Adolescents
Index Diagnostic Test Strategies	DiaskinTest	DiaskinTest	DiaskinTest TST + DiaskinTest
Alternative Provided	TST	TST	TST
Type of economic evaluation	CEA	Cost Analysis	CEA
Source of costing	Studies varying DiaskinTest costs in relation to ICER of Mantoux.	Empirical Data Collection	Published Literature
Primary outcome	ICER: RUB/additional TB case diagnosed	Retrospective costing.	ICER: RUB/active TB case averted.
Type of model	Decision Tree	N/A	Decision Tree
Time Horizon	Not Stated.	N/A	Not stated.
Discounting	Not Stated.	N/A	Not stated.
Sensitivity analysis	Univariate	N/A	One-way.
Key scenarios/variables in sensitivity analysis.	Costs of tests and treatments.	N/A	Unit costs of DiaskinTest and TST.
WTP Threshold	Not Stated.	N/A	Not Stated.
Unit Costs	Breakdown of TST: Syringe 1.46 RUB, TST test and readout 26.14 RUB, one 0.2 ml dose of TST 1.5 RUB. DiaskinTest: 256.71	Strategy/Examination TST: 107926 / 132 DiaskinTest: 52128 / 64	Strategy: DiaskinTest: 380.36 TST: 218.81 TST+ DiaskinTest: 220.46 Unit Cost of Test: DiaskinTest: 148 TST: 104.7

Continuation of Table A2

Study Characteristics	Moiseeva 2014	Solodun 2017 [16]	Sinitsyn 2018
Country Setting	Russian Federation	Russian Federation	Russian Federation
Year of Cost Valuation	2013	Not stated.	2015
Currency	Rubles	Rubles	Rubles
Study Population	Children	Children and Adolescents	People living with HIV
Index Diagnostic Test Strategies	DiaskinTest 1. TST+ DiaskinTest	1) DiaskinTest 2) TST + DiaskinTest	DiaskinTest
Alternative Provided	TST	TST	Do nothing.
Type of economic evaluation	CEA/CUA Depends on effectiveness measure.	CEA	CBA
Source of costing	In accordance with the price list of GBUZ IC "KKPTD" as of 1.09.2013	State Register and Moscow Centre for Tuberculosis.	Empirical Data Collection
Primary outcome	ICER: RUB/case of active TB	ICER: RUB/case of active TB identified.	Net savings
Type of model	Unclear	Decision Tree	N/A
Time Horizon	Not Stated.	Not stated.	Not Stated.
Discounting	Not Stated.	Not stated.	Not Stated.
Sensitivity analysis	None carried out.	One-way.	None carried out.
Key scenarios/variables in sensitivity analysis.	N/A	Cost of tests and treatment, sensitivity, and specificity of tests.	N/A
WTP Threshold	Not stated.	Not stated.	Not stated.
Unit Costs	TST: 85 74 RUB DiaskinTest: 118.48 RUB	Strategy (RUB/100 diagnoses): TST: 18,555.18 TST+ DiaskinTest: 16,311.93 DiaskinTest: 14,811.92 Unit Cost of Tests: DiaskinTest: 95.04 TST: 90.00	Test/Strategy: DiaskinTest: 720 RUB/2363.26

Continuation of Table A2

Study Characteristics	Chugaev 2020 [17]	Steffen (2020) [12]
Country Setting	Russian Federation	Brazil
Year of Cost Valuation	2013 for TST 2019 for DiaskinTest	2020
Currency	Rubles	USD
Study Population	Children	Adults living with HIV
Index Diagnostic Test Strategies	DiaskinTest	1. TST, 2. QFT-GIT, e. EC test
Alternative Provided	TST	DiaskinTest
Type of economic evaluation	Cost analysis	CUA
Source of costing	Retrospective cohort study.	Ministry of Health and
Primary outcome	Cost of strategy and cost of additional diagnosis of TB.	ICER: $/QALY gained
Type of model	N/A	Markov.
Time Horizon	N/A	20 years.
Discounting	N/A	5%
Sensitivity analysis	None.	One-way, Two-way and PSA
Key scenarios/variables in sensitivity analysis.	Not stated.	TST and DiaskinTest sensitivity and specificity, prevalence LTBI
WTP Threshold	Not Stated.	$7544
Unit Costs	No breakdown provided.	Yes

Table A3: Data extraction results for all the entries from the secondary systematic review

Study Characteristics	Linas (2011) [22]	Pareek (2011)[38]	del Campo (2012) [25]
Country Setting	USA	United Kingdom	Spain
Year of Cost Valuation	2011	2010	2012
Currency	USD	GBP	Euros
Study Population	1) Immigrant/Migrants 2) Immunocompromised 3) Vulnerable	Recently arrived immigrants (<16y.o. and 16-35y.o.)	Healthcare workers.
Index Diagnostic Test Strategies	1. QFT 2. TST	IGRA	1. TST (10mm) 2. QFT 3. TST (5mm) + QFT
Alternatives Provided	No screening.	No screening	TST (5mm)
Type of economic evaluation	CUA	CEA	CEA
Source ofcosting	Published Literature	Empirical data collection	Published Literature (tests) and Empirical data collection for other costs.
Primary outcome	ICER: $/QALYs gained	ICER: £/active TB case averted.	ICER: Euro/active TB case averted.
Type of model	Markov.	Decision tree.	Decision Tree
Time Horizon	Lifetime.	20 years.	2 years.
Discounting	1.5%	3.5%	Not stated.
Sensitivity analysis	One-way, Two-way deterministic.	One-way deterministic.	One-way deterministic.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity, specificity, and test costs.	Reactivation Rate, sensitivity, and specificity of tests.	TST and QFT sensitivity, specificity and LTBI prevalence.
WTP Threshold	$100,000	Not Stated.	Not Stated.

Continuation of Table A3

Study Characteristics	Eralp. (2012) [26]	Shah (2012) [39]	Mandalakas (2013) [27]
Country Setting	United Kingdom	USA	South Africa
YearofCostValuation	2011	2012	2009
Currency	GBP	USD	USD
Study Population	Healthcare Workers	Individuals at primary health clinic with positive TST.	Children (0-2 and 3-5y.o.) contacts.
Index Diagnostic Test Strategies	1. QFT 2. TST	TST + QFT-GIT	1) TST 2) TST+QFT. 3) TST-QFT. 4) TST+ T-SPOT 5) TST- T-SPOT 6) QFT. 7) T-SPOT
Alternative Provided	TST + QFT	TST	No Screening
Type of economic evaluation	CEA	CEA	CEA
Source ofcosting	Published Literature	Data from Boston Health Dep.	Published Literature.
Primary outcome	ICER: £/Life Year Gained.	ICER: $/QALY gained.	ICER: $/active TB case averted.
Type of model	Markov.	Decision tree.	Markov
Time Horizon	20 years.	1 year for those without LTBI and lifetime for those with LTBI.	15 years.
Discounting	5%	3%	3%
Sensitivity analysis	One-way deterministic and PSA.	Two-way and PSA.	One-way deterministic.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.	QFT-GIT sensitivity, specificity and LTBI prevalence.	TST and QFT sensitivity, specificity and LTBI prevalence.
WTP Threshold	£30,000	$50,000	Not Stated.

Continuation of Table A3

Study Characteristics	Pareek (2013) [29]	Steffen (2013) [7]	Swaminath (2013) [40]
Country Setting	United Kingdom	USA	USA
Year of Cost Valuation	2011	2012	Not stated.
Currency	GBP	USD	USD
Study Population	Recently arrived immigrants (≤35 y.o.).	35y.o. close contacts.	Immunosuppressed with IBD
Index Diagnostic Test Strategies	1. T-SPOT 2. QFT 3. TST + T-SPOT 4. TST + QFT	1. QFT 2. TST + QFT	QFT
Alternative Provided	TST	TST	TST
Type of economic evaluation	CEA	CEA	CEA
Source ofcosting	Published Literature	Published Literature.	Published Literature.
Primary outcome	ICER: £/active TB case averted.	ICER: $/active TB case averted.	Cost and TB deaths.
Type of model	Decision Tree	Decision tree.	Decision Tree
Time Horizon	20 years.	2 years.	1 year.
Discounting	3.5%	None.	None.
Sensitivity analysis	One-way deterministic.	One-way and two-way and PSA.	One-way deterministic.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.	TST and QFT sensitivity, specificity, test, and treatment. costs and LTBI prevalence.
WTP Threshold	Not Stated.	$50,000	Not Stated.

Continuation of Table A3

Study Characteristics	Verma (2013) [30]	Capocci, (2015) [21]	Wingate (2015) [31]
Country Setting	Canada	United Kingdom	USA
Year of Cost Valuation	2012	2012	2012
Currency	Canadian Dollars	Euro	USD
Study Population	>65y.o. in long term care	Adults living with HIV	Pre arrival refugees to USA.
Index Diagnostic Test Strategies	TST	1. QFT + TST 2. QFT (for higher risk) 3. QFT (for all)	TST
Alternative Provided	No screening	No Screening	No Screening
Type of economic evaluation	CEA	CEA/CUA	CBA
Source of costing	Published Literature	Published Literature and clinic data.	Published Literature and experts.
Primary outcome	ICER: $/active TB case averted.	ICER: Euro/QALY gained or active TB case averted.	Net benefit (Cost)
Type of model	Markov	Markov.	Decision tree and Markov.
Time Horizon	4 years	Lifetime	20 years.
Discounting	3%	3.5%	3%
Sensitivity analysis	One-way deterministic.	PSA.	One-way deterministic.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity, specificity and LTBI prevalence.	TST and QFT sensitivity, specificity, test, and treatment. costs and LTBI prevalence.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.
WTP Threshold	Not Stated.	Euro 24,000	Not Stated.

Continuation of Table A3

Study Characteristics	Auguste. (2016)[19]	Nijhawan (2016) [32]	Campbell (2017) [41]
Country Setting	United Kingdom	USA	Canada
Year of Cost Valuation	2012	2013	2016
Currency	Euro	USD	Canadian Dollars
Study Population	1) Children 2) Immunocompromised 3) Recently Arrived Immigrants 4) General Population.	Adults entering jail.	Pre-arrival refugees to USA.
Index Diagnostic Test Strategies	1. IGRA 2. TST + IGRA 3. Simultaneous.	QFT	1. TST 2. IGRA 3. TST + IGRA
Alternative Provided	TST	TST	No Screening
Type of economic evaluation	CUA	Cost Analysis.	CUA
Source of costing	Published Literature, NHS data and assumptions.	Published Literature and empirical data collection.	British Columbia Centre for Disease control
Primary outcome	ICER: £/QALY gained	Cost difference per active TB case detected.	ICER: $/QALY gained.
Type of model	Decision tree and discrete event simulation.	Decision tree.	Discrete event simulation.
Time Horizon	Lifetime	Not Stated.	10 years.
Discounting	3.5%	Not Stated.	1.5%
Sensitivity analysis	PSA.	One-way.	One-way and PSA.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity, specificity, test, and treatment. costs and LTBI prevalence.	Cost of labour, unit cost of tests.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.
WTP Threshold	£30,000	Not Stated.	$100,000

Continuation of Table A3

Study Characteristics	Haukaas (2017) [33]	Mullie (2017) [34]	Tasillo, (2017) [18]
Country Setting	Norway	Canada	USA
Year of Cost Valuation	2013	2015	2015
Currency	Euro	Canadian Dollars	USD
Study Population	Recently arrived immigrants <35y.o.	HCW with negative TST at time of employment.	US-born or migrants living with or without comorbidities.
Index Diagnostic Test Strategies	1. QFT (for those with risk factors) 2. TST + QFT 3. QFT (for all)	QFT-GIT	1. TST 2. IGRA 3. IGRA + TST 4. IGRA - TST
Alternative Provided	No Screening	TST	No Screening
Type of economic evaluation	CEA	CUA	CUA
Source of costing	Published Literature and expert opinion.	Published Literature.	Assumptions and published literature.
Primary outcome	ICER: £/QALY gained	ICER: $/QALY gained.	ICER: $/QALY gained.
Type of model	Decision Tree and Markov	Decision tree.	Decision tree and Markov.
Time Horizon	10 years.	20 years.	Lifetime
Discounting	4%	3%	3%
Sensitivity analysis	One-way.	One-way and two scenario analyses.	One-way and PSA.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity, specificity, test, costs and LTBI prevalence.	TST and QFT sensitivities and specificities.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.
WTP Threshold	28,400 Euros	Not Stated.	$100,000

Continuation of Table A3

Study Characteristics	Abubakar (2018) [24]	Li (2018) [42]	Sohn (2018) [35]
Country Setting	United Kingdom	China, Hong Kong SAR	Japan
Year of Cost Valuation	Not Stated.	Not stated.	2015
Currency	GBP	USD	USD
Study Population	1. Recent immigrants. 2. Contacts	Elderly (>65y.o.) at admission to residential care home.	Adolescents (13-18y.o.) contacts.
Index Diagnostic Test Strategies	1. T-SPOT.TB 2. QFT-GIT 3. TST (varying cut offs) 4. Confirm positive or negative T-SPOT or QFT-GIT after TST	IGRA	1. TST + QFT 2. QFT
Alternative Provided	No Screening	No screening.	TST
Type of economic evaluation	CUA	CUA	CEA
Source of costing	Published literature and NHS Data.	Estimation.	Published literature.
Primary outcome	ICER: £/QALY gained.	ICER: $/QALY gained.	ICER: $/QALY gained.
Type of model	Decision Tree.	Markov.	Decision tree.
Time Horizon	Lifetime	20 years.	2 years.
Discounting	3.5%	5%	3% (overhead costs only)
Sensitivity analysis	PSA.	One-way and PSA.	One-way and PSA.
Key scenarios/variables in sensitivity analysis.	Test sensitivity and specificity.	TST and QFT sensitivities and specificities, reactivation rate.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.
WTP Threshold	£20,000	$50,000	$50,000

Continuation of Table A3

Study Characteristics	Campbell (2019a) [36]	Campbell (2019b) [23]	Loureiro (2019) [37]
Country Setting	Canada	Canada	Brazil
Year of Cost Valuation	2016	2016	2016
Currency	Canadian Dollars	Canadian Dollars	USD
Study Population	Pre-arrival immigrants.	Migrants with either late-stage CKD or beginning dialysis.	Primary HCW.
Index Diagnostic Test Strategies	1. QFT-GIT 2. TST + QFT-GIT 3. TST (>10mm)	1. QFT-GIT 2. TST (>10mm)	1. TST (>10mm) 2. TST (>10mm) + QFT 3. TST (>5mm) + QFT QFT
Alternative Provided (Baseline for Incremental Analysis)	No screening	No screening	TST (>5mm)
Type of economic evaluation	CUA	CUA	CEA
Source of costing	British Columbia Centre for Disease control and expert opinion.	British Columbia Centre for Disease control and expert opinion.	Ministry of Health and estimations.
Primary outcome	ICER: $/QALY gained	ICER: $/QALY gained.	ICER: $/active TB case averted.
Type of model	Discrete event simulation	Markov.	Decision tree.
Time Horizon	25 years.	25 years.	1 year.
Discounting	3%	3%	None.
Sensitivity analysis	PSA.	PSA.	One-way.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity and specificity, incidence rate of country of origin of immigrant, reactivation rate.	TST and QFT sensitivities and specificities, cost of tests and reactivation rate.	TST and QFT sensitivity, specificity, test costs and LTBI prevalence.
WTP Threshold	$50,000	$50,000	Not stated.

Continuation of Table A3

Study Characteristics	Png (2019) [43]	Al Abri (2020) [44]	Jo (2020) [20]
Country Setting	Singapore	Oman	USA
Year of Cost Valuation	2016	2020	2018
Currency	USD	USD	USD
Study Population	HCW.	20-year-old recent immigrants.	(1) non–US-born, (2) living with diabetes, (3) HIV-positive, (4) experiencing recent homelessness, (5) incarcerated
Index Diagnostic Test Strategies	QFT-GIT either annually or every three years for combination of 1) new hires 2) high risk 3) international 4) universal.	1. QFT 2. TST 3. CXR All with varying treatments.	IGRA
Alternative Provided (Baseline for Incremental Analysis)	No Screening	QFT (4-month RIF)	Not Stated.
Type of economic evaluation	CUA	CUA	CUA
Source of costing	National University Hospital and Published Literature.	Not Stated.	Published Literature.
Primary outcome	ICER: $/QALY gained	ICER: $/QALY gained	ICER: $/QALY gained.
Type of model	Decision Tree.	Markov.	Individual-based TB epidemiological model.
Time Horizon	3 years.	Lifetime	30 years.
Discounting	3%	Not stated.	3%
Sensitivity analysis	PSA.	One-way, Two-way and PSA.	One-way and PSA.
Key scenarios/variables in sensitivity analysis.	TST and QFT sensitivity and specificity, incidence rate of country of origin of immigrant, reactivation rate.	TST and QFT sensitivity and specificity, incidence rate of country of origin of immigrant, reactivation rate.	Cost of tests, treatment, and completion of treatment probability.
WTP Threshold	$50,000	$100,000	Not stated.

Continuation of Table A3

Study Characteristics	Steffen (2020) [12]	Kim (2018) [28]
Country Setting	Brazil	South Africa
Year of Cost Valuation	2020	2016
Currency	USD	USD
Study Population	Adults living with HIV	HIV+ pregnant women
Index Diagnostic Test Strategies	1. TST 2. QFT-GIT 3. EC Test	1. TST
Alternative Provided (Baseline for Incremental Analysis)	DiaskinTest	QFT-GIT
Type of economic evaluation	CUA	CEA
Source of costing	Ministry of Health and market value.	National Health Laboratory service
Primary outcome	ICER: $/QALY gained	ICER: $/DALY averted
Type of model	Markov.	Decision tree
Time Horizon	20 years.	12 months
Discounting	5%	3%
Sensitivity analysis	One-way, Two-way and PSA.	One-way, Two-way and PSA.
Key scenarios/variables in sensitivity analysis.	TST and DiaskinTest sensitivity and specificity, prevalence of LTBI.	TST and QFT sensitivity and specificity, and other highly sensible parameters
WTP Threshold	$7544	$12,860/DALY

Table A7. Drummond checklist for studies quality: Cost/Cost-effectiveness analyses for Novel skin tests for diagnosing TBI

Drummond Checklist Questions	Kulikov 2009	Aksenova 2011	Yagudina 2013
1. Was a well-defined question posed in answerable form?			
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	No, only presents costs	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes, children and adolescents	Yes, children	Yes
2. Was a comprehensive description of the competing alternatives given?			
2.1. Were there any important alternatives omitted?	No	No	No
2.2. Was (should) a do-nothing alternative be considered?	No	No	No
3. Was the effectiveness of the programme or services established?			
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No

Drummond Checklist Questions	Kulikov 2009	Aksenova 2011	Yagudina 2013
3.2. Was effectiveness established through an overview of clinical studies?	Yes. Sensitivity and specificity of tests values from number of clinical studies.	No, cost analysis, no CE	Yes
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes. Observational data, no time horizon/discounting provided.	N/A	Assumption on the proportion of patients with a dubious and positive test with 2TE PPD-L and Diaskintest drug® to be equal to two different existing artiles, which might lead to biased conclusion depending on the settings analysed and population characteristics
4. Were all the important and relevant costs and consequences for each alternative identified?			
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes
4.3. Were the capital costs, as well as operating costs, included?	Test and treatment costs included, the rest is unclear	Only research costs, sample analyses costs, preventive treatment, consultation and Xrays	Yes, it included treatment, diagnostic, drugs and chemoteraphy costs, registering and operating costs, among others

Drummond Checklist Questions	Kulikov 2009	Aksenova 2011	Yagudina 2013
5. Were costs and consequences measured accurately in appropriate physical units?			
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	Outcome costs is missing	The cost of test itself, it just measures the cost of diagnostic measures. And outcome costs	No
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No
6. Were the cost and consequences valued credibly?			
6.1. Were the sources of all values clearly identified?	Unclear, cost of valuation is not stated	Unclear, but only stated on the Table 3	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	Yes	Yes	Yes
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A
6.4. Was the valuation of consequences appropriate for the question posed?	Yes	N/A	Yes

Drummond Checklist Questions	Kulikov 2009	Aksenova 2011	Yagudina 2013
7. Were costs and consequences adjusted for differential timing?			
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	No	N/A	Not stated
7.2. Was there any justification given for the discount rate used?	Discounting rate No stated	N/A	N/A
8. Was an incremental analysis of costs and consequences of alternatives performed?			
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	No	Yes
9. Was allowance made for uncertainty in the estimates of costs and consequences?			
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	Yes	N/A	Yes
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	Yes, for treatment and test costs	N/A	No explanation, just stated the parameters over which the SA was employed (test cost)

Drummond Checklist Questions	Kulikov 2009	Aksenova 2011	Yagudina 2013
9.3. Were the study results sensitive to changes in the values?	Not that much. The conclusion was robust to changes in key parameters with the cost of a second clinical visit being the most influential to the cost-effectiveness ratios	N/A	No, it was stable
10. Did the presentation and discussion of study results include all issues of concern to users?			
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	Yes (2.28 rubles compared to 3.42)		Yes
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	No	No	No
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	No	Yes	No

Drummond Checklist Questions	Kulikov 2009	Aksenova 2011	Yagudina 2013
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	No	Yes	Yes
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	Yes	Yes	No

Continuation of Table A7

Drummond Checklist Questions	Moiseeva 2014	Solodun 2017	Sinitsyn 2018	Chugaev 2020
1. Was a well-defined question posed in answerable form?				
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes	Yes, children and adolescents	Yes, HIV patients	Yes, children
2. Was a comprehensive description of the competing alternatives given?				
2.1. Were there any important alternatives omitted?	No	No	Yes, the use of any other test rather than Diaskin solely	Yes
2.2. Was (should) a do-nothing alternative be considered?	No	No	Yes	No
3. Was the effectiveness of the programme or services established?				
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No	No
3.2. Was effectiveness established through an overview of clinical studies?	Yes	Yes	Yes, CB analysis	Yes

Drummond Checklist Questions	Moiseeva 2014	Solodun 2017	Sinitsyn 2018	Chugaev 2020
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Not clear CE method used	Yes, it uses observational data from specific settings	Yes, ICER is not computed and no other interventions were employed rather than do-nothing and Diaskintest	
4. Were all the important and relevant costs and consequences for each alternative identified?				
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes	Yes
4.3. Were the capital costs, as well as operating costs, included?	unclear	Yes, including staff costs, diagnostics, tests, operating costs, etc	No	No
5. Were costs and consequences measured accurately in appropriate physical units?				
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	No	No	No	No
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No	No

Drummond Checklist Questions	Moiseeva 2014	Solodun 2017	Sinitsyn 2018	Chugaev 2020
6. Were the cost and consequences valued credibly?				
6.1. Were the sources of all values clearly identified?	Unclear	Yes, most costs coming from the State Treasure Healthcare institution	Yes	
6.2. Were market values employed for changes involving resources gained or depleted?	Yes	Yes	Yes	Yes
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A	N/A
6.4. Was the valuation of consequences appropriate for the question posed?	Yes	Yes	Yes	Yes
7. Were costs and consequences adjusted for differential timing?				
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	No	Not stated	Not stated	No
7.2. Was there any justification given for the discount rate used?	N/A	N/A	N/A	N/A

Drummond Checklist Questions	Moiseeva 2014	Solodun 2017	Sinitsyn 2018	Chugaev 2020
8. Was an incremental analysis of costs and consequences of alternatives performed?				
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Non ICER computed, comparison between two alternatives	Yes Diaskin over TST, costs saved
9. Was allowance made for uncertainty in the estimates of costs and consequences?				
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	No	N/A	N/A	N/A
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	No sensitivity analyses carried out	Yes	Non sensitivity analyses employed	Non sensitivity analyses employed
9.3. Were the study results sensitive to changes in the values?	N/A	No, 1% change	N/A	N/A
10. Did the presentation and discussion of study results include all issues of concern to users?				
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	No	Not stated	No	No

Drummond Checklist Questions	Moiseeva 2014	Solodun 2017	Sinitsyn 2018	Chugaev 2020
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	No	No	No	No
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	No	No	No	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	No	No	No	No
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	No	No	No	Yes

Table A8. Drummond Checklist for studies quality: Cost/Cost-effectiveness analyses for TST or IGRA for diagnosing TBI

Drummond Checklist Questions	Linas (2011)	Pareek (2011)	del Campo (2012)	Eralp (2012)	Shah (2012)
1. Was a well-defined question posed in answerable form?					
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes Immigrant/migrants, immunocompromised and vulnerable populations in the USA.	Yes Recently arrived immigrants (<16y.o. and 16-35y.o.) in the United Kingdom.	Yes Healthcare workers in Spain.	Yes Healthcare workers in the United Kingdom.	Yes Individuals at primary health clinic with positive TST in the USA.
2. Was a comprehensive description of the competing alternatives given?					
2.1. Were there any important alternatives omitted?	No	No	No	No	No
2.2. Was (should) a do-nothing alternative be considered?	Yes, do nothing was included appropriately.	Yes, do nothing was included appropriately.	No	No	No
3. Was the effectiveness of the programme or services established?					
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No	No	No

Drummond Checklist Questions	Linas (2011)	Pareek (2011)	del Campo (2012)	Eralp (2012)	Shah (2012)
3.2. Was effectiveness established through an overview of clinical studies?	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes. Sensitivity and specificity of tests values from number of clinical studies.
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes. Assumed that quality of life with cured TB was the same as that for healthy individuals. All assumptions references/reasoned adequately.	Yes. Prospective cohort analysis performed for LTBI prevalence. All assumptions references/reasoned adequately.	Yes. Key assumption that no active cases of TB at the time of testing. All assumptions references/reasoned adequately.	Yes Key assumption that "LTBI generates a positive result at same probability that active TB" for test. All effectiveness assumptions clearly identified and references/ reasoned adequately.	Yes All effectiveness assumptions clearly identified and references/ reasoned adequately.
4. Were all the important and relevant costs and consequences for each alternative identified?					
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Linas (2011)	Pareek (2011)	del Campo (2012)	Eralp (2012)	Shah (2012)
4.3. Were the capital costs, as well as operating costs, included?	No. State only direct medical costs included.	No.	No.	Yes, stated overheads included but no description provided.	Yes, operating costs such as quality assurance, specimen transport, supply delivery, and estimates for rent and utilities included.
5. Were costs and consequences measured accurately in appropriate physical units?					
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	No	Yes. Excluded drug-resistant strains and HIV infection. Assumed minimal impact.	No	No	No
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No	No	No
6. Were the cost and consequences valued credibly?					
6.1. Were the sources of all values clearly identified?	Yes	Yes	Yes	Yes	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Linas (2011)	Pareek (2011)	del Campo (2012)	Eralp (2012)	Shah (2012)
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A	N/A	N/A
6.4. Was the valuation of consequences appropriate for the question posed?	Yes	Yes	Yes	Yes	Yes
7. Were costs and consequences adjusted for differential timing?					
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	Yes, 3%.	Yes, 3.5%.	No.	Yes, 5%.	Yes, 3%.
7.2. Was there any justification given for the discount rate used?	Yes, following Siegel et al. 1997 guidelines.	Yes, following NICE recommendations.	N/A	State "standard rate" with no reference in supplementary material.	No justification for discount rate of costs provided.
8. Was an incremental analysis of costs and consequences of alternatives performed?					
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Linas (2011)	Pareek (2011)	del Campo (2012)	Eralp (2012)	Shah (2012)
9. Was allowance made for uncertainty in the estimates of costs and consequences?					
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	Yes	Yes	Yes	Yes	Yes
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	Yes	Yes	Yes	Yes	Yes
9.3. Were the study results sensitive to changes in the values?	Yes – many conclusions for different populations sensitive to key parameters.	No	No	No	No
10. Did the presentation and discussion of study results include all issues of concern to users?					
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	Yes	Yes	Yes	Yes	Yes
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for	No	Yes	Yes	No	No

Drummond Checklist Questions	Linas (2011)	Pareek (2011)	del Campo (2012)	Eralp (2012)	Shah (2012)
potential differences in study methodology?					
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	No	Yes	No	No	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	No	Yes	No	Yes	No
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	No	Yes	No	Yes	Yes

Continuation of Table A8

Drummond Checklist Questions	Mandalakas (2013)	Pareek (2013)	Steffen (2013)	Swaminath (2013)	Verma (2013)
1. Was a well-defined question posed in answerable form?					
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes Children (0-2 and 3-5y.o.) contacts in South Africa.	Yes Recently arrived immigrants (≤35 y.o.) in the United Kingdom.	Yes 35y.o. close contacts of active TB cases in Brazil.	Yes Immunosuppressed with IBD in the USA.	Yes >65y.o. in long term care in Canada.
2. Was a comprehensive description of the competing alternatives given?					
2.1. Were there any important alternatives omitted?	No	No	No	No	No
2.2. Was (should) a do-nothing alternative be considered?	Yes, do nothing was included appropriately.	No	No	No	Yes, do nothing was included appropriately.
3. Was the effectiveness of the programme or services established?					
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No	No	No

Drummond Checklist Questions	Mandalakas (2013)	Pareek (2013)	Steffen (2013)	Swaminath (2013)	Verma (2013)
3.2. Was effectiveness established through an overview of clinical studies?	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes. Sensitivity and specificity of tests values from number of clinical studies.	Yes Sensitivity and specificity of tests values from number of clinical studies.
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes All effectiveness assumptions are said to be derived from published data obtained in high-burden settings, but no reference provided.	Yes Observational study performed to estimate LTBI prevalence.	Yes Many assumptions effecting effectiveness with no references/reasoning.	Yes Many assumptions to estimate effectiveness with only a small number with references/reasoning.	Yes Key assumption is TST characteristics used are for general population despite evidence of lower specificity and sensitivity for elderly. Could introduce overestimation of effectiveness hence underestimate of cost-effectiveness ratio.
4. Were all the important and relevant costs and consequences for each alternative identified?					
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Mandalakas (2013)	Pareek (2013)	Steffen (2013)	Swaminath (2013)	Verma (2013)
4.3. Were the capital costs, as well as operating costs, included?	Unclear. Outpatient hospitalisation costs included but no description provided.	No, state only direct medical costs included and clearly state breakdown of hospitalisation costs.	No.	No, authors use Linas 2011 hospitalisation costs.	No. Hospitalisation costs breakdown do not state any overheads/operating costs.
5. Were costs and consequences measured accurately in appropriate physical units?					
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	Yes Excluded adverse reaction costs as very rare events. Assumed no impact.	No	No	Yes. Excluded MDR-TB due to low prevalence and secondary reactivation of TB. Also did not include patient data for those who did not attend second clinic visit for TST reading. Assumed no impact.	No
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No	No	No
6. Were the cost and consequences valued credibly?					

Drummond Checklist Questions	Mandalakas (2013)	Pareek (2013)	Steffen (2013)	Swaminath (2013)	Verma (2013)
6.1. Were the sources of all values clearly identified?	Yes	Yes	Yes	Yes but no year of valuation stated.	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	Yes	Yes	Yes	Yes	Yes
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A	N/A	N/A
6.4. Was the valuation of consequences appropriate for the question posed?	Yes	Yes	Yes	Yes	Yes
7. Were costs and consequences adjusted for differential timing?					
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	Yes, 3%.	Yes, 3.5%.	No.	No.	Yes, 3%.

Drummond Checklist Questions	Mandalakas (2013)	Pareek (2013)	Steffen (2013)	Swaminath (2013)	Verma (2013)
7.2. Was there any justification given for the discount rate used?	Yes, state "standard rate" and referenced.	Yes, reference NICE recommendations.	N/A	N/A	No.
8. Was an incremental analysis of costs and consequences of alternatives performed?					
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes	Yes
9. Was allowance made for uncertainty in the estimates of costs and consequences?					
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	Yes	Yes	Yes	Yes	Yes
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	No Stated "reasonable range" but no reference.	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Mandalakas (2013)	Pareek (2013)	Steffen (2013)	Swaminath (2013)	Verma (2013)
9.3. Were the study results sensitive to changes in the values?	Yes Sensitive to LTBI rate.	Yes Sensitive to specificity of QFT-GIT.	Yes Sensitive to QFT costs and TST specificity.	No	Yes Sensitive to TB re-activation rate.
10. Did the presentation and discussion of study results include all issues of concern to users?					
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	Yes	Yes	Yes	No No cost-effectiveness ratio was used. Conclusions were based on "highest benefits" and "lowest cost" strategy.	Yes
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	No	Yes	Yes	No	No
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	No	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Mandalakas (2013)	Pareek (2013)	Steffen (2013)	Swaminath (2013)	Verma (2013)
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	Yes	Yes	Yes	Yes	Yes
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	Yes	Yes	Yes	No	No

Continuation of Table A8

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
1. Was a well-defined question posed in answerable form?					
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes Adults living with HIV in the United Kingdom.	Yes Pre arrival refugees to USA.	Yes Children, immunocompromised people and recently Arrived immigrants and General Population in the United Kingdom.	Yes Adults entering jail in the USA.	Yes Pre-arrival refugees to Canada.
2. Was a comprehensive description of the competing alternatives given?					
2.1. Were there any important alternatives omitted?	No	No	No	No	No
2.2. Was (should) a do-nothing alternative be considered?	Yes, do nothing was included appropriately.	Yes, do nothing was included appropriately.	No	No	Yes, do nothing was included appropriately.
3. Was the effectiveness of the programme or services established?					

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No	No	No
3.2. Was effectiveness established through an overview of clinical studies?	Somewhat. Sensitivity of IGRA used from one clinical study.	Yes Sensitivity and specificity of TST values from number of clinical studies	Yes Systematic review carried out to establish sensitivity and specificity of tests.	Yes Sensitivity and specificity of test values from clinical studies and systematic review.	Yes Sensitivity and specificity of test values from clinical studies and systematic reviews.
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes Both observational data from clinic and referenced assumptions used. Unclear method of calculating effectiveness outcomes.	Yes Key assumption is proportions of population with/without BCG vaccination with adequate references/reasoning. Authors use this assumption and one clinical study to estimate TST sensitivity.	Yes Many assumptions for effectiveness with no references/reasoning.	No	Yes Many assumptions for effectiveness all with adequate references/reasoning.

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
4. Were all the important and relevant costs and consequences for each alternative identified?					
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes	Yes	Yes
4.3. Were the capital costs, as well as operating costs, included?	Unclear breakdown of costs provided.	No. Includes labour but no indication of other operating/capital costs.	No. Clear breakdown provided and capital/overheads not included.	Yes. For QFT-GIT laboratory operations included.	No.
5. Were costs and consequences measured accurately in appropriate physical units?					
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	No	No	No	Yes Treatment and outcome costs. Potential underestimate of benefits of QFT-GIT.	No

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No	No	No
6. Were the cost and consequences valued credibly?					
6.1. Were the sources of all values clearly identified?	Yes	Yes	Yes	Yes	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	Yes	Yes	Yes	Yes	Yes
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A	N/A	N/A

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
6.4. Was the valuation of consequences appropriate for the question posed?	Yes	Yes	Yes	Yes	Yes
7. Were costs and consequences adjusted for differential timing?					
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	Yes, 3.5%	Yes, 3%	Yes, 3.5%	No	Yes, 1.5%
7.2. Was there any justification given for the discount rate used?	Yes, reference NICE recommendations.	Yes, reference Haddix et al. 2003.	Yes, reference NICE recommendations.	N/A	Yes, reference Canadian Agency for Drugs and Technologies in Health recommendations.
8. Was an incremental analysis of costs and consequences of alternatives performed?					
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
9. Was allowance made for uncertainty in the estimates of costs and consequences?					
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	Yes	Yes	Yes	Yes	Yes
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	Yes	Yes	Yes	No	Yes
9.3. Were the study results sensitive to changes in the values?	No	No	No	No	Yes Sensitive to treatments received.
10. Did the presentation and discussion of study results include all issues of concern to users?					
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	Yes	Yes	Yes	No	Yes

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	No	Yes	Yes	Yes	Yes
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	Yes	No	Yes	Yes	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	Yes	No	Yes	Yes	Yes

Drummond Checklist Questions	Capocci, (2015)	Wingate (2015)	Auguste. (2016)	Nijhawan (2016)	Campbell (2017)
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	Yes	No	Yes	Yes	Yes

Continuation of Table A8

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
1. Was a well-defined question posed in answerable form?					
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes Recently arrived immigrants <35y.o to Norway	Yes HCW with negative TST at time of employment.	Yes US-born or migrants living with or without comorbidities.	Yes Recent immigrants and contacts in the United Kingdom.	Yes Elderly (>65y.o.) at admission to residential care home in China, Hong Kong SAR.
2. Was a comprehensive description of the competing alternatives given?					
2.1. Were there any important alternatives omitted?	No	No	No	No	No

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
2.2. Was (should) a do-nothing alternative be considered?	Yes, do nothing was included appropriately.	No	Yes, do nothing was included appropriately.	Yes, do nothing was included appropriately.	Yes, do nothing was included appropriately.
3. Was the effectiveness of the programme or services established?					
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No		No
3.2. Was effectiveness established through an overview of clinical studies?	Somewhat. Sensitivity and specificity of TST values from only one study.	Yes Sensitivity and specificity of tests values from number of clinical studies.	Yes Sensitivity and specificity of tests values from number of clinical studies.	No	Yes Sensitivity and specificity of tests values from number of clinical studies.

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes Expert opinion and estimation used for multiple key effectiveness parameter. Other assumptions provided with no references/reasoning.	Yes Key assumption is no loss of QALYs with uncomplicated treatment of LTBI. All other assumptions references/reasoned adequately.	Yes Key assumption is that LTBI treatment without adverse effects causes no change in quality of life. Other assumptions provided with no references/reasoning.	Yes Cohort trial used to estimate sensitivity and specificity of tests. Reporting and selection bias possible.	Yes Many assumptions for effectiveness all with adequate references/reasoning.
4. Were all the important and relevant costs and consequences for each alternative identified?					
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
4.3. Were the capital costs, as well as operating costs, included?	Unclear. Hospitalisation costs included but no breakdown provided.	Unclear. Hospitalisation costs included but no breakdown provided.	Unclear. Treatment costs included but no breakdown provided.	No. Clear breakdown provided and capital/overheads not included	No. Clear breakdown provided and capital/overheads not included
5. Were costs and consequences measured accurately in appropriate physical units?					
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	Yes VAT, MDR-TB assumed minimal effect on analysis.	No	No	No	Yes. Costs of minor adverse events or additional radiologic tests. Underestimation of costs.

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No	No	No
6. Were the cost and consequences valued credibly?					
6.1. Were the sources of all values clearly identified?	Yes	Yes	Yes	Yes, but year of valuation not stated.	Yes, but year of valuation not stated.
6.2. Were market values employed for changes involving resources gained or depleted?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A	N/A	N/A
6.4. Was the valuation of consequences appropriate for the question posed?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
7. Were costs and consequences adjusted for differential timing?					
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	Yes, 4% Only costs discounted in base case analysis.	Yes, 3%	Yes, 3%	Yes, 3.5%	Yes, 5%
7.2. Was there any justification given for the discount rate used?	Yes, reference Norwegian Directorate of Health recommendations.	Yes, reference Sanders et al. 2016 recommendations.	State current recommendations.	State current recommendations.	Yes, reference Drummond et al. 2005.
8. Was an incremental analysis of costs and consequences of alternatives performed?					
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
9. Was allowance made for uncertainty in the estimates of costs and consequences?					
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	Yes	Yes	Yes	Yes	Yes
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	No	Yes	Yes	Yes	Yes
9.3. Were the study results sensitive to changes in the values?	Yes. Sensitive to IGRA cost.	No	Yes Sensitive to LTBI rate and TST sensitivity.	No	No.

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
10. Did the presentation and discussion of study results include all issues of concern to users?					
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	Yes	Yes	Yes	Yes	Yes
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	Yes	Yes	No	No	Yes

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	No	Yes	No	Yes	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Haukaas (2017)	Mullie (2017)	Tasillo, (2017)	Abubakar (2018)	Li (2018)
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	Yes	No	No	Yes	Yes

Continuation of Table A8

Drummond Checklist Questions	Sohn (2018).	Campbell (2019a).	Campbell (2019b).	Loureiro (2019).	Png (2019).
Was a well-defined question posed in answerable form?					
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes Adolescents (13-18y.o.) contacts in Japan.	Yes Pre-arrival immigrants to Canada	Yes Migrants with either late-stage CKD or beginning dialysis.	Yes Primary HCW in Brazil.	Yes HCW in Singapore.
2. Was a comprehensive description of the competing alternatives given?					
2.1. Were there any important alternatives omitted?	No	No	No	No	No
2.2. Was (should) a do-nothing alternative be considered?	No	Yes	Yes, do nothing was included appropriately.	No	Yes, do nothing was included appropriately.
3. Was the effectiveness of the programme or services established?					
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No	No	No

Drummond Checklist Questions	Sohn (2018).	Campbell (2019a).	Campbell (2019b).	Loureiro (2019).	Png (2019).
3.2. Was effectiveness established through an overview of clinical studies?	Yes Sensitivity and specificity of tests values from number of clinical studies and systematic reviews.	Yes Sensitivity and specificity of tests values from number of clinical studies.	Yes Sensitivity and specificity of tests values from number of clinical studies.	Yes Sensitivity and specificity of tests values from number of clinical studies.	Yes
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes Key assumption is the proportion of recent infection among contacts with LTBI. All assumptions for effectiveness all have adequate references/reasoning.	Yes All assumptions for effectiveness all have adequate references/reasoning.	Yes Key assumption is that no individuals had active TB at the time of LTBI screening. All assumptions for effectiveness all have adequate references/reasoning.	Yes Multiple assumptions for effectiveness all with no references/reasoning.	Yes Assumed 100% specificity and sensitivity of QFT-G with reasoning and references. All assumptions are stated to be from "published literature or expert opinion"
4. Were all the important and relevant costs and consequences for each alternative identified?					
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes	Yes	Yes
4.3. Were the capital costs, as well as operating costs, included?	Yes. Includes overheads and vehicle operations.	Unclear. Hospitalisation costs included but no breakdown provided.	Unclear. Hospitalisation costs included but no breakdown provided.	Yes. No capital costs stated but equipment use costs clearly included.	Yes. State that hospitalisation and testing costs include labour and overhead costs.

Drummond Checklist Questions	Sohn (2018).	Campbell (2019a).	Campbell (2019b).	Loureiro (2019).	Png (2019).
5. Were costs and consequences measured accurately in appropriate physical units?					
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	Yes Cost of referral to clinic assumed relatively minimal impact.	No	No	Yes MDR-TB due to low prevalence. Assumed minimal impact.	No
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No	No	No
6. Were the cost and consequences valued credibly?					
6.1. Were the sources of all values clearly identified?	Yes	Yes	Yes	No	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	Yes	Yes	Yes	No	Yes
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A	N/A	N/A

Drummond Checklist Questions	Sohn (2018).	Campbell (2019a).	Campbell (2019b).	Loureiro (2019).	Png (2019).
6.4. Was the valuation of consequences appropriate for the question posed?	Yes	Yes	Yes	Yes	Yes
7. Were costs and consequences adjusted for differential timing?					
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	No	Yes, 3%	Yes, 1.5%	No	Yes, 3%
7.2. Was there any justification given for the discount rate used?	State due to short horizon of analysis.	Yes, reference Sanders et al. 2016.	Yes, reference Canadian Agency for Drugs and Technologies in Health recommendations.	State due to short horizon of analysis.	Only state "commonly used rate".
8. Was an incremental analysis of costs and consequences of alternatives performed?					
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes	Yes
9. Was allowance made for uncertainty in the estimates of costs and consequences?					
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Sohn (2018).	Campbell (2019a).	Campbell (2019b).	Loureiro (2019).	Png (2019).
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	Yes	Yes	Yes	Yes	Yes
9.3. Were the study results sensitive to changes in the values?	Yes Sensitive to QFT-GIT cost.	Yes For migrants from low incidence conclusion only cost-effective 50% of time in probabilistic sensitivity analysis.	No	No	No
10. Did the presentation and discussion of study results include all issues of concern to users?					
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	Yes	Yes	Yes	Yes	Yes
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	Yes	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Sohn (2018).	Campbell (2019a).	Campbell (2019b).	Loureiro (2019).	Png (2019).
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	Yes	Yes	Yes	No	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	Yes	Yes	Yes	Yes	No
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	Yes	Yes	Yes	Yes	Yes

Continuation of Table A8

Drummond Checklist Questions	Al Abri (2020).	Jo (2020).	Steffen (2020)	Kim (2018)
1.Was a well-defined question posed in answerable form?				
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes	Yes	Yes	Yes
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	Yes 20-year-old recent immigrants in Oman.	Yes General population with vulnerable sub-populations.	Yes Adults living with HIV in Brazil.	Yes HIV+ Pregnant women
2. Was a comprehensive description of the competing alternatives given?				
2.1. Were there any important alternatives omitted?	No	Unclear alternatives provided.	No	No
2.2. Was (should) a do-nothing alternative be considered?	No	No	No	No
3. Was the effectiveness of the programme or services established?				
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	No	No	No	No
3.2. Was effectiveness established through an overview of clinical studies?	Yes Sensitivity and specificity of tests values from number of clinical studies and systematic reviews.	Yes Sensitivity and tests values from a single systematic review.	Yes Sensitivity and specificity of tests values from number of systematic reviews.	Yes Sensitivity and specificity of tests values from number of systematic reviews.

Drummond Checklist Questions	Al Abri (2020).	Jo (2020).	Steffen (2020)	Kim (2018)
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes No clear assumptions stated. TB incidence and the LTBI rate, BCG vaccination rate, (MDR-TB) rate, TB mortality, and rate of adherence to treatment were imputed from Omani data.	Yes Most assumptions for effectiveness all have adequate references/reasoning.	Yes All assumptions for effectiveness all have adequate references/reasoning.	Yes All assumptions for effectiveness all have adequate references/reasoning.
4. Were all the important and relevant costs and consequences for each alternative identified?				
4.1. Was the range wide enough for the research question at hand?	Yes	No	Yes	Yes
4.2. Did it cover all relevant viewpoints?	Yes	Yes	Yes	Yes
4.3. Were the capital costs, as well as operating costs, included?	Unclear breakdown of costs provided.	Unclear breakdown of costs provided.	Yes. No capital costs stated but equipment use costs clearly included.	Yes Costs of overheads, building space, equipment, staff, and consumables included.
5. Were costs and consequences measured accurately in appropriate physical units?				
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	No	No	No	No

Drummond Checklist Questions	Al Abri (2020).	Jo (2020).	Steffen (2020)	Kim (2018)
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	No	No
6. Were the cost and consequences valued credibly?				
6.1. Were the sources of all values clearly identified?	No	Yes	Yes	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	No	Yes	Yes	Yes
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	N/A	N/A	N/A	N/A
6.4. Was the valuation of consequences appropriate for the question posed?	No	Yes	Yes	Yes
7. Were costs and consequences adjusted for differential timing?				
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	No	Yes, 3%	Yes, 3%	Yes, 3%
7.2. Was there any justification given for the discount rate used?	No	Yes, reference Sassi 2006.	Yes, reference Brazilian Ministry of Health recommendations.	No
8. Was an incremental analysis of costs and consequences of alternatives performed?				
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes

Drummond Checklist Questions	Al Abri (2020).	Jo (2020).	Steffen (2020)	Kim (2018)
9. Was allowance made for uncertainty in the estimates of costs and consequences?				
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	Yes	Yes	Yes	Yes
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	No	Yes	Yes	Yes
9.3. Were the study results sensitive to changes in the values?	Yes Sensitive to treatments received.	Yes Specificity of QFT-GIT.	Yes Cost of DiaskinTest.	Yes Sensitive to the probability of developing TB and LTBI prevalence
10. Did the presentation and discussion of study results include all issues of concern to users?				
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	Yes	Yes	Yes	Yes
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	Yes	Yes	Yes	Yes
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	No	Yes	Yes	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	No	Yes	Yes	Yes

Drummond Checklist Questions	Al Abri (2020).	Jo (2020).	Steffen (2020)	Kim (2018)
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	No	Yes	Yes	Yes

Table A9. Summary of the proportion of articles accomplishing each of the Drummond's criteria

Drummond Checklist Questions	% Of papers accomplishing each criterion		
	All articles	DiaskinTest-related	TST or IGRA
1. Was a well-defined question posed in answerable form?			
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	97.1%	87.5%	100.0%
1.2. Did the study involve a comparison of alternatives?	100.0%	100.0%	100.0%
1.3. Was a viewpoint for the analysis stated and was the study placed in any decision-making context?	100.0%	100.0%	100.0%
2. Was a comprehensive description of the competing alternatives given?			
2.1. Were there any important alternatives omitted?	5.7%	25.0%	0.0%
2.2. Was (should) a do-nothing alternative be considered?	42.9%	12.5%	50.0%
3. Was the effectiveness of the programme or services established?			
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	0.0%	0.0%	0.0%
3.2. Was effectiveness established through an overview of clinical studies?	91.4%	87.5%	92.9%
3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	85.7%	62.5%	92.9%
4. Were all the important and relevant costs and consequences for each alternative identified?			
4.1. Was the range wide enough for the research question at hand?	97.1%	100.0%	96.4%
4.2. Did it cover all relevant viewpoints?	100.0%	100.0%	100.0%
4.3. Were the capital costs, as well as operating costs, included?	31.4%	62.5%	25.0%
5. Were costs and consequences measured accurately in appropriate physical units?			
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	25.7%	25.0%	25.0%
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	0.0%	0.0%	0.0%
6. Were the cost and consequences valued credibly?			
6.1. Were the sources of all values clearly identified?	82.9%	50.0%	92.9%
6.2. Were market values employed for changes involving resources gained or depleted?	94.3%	100.0%	92.9%
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	0.0%	0.0%	0.0%

6.4. Was the valuation of consequences appropriate for the question posed?	94.3%	87.5%	96.4%
7. Were costs and consequences adjusted for differential timing?			
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	60.0%	12.5%	75.0%
7.2. Was there any justification given for the discount rate used?	60.0%	12.5%	75.0%
8. Was an incremental analysis of costs and consequences of alternatives performed?			
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	94.3%	75.0%	100.0%
9. Was allowance made for uncertainty in the estimates of costs and consequences?			
9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	77.1%	37.5%	89.3%
9.2. If sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	71.4%	37.5%	82.1%
9.3. Were the study results sensitive to changes in the values?	34.3%	12.5%	42.9%
10. Did the presentation and discussion of study results include all issues of concern to users?			
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)?	77.1%	50.0%	85.7%
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	40.0%	12.5%	50.0%
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	57.1%	37.5%	64.3%
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	62.9%	37.5%	71.4%
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	62.9%	50.0%	67.9%

World Health Organization

9789240056626