WHO consolidated guidelines on tuberculosis

Module 3: diagnosis. Tests for TB infection

Web Annex F

Qualitative evidence for the use of *Mycobacterium tuberculosis* antigen-based skin tests

Review team:

Amy Hai Yan Chan[1,2] , Kate Loveys[1], Matthew Quaife[3], Amy Clarke[2], Liliya Eugenevna Ziganshina[4], Irina Kontsevaya[5], Chloe Tuck[6], Elena Surkova[7], Ting Ting Wang[2], Liu (Victoria) Wan-Hsin[3], Yohhei Hamada[2], Molebogeng X Rangaka[2]

1. The University of Auckland, Auckland, New Zealand
2. University College London, London, United Kingdom
3. London School of Hygiene and Tropical Medicine, United Kingdom
4. Cochrane Russia, Russian Federation
5. Research Center Borstel, Germany
6. The University of Sheffield, Sheffield, United Kingdom
7. Royal Brompton Hospital, Part of Guy's and St Thomas' NHS Foundation Trust 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

World Health Organization

WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for TB infection. Web Annex F. Qualitative evidence for the use of *Mycobacterium tuberculosis* antigen-based skin tests/ Amy Hai Yan Chan, Kate Loveys, Matthew Quaife, Amy Clarke, Liliya Eugenevna Ziganshina, Irina Kontsevaya et al.

ISBN 978-92-4-005664-0 (electronic version)

Suggested citation. Chan AHY, Loveys K, Quaife M, Clarke A, Ziganshina LE, Kontsevaya I et al. Web Annex F. Qualitative evidence for the use of *Mycobacterium tuberculosis* antigen-based skin tests. 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 Clamette-Guérin

DCE – Discrete Choice Experiment

GRADE-CERQual – Confidence in the Evidence from Reviews of Qualitative Research

HIC – High Income Country

HIV – Human Immunodeficiency Virus

IGRA – Interferon gamma release assays

LMIC – Low-Middle Income Country

PPD-TST – Purified protein derivative tuberculin skin tests

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

QFT-gold – QuantiFERON-TB Gold

TBI – Tuberculosis infection

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

WHO – World Health Organization

Executive Summary

Background: New *Mycobacterium tuberculosis* specific antigen-based skin tests (TBST) have been developed. However, end-user values and preferences, acceptability and feasibility for use in clinical care, and possible impact on health equity outcomes are unknown. Understanding the factors that affect these key outcomes may help to improve utilisation and programmatic implementation.

Methods: Objective 4 was informed by three work packages – A, B, C. This Executive Summary draws on the findings from all three work packages: A-Systematic Review, B-User Interviews), C End-user values and preferences: Discrete Choice Experiment (DCE).

For **work package A**, two systematic reviews were conducted that synthesized the qualitative research evidence on end-user values and preferences for the use of specific TBST for latent TB infection, compared to existing tests (IGRA, TST). One review aimed to capture data on TBST specifically as compared to IGRA and /or TSTs (Review A), and another focused on IGRA and TST only because we anticipated few studies on novel TBST (Review B). A public call for data relating to TBST was also announced by the WHO. Data were extracted from eligible studies, synthesised then analysed thematically. Study quality and confidence in the evidence was evaluated in accordance with the GRADE-CERQual. The reviews were registered in PROSPERO: Systematic review A, specific TBST compared to the IGRA and/or the TST (CRD42021273744); Systematic review B, IGRA and TST compared to each other or without a comparator (CRD42021273952). A qualitative coding framework was developed from the themes and subthemes that emerged from these reviews. This was then used for analysis of data from work package B.

For **work package B**, 20 semi-structured interviews were conducted with a diverse range of clinicians, laboratory staff, programme officers, and individuals living with LTBI (referred to as 'consumers' throughout this report), with the aim of understanding their experiences of using TBI tests, including the IGRA, TST, and TBST. Findings from these interviews were synthesised and analysed thematically using the coding framework from work package A, with adaptations to accommodate additional key themes and subthemes influencing TBST acceptability, feasibility and health equity identified from the interviews.

For **work package C**, an online cross-sectional discrete choice experiment (DCE) survey was conducted drawing from themes derived in work packages A and B. DCE methodology was used to elicit stated values and preferences from participants (end users) without directly asking them to state their preferred options. Consumers (people living with TBI/eligible for TB infection testing or underwent testing previously) or providers (healthcare providers involved in the TB care pathway) were invited to participate via global networks of TB advocates, researchers and policy makers. The survey was delivered in English and required access to the internet to complete. Different designs were developed for consumers and providers to represent the different potential determinants of choice to either, and a minimum sample size of 45 was deemed sufficient to draw statistical inference in both designs. Choice data were analysed using multinomial logistic regression models, and willingness-to-accept analyses used to compare results within and across designs.

Results:

Characteristics of data included in each work packages

For **work package A**, four studies were identified that met the inclusion criteria for both systematic reviews. From Review A on specific TBSTs, only one data sources was identified from the Russian Federation, which came from a WHO public call for data relating to the feasibility and acceptability

of antigen-based TB skin tests. Participants were parents of children with TBI. From Review B on current IGRA and TST, 3 peer-reviewed articles were found to meet the inclusion criteria; these 3 papers were from the United States, South Africa and the Netherlands. Participants included a range of health professionals involved in TB care (US and Netherlands) and people living with HIV (South Africa). The overall confidence in the quality of the evidence from the studies was low to moderate based on the GRADECerQual assessments, as the data lacked richness, with most studies reporting only summaries of participant quotes or limited direct quotes. All studies were conducted on specific subgroups (e.g. people living with HIV, or parents of children and adolescents with TBI).

For **work package B** (user interviews), we recruited 20 participants– 13 were TB healthcare providers (8 from low middle-income countries) and 7 TB consumers (3 from low-middle income countries). Healthcare providers included program executives and decision-makers, public health practitioners and advocates, physicians, researchers and lab technicians, and a nurse.

For **work package C** (End-user values and preferences: DCE), we had a total of 234 participants complete the DCE (186 providers and 48 consumers). Overall, 59% of respondents were female, 56% were aged 36-55 years, and the main countries in which respondents were based were India (26%), the United States (16%), South Africa (9%), Pakistan (8%) and Zimbabwe (7%).

End-user values and preferences

Qualitative data from the systematic reviews and end-user interviews, and quantitative data from the DCE, indicated that healthcare consumers and providers had similar values and preferences in terms of TBI tests. Key end-user values included test accuracy, convenience, positive patient experience, cost, and resource requirements. In particular, end-users valued tests with high accuracy such as the TBSTs and IGRA (i.e., low false-positive and negative rates), as they reduce the risk of downstream consequences associated with false positive and negative results (e.g., anxiety, the need for additional testing or unnecessary treatment). End users also preferred having a test that was convenient to administer and access. This included valuing tests that can be accessed in a community or primary care setting, that avoid follow-up visits to read test results, and that can be administered without the need for additional systems or infrastructure to be developed. These findings were initially identified from themes emerging from the systematic reviews and end-user interviews, and confirmed by the DCE findings.

From the qualitative data from the reviews and interviews, all TBI test options were found to have strengths and limitations in terms of convenience. A positive consumer experience was valued by end-users. This meant that tests with fewer psychological (e.g., anxiety, stigma, stress) and physical consequences (e.g., discomfort) were preferred. Tests that were more accurate tended to be associated with better consumer experience, although some aspects of consumer experience were worse in skin tests (e.g., stigma from the welt, discomfort) compared to non-skin-based tests. Low-cost tests were generally preferred due to greater accessibility in resource-limited contexts (e.g., the TBSTs, TST). Tests with lower resource requirements were preferred in resource-limited settings (e.g., the TBSTs, TST), however, this appeared to be less of a consideration in high-income countries. End users showed a preference toward TBI tests that used existing infrastructure in their healthcare setting. Data from the DCE confirmed that a lack of a need for an in-person follow-up appointment and not requiring specialist staff or equipment to interpret or administer the test, were important end-user preferences for TB testing.

Acceptability

Qualitative data from systematic reviews and end-user interviews suggests TBST were perceived to have greater specificity and sensitivity than the TST. Having greater test accuracy was deemed desirable to avoid the negative consequences of false positives or negatives. However, the TBST was expected to have many of the same limitations as skin tests in terms of patient experience (e.g., a return visit required, discomfort, a welt on the arm, stigma) compared to the IGRA. The IGRA was deemed the preferred test option in countries which already have the IGRA in use, as the required supporting infrastructure are already in place, and because the TBST would have comparable accuracy and performance, thus not adding value. There were also concerns about skin tests as these tests were viewed as a dated, basic technology subject to human error and/or interpretation. Suggestions for improving the acceptability of TBST included careful communication during its implementation with endorsement by healthcare providers and organizations (e.g., the WHO). Data from the DCE found strong and consistent preferences among both healthcare providers and consumers for tests which minimize false positive and negative results. The DCE also found that consumers had a strong preference for testing in the community and primary care settings compared to hospital locations.

Feasibility

Findings from the qualitative evidence synthesis (reviews and end-user interviews) support the feasibility of use of the TBST but only in settings where the TST is already in use, as the required resourcing and training is already in place. The TBST is likely to be a low-cost, portable test that can be well-suited for low-resource healthcare settings, which may not be able to support the IGRA, due to the comparably greater cost and resources required to implement the IGRA. However, if healthcare settings already have the necessary infrastructure in place to implement the IGRA, then the IGRA are a more feasible test option than either skin tests because they do not require a return visit to read the result, a step where patients may be lost to follow-up. Results from the DCE found that lack of a need for an in-person follow-up appointment and not requiring specialist staff or equipment to interpret or administer the test, were important preferences for TB testing that would influence feasibility. There was some suggestion that providers preferred more expensive tests (when offered a choice based on a hypothetical cost of USD 50 compared to USD 25), though test cost was the least important determinant of test choice.

Health equity

Qualitative evidence from reviews and end-user interviews indicate that it is unlikely that specific TBST would create any new equity issues. Rather, the TBST are likely to improve health equity through the provision of a more accurate, low-cost test for resource-limited settings where the PPD-TST is already in use. Moreover, their portability and low cost make them suited to use in large scale screening programs in vulnerable, hard-to-reach communities. However, it is possible that TBST may not affect health equity in low-resource settings that do not already use the TST, as barriers to accessing skin and other healthcare tests in these settings exist, which would need to be addressed first, regardless of the type of TB test available. In terms of test accessibility, the data from the DCE found that consumers had a strong preference for testing in the community and primary care settings compared to hospital locations, which could have health equity implications.

Conclusion:

Findings across the three work packages were consistent with each other in the themes, and subthemes identified that would influence TBST acceptability and feasibility. The qualitative evidence synthesis from the systematic reviews and the user interviews, as well as the quantitative

findings from the DCE highlight the potential for novel antigen-based skin tests (TBST) as an acceptable alternative to the TST. The findings from the DCE support the data from the reviews and interviews, even though published review data from work package A were limited. Our findings were generally consistent across both healthcare providers and consumers.

In conclusion, specific TBST were found to be more acceptable and equally feasible to the TST, with greater benefits for health equity driven by higher sensitivity and specificity. TBST may be more preferrable for use in more resource-limited settings compared to the TST due to the higher accuracy whilst retaining portability and ability to administer in a community / primary care setting. Test preferences were influenced by end-user values such as test accuracy, convenience, cost, positive patient experience, and resource requirements. IGRA remain the preferred test option in high-income countries with existing infrastructure to support their use; it is unlikely that widespread use of the IGRA would change in these settings where IGRA is mostly the standard of care, particularly due to the strong provider preference for less of a need for an in-person follow-up appointment for all tested individuals, so resources can focus on those who do require follow-up such as those who test positive.

1. Introduction

Tuberculosis (TB) is an airborne transmissible infectious disease caused by *Mycobacterium tuberculosis*. TB is a major cause of health loss (i.e., healthy life years lost to death, illness, or disability), affecting approximately one quarter of the global population and ranking as one of the top ten causes of death[1]. Approximately one quarter to one-third of the global population is estimated to have latent TB infection (TBI) [2, 3]. TBI may progress to active disease, especially in those with pre-disposing factors such as HIV infection, diabetes, and immunosuppressive conditions, as well as individuals who are contacts of people with TB especially children [1,4]. Testing for TBI is desirable for diagnosis and to guide preventative treatment to curtail progression to active disease and onward disease transmission[5].

Current tests for TBI include the purified protein derivative (PPD) tuberculin skin tests (TST) and interferon gamma release assays (IGRA), which include the WHO-endorsed QuantiFERON-TB Gold (QFT-gold; QFT-gold-in-tube; QFT-plus) and the T-SPOT TB technologies. Despite the importance of TBI testing, there is a large gap in the number of people tested, treated and the numbers estimated to have TBI[1] **.** This issue might be due to poor access to TB testing and / or poor uptake and engagement with TB testing. Both of these can be improved by increasing available options and platforms for TBI testing, improving test performance, and addressing barriers to use. Understanding the factors that contribute to the acceptability and feasibility of TBI tests may help to improve utilisation.

Acceptability and feasibility of tests for TBI may be affected by a variety of factors[6]. These include factors pertaining to the TBI test (e.g., time to screen, invasiveness, safety, prognostic accuracy, self-test ability), the testing context (e.g., cost, convenience, privacy), the test administrator (e.g., rapport, opportunity to have questions answered), the individual living with TBI (e.g., risk perception, fear of TB), the societal context (e.g., stigma, cultural appropriateness), and the implications of the test result (e.g., cost and availability of treatment, negative social consequences)[6]. Several features of TST and IGRA may impair their acceptability and feasibility. TST, although widely available at low cost, require two clinic visits by the individual (one to administer the test, one to read the result 48-72 hours later), which may be expensive and impracticable for both people accessing or paying for healthcare as well as healthcare providers[7]. Moreover, TST has been shown to have low specificity as they provide false-positive results in individuals with previous recent Bacillus Calmette-Guérin (BCG) vaccination[8], and low sensitivity in immunosuppressed people (e.g., individuals with HIV), which may further decrease acceptability. Recent global shortages in the supply of PPD may also reduce test accessibility. IGRA, on the other hand, have better specificity and sensitivity than TST as predictive performance is not impaired by prior BCG vaccination or immunosuppressive conditions[9, 10]. Moreover, IGRA requires one visit by individuals and results can be generated more rapidly than TST and remain on record. However, IGRA can be considerably more expensive than TST, which may reduce feasibility[11]. This is because they traditionally require an accredited laboratory, specialised kits, a phlebotomist to draw a blood sample, is time-sensitive between collection and incubation at the laboratory, and qualified technicians are needed to analyse the sample. Tests are also often run in batches to reduce health system costs; thus, it is not necessarily true that results would be immediate. IGRA are also more invasive, requiring venepuncture, which may impair acceptability.

Over the past decade, *Mycobacterium tuberculosis* specific antigen-based skin tests (TBST) have been developed with the aim of improving diagnostic performance while using the same platform as the PPD-TST. These antigen-based skin tests include the C-Tb (Staten Serum Institut, Denmark), the Diaskintest (Generium, Russain Federation), and the ESAT6-CFP10 test (Anhui Zhifei Longcom,

China). Emerging evidence suggests that specific antigen-based skin tests may have similar diagnostic performance to IGRA tests[12]. However, like the TST, they are relatively portable and expected to have a low cost. Although, they require two clinic visits to administer the test and evaluate the results, which may pose feasibility and equity concerns. As these tests are relatively new, it is not known whether they are acceptable or feasible for use in routine practice for end users (e.g., individuals and healthcare providers), and whether use of these tests would impact health equity.

To address this knowledge gap, World Health Organization (WHO) commissioned systematic reviews into the perspectives, preferences, and experiences of users of TBST and explore health equity. The reviews aim to qualitatively explore what the end user perspectives, preferences and experiences are of TBST, particularly regarding their acceptability and feasibility, and any impacts on health equity. Anticipating a paucity of papers to inform the systematic reviews, we further sought to collect primary data from in-depth interviews. The findings from the systematic reviews and in-depth interviews were then used to design a pilot discrete choice experiment (DCE) to identify the key attributes and trade-offs that inform user choices and preferences for TBST.

2. Overall Aim

The aim of this objective was to qualitatively explore and understand perspectives on the acceptability, feasibility, and health equity impacts of antigen-based skin tests (TBST) for TB infection.

Participants: End-users (individuals living with TB or health providers involved in the care of people with TB)

Intervention: Antigen-based skin tests for TBI (TBST)

Comparator: Any other test for TBI (e.g. IGRA, or TST)

Outcomes: Acceptability, feasibility, health equity, and end-user values and preferences.

Specific objectives

A- To conduct a systematic review that summarizes the evidence on end-user values and preferences regarding the acceptability and feasibility of antigen-based skin tests for TB infection, compared to current tests of infection (IGRA, TST).

B- To conduct in-depth interviews to obtain data on views and perspectives of TBST compared to existing TBI tests during qualitative interviews with end-users and key informants (primary data collection).

C- To evaluate any potential impacts of TBST on health equity, based on findings from A and B.

D- To explore end-user values and preferences for TB testing by investigating end-user values and preferences for TB testing by investigating the trade-offs end users are willing to make between test attributes derived from systematic reviews and in-depth user interviews, and determine the order of preference of different test attributes using a discrete choice experiment.

3. Approach

Proposed activities

Three studies were designed and conducted to achieve the overall project aim, with findings from the systematic review feeding into activities B and C. Methods and results of each are reported in the relevant section describing each of the following work packages:

A. Systematic review of qualitative evidence
B. End-user interviews
C. Discrete choice experiment (DCE)

An expert technical advisory panel comprised of multidisciplinary experts with skills in global health, infectious diseases and behavioural medicine contributed to the design of all three studies.

3.1 Work Package A-Systematic review

This activity intends to answer the following objectives:

Objective A: To conduct a systematic review that summarizes the evidence on end-user values and preferences regarding the acceptability and feasibility of antigen-based skin tests for TB infection, compared to current tests of infection (IGRA, TST).

Objective C: To evaluate any potential impacts of TBST on health equity

Two systematic reviews were conducted that synthesized the qualitative research evidence to date on end-user values and preferences for the use of TBST for latent TB infection, compared to existing tests (IGRA, TST). In addition to a review to capture data on TBST specifically as compared to IGRA and /or TSTs, we conducted another one focusing on IGRA and TST only because we anticipated few studies on TBST. We considered that given that the operational and logistic requirements are the same for TBST and TST, the findings from the second review would supplement those from the first review. Study quality and confidence in the evidence was evaluated in accordance with the GRADE-CERQual. The reviews were registered in PROSPERO on 02/09/2021 under separate records that focused on: (Systematic review A) TBST compared to the IGRA and/or the TST (CRD42021273744); and (Systematic review B) The IGRA and TST compared to each other or without a comparator (CRD42021273952). Methods were similar for each SR except where indicated.

3.1.1 Search strategy for both Systematic Reviews A and B

The systematic search terms are presented in Appendix A. The search strategy was informed by a recent systematic review that investigated the diagnostic performance of antigen-based skin tests for TB infection[13], by adapting the search terms used for TBST, TST and IGRA, and a prior qualitative evidence synthesis on the feasibility and acceptability of nucleic acid amplification tests (NAATs) for TB detection conducted for WHO[14], adapting their search terms for acceptability and feasibility. The search terms for tuberculosis and test types were consistent with objectives 1-3 in this work.

Electronic searches

The following databases were searched on 07 July, 2021 using the search terms and strategy outlined in Appendix A:

- EMBASE (Ovid) (Systematic Reviews A and B)
- MEDLINE (Ovid) (Systematic Reviews A and B)
- e-library (via our Russian collaborators – Russian versions of Diaskintest OR "Recombinant tuberculosis allergen": Диаскинтест* или «Аллерген* туберкулезн* рекомбинантн*) (Systematic Review A)
- Chinese Biomedical Literature Database (using CFP-10 and ESAT6) (Systematic Review A)
- China National Knowledge Infrastructure Database (using CFP-10 and ESAT6) (Systematic Review A)
- CINAHL (EBSCOHost; Cumulative Index to Nursing and Allied Health Literature) (Systematic Review B)
- PsycInfo (EBSCOHost) (Systematic Review B)

Manual searches

Manual searches were conducted on the reference lists and author publications of included studies and any related reviews to identify additional literature. Experts in the field were consulted to assist with identifying additional eligible literature. Studies were also reviewed that were identified through a public call for data by the 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).

Grey literature

Grey literature was not searched as part of the review.

3.1.2 Eligibility criteria

Study design

Included studies were primary research studies that used qualitative methodologies or study designs. Studies were required to use qualitative methods for data collection (e.g., interviews, focus groups, open-ended survey questions) and analysis (e.g., thematic analysis, grounded theory). Mixed methods studies that used qualitative methods for data collection and analysis were included. Survey studies that presented qualitative free text that was collected in response to open-ended questions were eligible for inclusion.

Studies were excluded that analysed qualitative data using quantitative methods (e.g., descriptive statistics), as insufficient detail was provided to synthesize qualitative themes in the data[15]. Other excluded study designs were review papers, letters without original data, abstracts, case reports, and studies reporting insufficient data for qualitative synthesis (e.g., only dropout rates as an acceptability measure).

Language

No language limits were placed on the search results.

Publication year

Search results were restricted to 2011 – present.

Publication status and type

Included studies were published, peer-reviewed journal articles and proceedings, conference proceedings, and in press literature.

Participants

Included studies focused on individuals living with TB and/or their caregivers, laboratory technicians, and/or healthcare providers who were users or potential users of TB tests. Potential users refer to groups who would use a TB test, but they could not access the test, or the test was not feasible to use in their setting. Studies were excluded that focused on animals.

Intervention

Included studies evaluated at least one of the following TBST including C-TB, Diaskintest, and ESAT6-CFP10 test (EC). Studies were also included that focused on comparator tests including the TST and interferon gamma release assay (IGRA) tests, including the QuantiFERON-TB Gold test (QFT-gold; QFT-gold-in-tube; QFT-plus), and the T-SPOT TB test.

Outcomes

Studies were included that evaluated end-users values and/or preferences, acceptability, feasibility, or health equity.

Setting

Included studies were conducted in any low-, middle-, or high-income country and in any setting, including any type of healthcare facility (e.g., hospital, clinic, community health centre, mobile testing van).

3.1.3 Data management

Two independent reviewers (AC, KL) screened the English search results in two stages (title and abstract screen, and a full-text screen) using Covidence software (Veritas Health Innovation, Melbourne, Australia). Research assistants who were native speakers of Russian and Chinese screened the results of the Russian and Chinese databases from Review A in duplicate using Rayyan software[16].

Data extraction

Qualitative data pertaining to test acceptability, feasibility, and health equity were extracted from the included papers independently in duplicate by two raters (AC, KL) using Covidence software. Data were also extracted on study characteristics, including publication language, country, type of facility (including public/private, rural/urban), study objectives, methods of data collection and analysis, study duration, conceptual framework, conflicts of interest, study participant type and mean age, percentage male, ethnicity, diagnostic tests involved, programmatic features of the intervention, comparator conditions.

Data synthesis

A qualitative evidence synthesis from Systematic Reviews A and B was conducted to investigate end users' values/preferences, and user perspectives on the acceptability, feasibility and health equity impacts of TBST compared to the IGRA and TST. A thematic analysis was conducted to identify themes that contributed to the perceived acceptability, feasibility, and health equity from end-users of the TBST, IGRA and TST. Due to the small number of eligible papers (see Appendix B), results were collapsed for the two reviews. Two researchers (AC, KL) coded the data independently in duplicate using a coding scheme iteratively developed by the research team in NVivo software (QSR International Ltd, Melbourne, Australia). Multiple rounds of data analysis were undertaken to refine the coding scheme. The final themes were refined by feedback from a multidisciplinary advisory group and the wider research team.

Data quality

The GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) tool was used to assess confidence in the systematic review findings from both Review A and B by two raters (AC, KL) independently in duplicate spreadsheet forms[17]. The GRADE-CERQual can only be used for systematic reviews, so it has been used only for work package A. The tool examines the methodological limitations of the included studies, the coherence of each review finding, the adequacy of the data in support of a review finding, and the relevance of the included studies to the review research questions. Assessments were made for each study and review finding against the criteria, and an overall judgment was made about confidence in the evidence supporting the review findings. Final assessments were made through a consensus discussion by the raters (AC, KL).

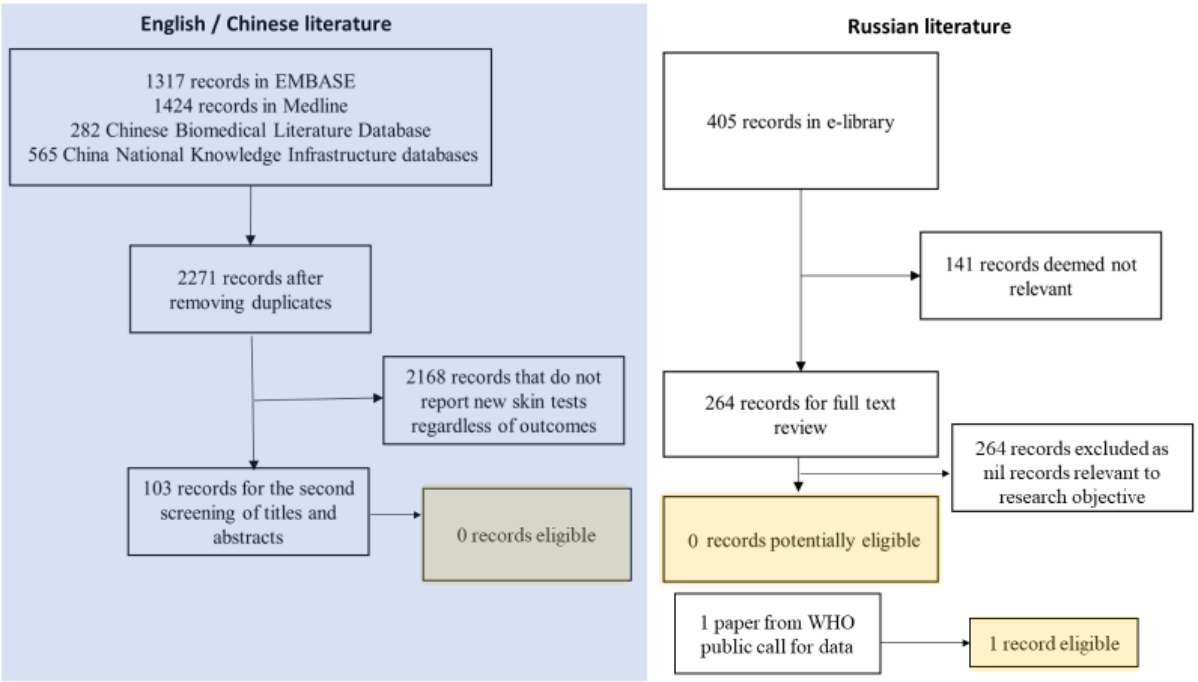
3.1.4 Results from Systematic Review A and B

3.1.4.1 Results of the Search

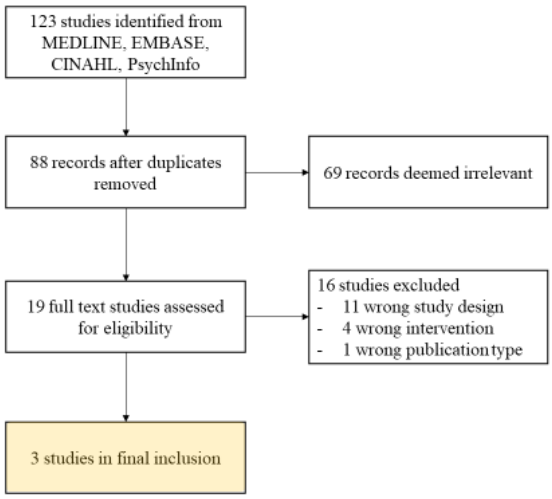
Four studies were identified that met the inclusion criteria (see Figure A1 for Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow charts). Review A included papers focused on specific TBST compared to the IGRA and/or the TST, while Review B focused on papers on the IGRA and TST compared to each other or without a comparator. From Review A for TBSTs, only one study was identified[18], which came from a WHO public call for data relating to the feasibility and acceptability of antigen-based TB skin tests. From Review B, which focused on IGRA and TST, 3 peer-reviewed articles were found to meet the inclusion criteria[19,20,21]. The included studies were published between 2014-2021.

Figure A1: PRISMA Flow Charts

Review A: Papers on TBST compared to the IGRA and/or the TST



Review B: Papers on the IGRA and TST compared to each other or without a comparator



3.1.4.2 Description of the Studies

A table summarizing the Characteristics of the included studies identified from Review A and B are presented in Table A1.

Test

From Review A, only one study reported on a TBST which was the Diaskintest[18], which was from the WHO public call for data. The Diaskintest was evaluated as part of a large-scale screening program for children and adolescents in the Penza region of the Russian Federation. From Review B, there were three published studies reported on the IGRA and the TST[19,20,21]. Two studies focused on the use of the IGRA and TST as part of clinic services for HIV individuals[19,20]; the third study explored the IGRA and TST in the context of Public Health Services in the Netherlands, which was targeted at new immigrants from non-EU countries with a high TB incidence[21].

Country and healthcare setting

The public call for information for TBSTs for Review A attracted data from an upper-middle income country, the Russian Federation[18]. Two studies from Review B were based in high-income countries (USA, The Netherlands[20,21]), and one in an upper-middle income country (South Africa[19]). One study was conducted in an urban setting, one was across both rural and urban areas, and two did not report on the setting. Three studies were conducted in publicly funded healthcare settings, and one did not report on the funding of the service. The included studies were conducted across different types of healthcare facilities: one was based in a public health service, one across 14 primary care clinics, one in a range of facilities (e.g., 2 hospitals, 1 private community clinic, 1 health department clinic), and one did not report on the type of healthcare facility.

Participants

Participants in the included studies from Review A were parents (n=not reported)[18]. Participants from Review B were clinic staff (n=20)[20], a combination of clinical care providers (comprised of 28 physicians, 12 TB specialist nurses, 11 general nurses, 3 counsellors, 1 operational manager, 1 community health worker; n=28) and HIV individuals (n=28)[19] and 5 teams of Public Health Service staff (comprised of at least 1 Medical Technical Assistant, 1 TB nurse, and 1 TB physician)[21]. Only one study provided the mean age of participants which was 38.4 years for the HIV individual interviewees[19].

Study design

Most (3 of 4) studies from Review A and B adopted a mixed-method design[18,20,21]. Only 1 study from Review B was purely qualitative with interviews of individuals living with HIV and their healthcare providers, who were taking part in a larger trial (TEKO)[19]. Most studies were descriptive in nature; only 1 from Review B adopted a conceptual framework to inform the analyses, which was Levesque's Conceptual Framework for Access to Care[21]. Thematic data analysis was used by 2 of the studies[19,21], whereas the 2 other studies did not report on their approach to data analysis. For all studies, the objectives were not directly relevant to the aims of this report (e.g. they focused on specific subgroups, or TB screening and treatment programs of which TBI tests were only a part).

3.1.4.3 Methodological Limitations of the Studies

Overall, there were moderate concerns about the methodological quality of the studies included in review A and B. None of the included 4 studies reported on the relationship between the researcher and the participants. Half (2 of 4) of the studies (1 from Review A[18], and 1 from Review B[21]) did not report on the recruitment or data analysis methods. For half of the studies (1 from Review A[18] and 1 from Review B[20]), direct participant quotes were not reported thus, it was not possible to evaluate the fit of the data to the research findings. The study from Review A[18] did not report on ethical considerations (e.g., IRB approval, informed consent).

3.1.4.4 Confidence in the Review Findings

Confidence in the Review Findings was determined based on the data from all 4 studies combined from Review A and B. Of the 4 review findings (end user values/preferences, acceptability, feasibility, health equity), 3 findings (end user values and preferences, acceptability, feasibility) were rated with moderate confidence based on the findings from Review A and B. Health equity received a low confidence rating. Appendix B provides a detailed summary table of the GRADE-CERQual analysis. There were serious concerns about data adequacy for the feasibility and health equity findings due to the small number of studies supporting each finding and the low data richness (i.e., small amount of relevant data) in 3 of 4 studies (1 from Review A[18], and 2 from Review B[20, 21]). Moreover, there were moderate concerns about the relevance of the evidence for all four findings as all of the included studies from Review A and B focused on specific subgroups (e.g,. persons with HIV, new immigrants, children and adolescents), which may reduce the generalizability of the findings to broader populations.

Table A1: Systematic Reviews A and B- Characteristics of Included Studies

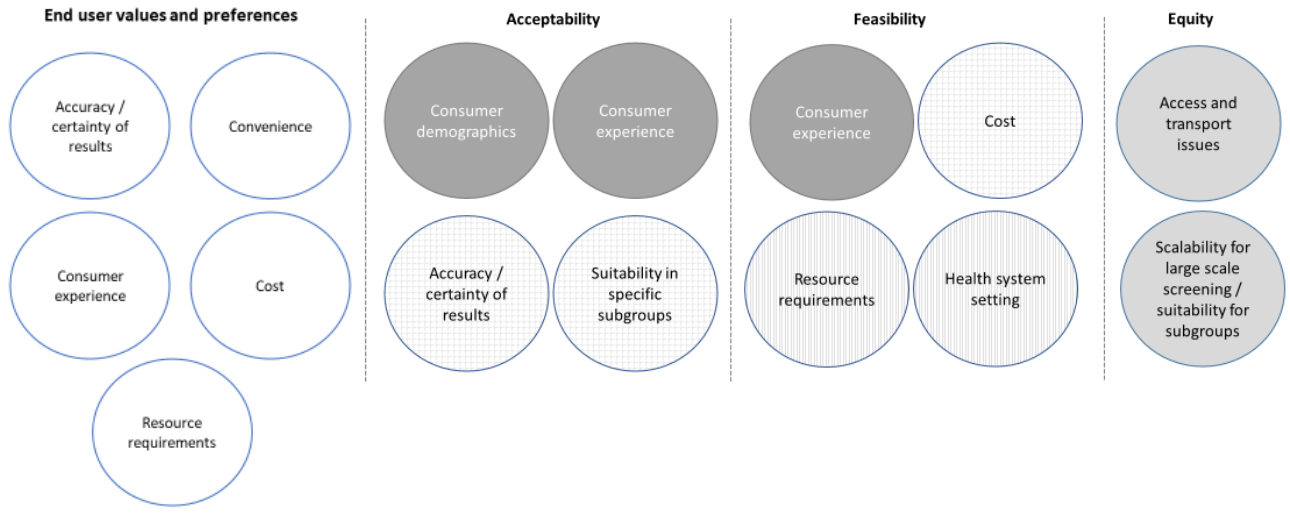
Study ID	Publication type	TBI Tests	Country	Facility Type	Public/ Private	Rural/ Urban	Participants	Programmatic features	Comparators	Data collection	Study duration	Method of data analysis	Conceptual framework	Conflicts of Interest
Review A findings on TBST														
Aksenova 2021 (Review A)	Report (WHO call for data)	Diaskintest	Russian Federation	Not reported	Not reported	Not reported	Parents of children and adolescents undergoing TB screening. Demographic data not reported.	Tests were administered as part of a large-scale screening program for TBI in children and adolescents in the Penza region of the Russian Federation.	No comparator	Not reported	Not reported	Not reported	N/A	Not reported
Review B findings on TST and IGRA														
Kerrigan 2018 (Review B)	Peer-reviewed article	IGRA (QuantiFERON TB Gold In-Tube [QGIT])	South Africa	14 primary care clinics	Public	Not reported	28 HIV patients and 28 clinical care providers (12 TB nurses, 11 professional nurses, 3 counsellors, 1 operational manager, 1 community health worker). Mean age of patients 38.4 years. 50% male patients.	Clinics were participating in the TEKO trial (investigating cost-effectiveness of the IGRA during routine CD4 count blood draws in HIV-infected adults compared to the TST).	TST	Individual interviews using a semi-structured guide	2 months (April – June, 2014)	Thematic content analysis	No formal framework. Interviews followed a semi-structured guide.	None to declare
Pascopella 2014 (Review B)	Peer-reviewed article	IGRA	United States	4 clinics (2 hospital-based; 1 private clinic; 1 health department clinic)	Public	Urban	20 clinic staff. Demographic data not reported.	All clinics took part in the Health Resources & Services Administration (HRSA) quality assurance program for persons with low-income living with HIV/AIDS. Two clinics were part C funded by the Ryan White HIV/AIDS Program of the HRSA.	TST	Individual interviews using a semi-structured guide	2 years (2008 – 2010)	Not explicitly reported (describes using data to inform themes but does not describe the analytic process)	N/A	None to declare
Spruijt 2019 (Review B)	Peer-reviewed article	IGRA (QuantiFERON TB Gold Plus [QFT-Plus])	The Netherlands	5 Public Health Services	Public	Both urban and rural	5 PHS teams (each comprised of at least 1 Medical Technical Assistant, 1 TB nurse, 1 TB physician). No demographic data reported.	Public Health Services focused on treating new immigrants (not applying for asylum) from non-EU countries with a TB incidence of >50/100,000 population.	TST	Semi-structured group interviews	6 months (March – September, 2016)	Thematic analysis (although not explicitly reported; describes development of a coding scheme refined through author agreement)	Levesque's conceptual framework for access to care	Qiagen provided the IGRA kits free of charge but did not contribute to study design or analysis.

3.1.4.5 Review Findings

Our qualitative evidence synthesis from Work Package A draws on the findings from both Systematic Reviews A and B. Overall, there were limited data available hence the decision to combine findings from both Reviews, with Review A on TBST retrieving only one relevant data report, and Review B on TST and IGRA retrieving only three papers.

The findings identified factors contributing to end-user values/preferences, acceptability, feasibility, and health equity impacts of TBST, IGRA and TST, corresponding to our four objectives. We present each finding under each of the four objective areas of end-user values/preferences, acceptability, feasibility and equity, along with the underlying themes and representative quotes for illustration. Further details on the coding framework used can be found in Appendix C. Overall, there were three types of themes emerging: people-related factors, test-factors, and contextual factors. Note people-related factors include both consumers and providers, however as described below, the review findings only reported on consumer perspectives or health providers reported on views of consumers. Figure A2 outlines themes and subthemes organized by objective area.

Figure A2: Themes and subthemes identified from Systematic Review A and B



Findings from Review A and B. Note shading of the circles correspond to the different types of subthemes for acceptability and feasibility: Dark solid grey refers to people factors; square patterns test-factors, and striped contextual factors. The unfilled circles refer to end-user values and preferences, and light solid grey refers to factors relating to health equity. This coding will be used throughout Work Packages A and B findings.

3.1.4.5.1 End User Values and Preferences for TBI Tests

End-user values and preferences for TBI tests were evaluated by all four studies in Review A and B[18,19,20,21]. One study focused on values and preferences for the TBST[18], and 3 studies looked at the IGRA and TST[19,20,21]. End user values and preferences for TBI tests were accuracy, convenience, positive patient experience, cost, and resource requirements, and are described in greater detail below. The GRADE-CERQual assessment revealed moderate confidence in the evidence (see Appendix B).

All 4 studies found that the accuracy of TBI tests affected user preferences[18,19,20,21]. Parents in the Russian Federation were more motivated to perform testing in 92% of cases due to the high accuracy of the Diaskintest (TBST)[18], indicating a preference towards the test, although no comparisons were made to other TBI tests. In 3 studies comparing the IGRA and TST, test accuracy was found to affect user preferences, although results were mixed[19,20,21]. In 2 studies, the IGRA was

preferred to the TST due to greater accuracy[19,21]. In these studies, the TST was viewed as more subjective and associated with a greater false-positive rate than the IGRA. However, one study reported that concerns around the accuracy of the IGRA reduced its desirability with clinic staff[20]. Concerns were driven by the perception that the IGRA was less accurate in HIV-infected adults, and that there was a high frequency of indeterminate results from the IGRA, although this was not supported in the clinic data where only 1% of indeterminate results were received[20].

> *"Sometimes we're not sure if they measure it correctly, or if the TST was done correctly, did they take enough, was it inserted sub-dermally? There can come human errors as well. With blood, I think the result is the result." (South Africa)*

Convenience was found to affect test preferences in all 4 studies[18,19,20,21]. Although limited data was provided, the Diaskintest was reported to be convenient for use in large scale screening programmes[18]. Preferences for the IGRA and TST were both negatively impacted by convenience factors, including long clinic wait times, finding transportation to clinics, and clinic crowdedness in patients living with HIV in South Africa[19]. However, providers in 3 studies reported a preference towards the IGRA as only one clinic visit was required for the test[19,20,21]. Providers reported an issue with the TST was losing patients to follow up which prevented reading their test result.

> *"There was a consensus that using the blood draws would help overcome many logistical and follow-up problems, indicating its ease compared with TST." (South Africa)*

> *"Some PHS staff said that LTBI screening with only IGRA would be a more efficient pathway: the client only must visit the PHS once instead of twice." (The Netherlands)*

Positive consumer experience was another end-user value that affected TBI test preferences in 1 study[19]. The TST was reported to be uncomfortable for consumers:

> *"Sometimes it's itching and irritating for the patients, TST." (South Africa)*

The only other consumer experience factors were transport and wait time which are already covered above under convenience. Across 3 studies, test cost was found to affect end-user preferences for TBI tests[18,20,21]. The high cost of IGRAs reduced their desirability comparative to the TST in healthcare providers in the United States and the Netherlands[20,21]. The low cost of the Diaskintest was reported to improve its suitability for use in large scale screening programmes[18].

> *"According to clinic staff, IGRAs, while desirable, were of limited utility because of their high cost…" (United States)*

Lastly, resource requirements of TBI tests affected user preferences in 2 studies[19,20]. Both the IGRA and TST require highly trained staff for administration, which can create issues in resource-limited settings where there may be trained staff shortages[19]. Having sufficient supplies of test equipment was also reported to affect preferences for both the IGRA and TST in resource-limited contexts[20].

> *"We are not maybe trained, the Sisters are not well trained… that can be another challenge, because maybe they will wait for another one [that is trained]. That Sister is not here, so there's nobody to do the TST." (South Africa)*

3.1.4.5.2 Acceptability of TBI Tests

All four studies from Review A and B explored the acceptability of TBI tests. One study investigated the TBST[18], and 3 studies focused on the IGRA and TST[19,20,21]. Evidence on the acceptability of TBI tests was graded to have moderate confidence, based on the GRADE-CERQual assessment (see Appendix B).

People factors

The acceptability of the TST was affected by consumer demographics (e.g., age) and aspects of the consumer experience (e.g., discomfort). Where providers were included, providers reported their views about TSTs based on the consumer perspective, rather than the use of the test in their practice. Limited data were available in both adults and children factors, however, in 1 study in a middle-income country (South Africa), the TST was reportedly used in both children and adults[19]. Acceptability of the TST was reduced by test discomfort. In this study by Kerrigan et al. conducted in South Africa, a healthcare provider reported[19]:

"Sometimes it's itching and irritating for the patients, TST." (South Africa)

No studies evaluated people-factors in relation to the TBST or the IGRA.

Test Factors

Acceptability was affected by test-related factors, including the accuracy of the test, the certainty of results, and the suitability of the test for use in particular subgroups (e.g., children and adolescents, persons with HIV, needle-phobic individuals).

TBST and IGRA were perceived as having greater specificity than the TST[18,21]. The TST was described as having a high rate of false positives; thus the IGRA was deemed preferable in terms of specificity[21]. From Review A, the high specificity of the TBST was found to improve the motivation of parents to perform TB testing for their children, indicating good acceptability in the population[18], see quote below (Russian Federation):

> *"A positive result of the test with the RTA (Diaskintest) is consistently recorded in less than 1% of subjects (2015- 0.97% and 2016- 0.86%). A differentiated approach to the prescription of preventive treatment in groups at risk for the development of tuberculosis significantly increases the motivation of parents to perform testing in 92.0% of cases." (Russian Federation)*

Perceptions about the subjectivity and objectivity of the results reading process affected acceptability of the TST and the IGRA. In Kerrigan et al., staff in an HIV clinic reported preferring the objectivity of the test results produced from an IGRA to the more subjective results read by a human with the TST[19]:

> *"We're all trained but sometimes we're not sure if they measure it correctly, or if the TST was done correctly, did they take enough, was it inserted sub-dermally? There can come human errors as well. With blood, I think the result is the result. You can see it in the blood."(South Africa)*
>
> *"I like black and white. So if it is done with a blood test, blood has never lied to anybody. If it is done with a blood test, I'm sure it is going to be much easier." (South Africa)*

On the other hand, healthcare providers in two studies (done in the United States and the Netherlands) reported lower perceived certainty in the results of the IGRA, which reduced acceptability[20,21]. In one case, this perception was not reflected in the clinic data[20]:

"[Staff] had a perception that IGRA produced an unacceptably high frequency of indeterminate results. However, study patient data revealed that only 1% of patients tested with IGRA had an indeterminate result."(United States)

A TB physician in a Public Health Service in the Netherlands reported[21]:

"IGRA results between 0.35 and 1.00- the so-called grey area- can go either way. You can repeat the test, or you can just say: no, it is not an LTBI if anamnesis[sic] factors are not in favor. So, basing LTBI diagnosis solely on an IGRA result of 0.35 is not enough." (Netherlands)

In terms of subgroup acceptability, the TST was reported to be acceptable for use in adults and children[19], while the TBST (Diaskintest) was found to be acceptable in children and adolescents as part of a large scale screening of TBI[18]. The IGRA was perceived to be less acceptable for use in persons with HIV due to the perceived uncertainty of results[20]. Moreover, the IGRA may be less acceptable to individuals with needle phobias as reported in 1 study[19], though the authors concluded that the IGRA might still be acceptable if completed as part of a routine blood draw:

"Patients indicated that while some people are not fond of blood draws, that since the CD4 count blood draw was happening anyway, it would be well received." (South Africa)

Contextual Factors

No contextual factors were found to affect acceptability across the 4 studies.

3.1.4.5.3 Feasibility of TBI Tests

Three studies – all from Review B - investigated the feasibility of TBI tests[19,20,21]. These studies evaluated the IGRA and TST. None of the included studies evaluated the feasibility of the TBST. Based on the GRADE-CERQual, there was moderate confidence in the evidence (see Appendix B for more details on confidence rating decision).

People factors

Aspects of the individual experience were found to affect the feasibility of the IGRA and TST. A common feasibility concern for the TST, based on findings from Review B, were that consumers were required to visit the clinic a second time within 72 hours to read the test result[19,20,21]. Kerrigan et al. described how in the North West Province of South Africa, many consumers experienced difficulties with visiting a clinic due to the long journey and wait time involved, as well as issues with accessing transportation[19]. Similar concerns were echoed in HIV clinics in the United States, although only 4% of consumers did not appear for a second appointment in this setting[20]. Indeed, Spruijt et al. described how a distinct advantage of the IGRA is that consumers only need to visit the clinic once, thus creating a more efficient care pathway[21].

A consumer from South Africa[19]:

"Sometimes you'd find that you have to queue from 4 o'clock in the morning, you go out here 4 o'clock afternoon. You find that twelve hours you're in the clinic. So it's like useless and what, but it's about your life. You must stay. So sometimes it's punishing our mind, we're hungry, we're thirsty. There's only this place, the clinic. There's no shop around. You must travel long so that you can't find the people behind you being upfront of you." (South Africa)

A quote from Spruijt et al. describing a Public Health Service in the Netherlands[21]:

"Some PHS staff said that LTBI screening with only IGRA would be a more efficient pathway: the client only must visit the PHS once instead of twice, and the PHS would save time in extra consultations because of the high rate of positive TST." (Netherlands)

Test Factors

Cost was another factor affecting TBI test feasibility. In particular, the IGRA was reported to be less feasible for targeted screening in at-risk groups due to its high cost compared to the TST[20,21]. These sentiments were reported from healthcare providers in a Public Health Service for new immigrants from high TB incidence countries in the Netherlands, and from healthcare providers in a series of publicly-funded HIV clinics in the United States.

A quote from Pascopella et al. describing staff attitudes in clinics for people living with HIV[20]:

> *"According to clinic staff, IGRA, while desirable, were of limited utility because of their high cost and frequency of indeterminate results." (United States)*

A quote from Spruijt et al. on the views of clinic staff in a Public Health Service for new immigrants from high incidence countries[21]:

> *"PHSs consider IGRA to be expensive at this moment (prices in the Netherlands range from 50-104 euros for a single IGRA test compared to 27 euros for a single TST test). They expected with an increased demand for IGRA, laboratories should be able to offer them cheaper." (Netherlands)*

Contextual Factors

Aspects of the health system setting (e.g. the ability to screen alongside other tests, aspects of the local healthcare system), and resource requirements (e.g., trained staff requirement, equipment or sample storage requirement) were found to affect the feasibility of the IGRA and TST and will be discussed in greater detail below.

A key advantage of the IGRA was the ability to screen for TBI alongside other tests during a routine blood draw. Kerrigan et al. reported that this would further help to avoid the logistical challenges involved with a return visit to read a TST result in North West South Africa[19]:

> *"Almost every provider and patient interviewed reported being open to the idea of screening for latent TB at the time of the blood draws for CD4 count. Providers in particular felt this would streamline the latent TB screening process, given that due to heavy patient loads and a lack of personnel, people were simply being missed for screening in a given visit. Providers also discussed the significant challenges they faced in ensuring that their patients returned to have the TST read. There was a consensus that using the blood draws would help overcome many logistical and follow-up problems, indicating its ease compared with TST." (South Africa)*

However, one consumer in this study expressed concern that this might delay detection of TBI given the low frequency with which they would typically have a blood draw:

> "*While there was general consensus on this issue, one patient did relay a concern about timing, as many patients may only have a CD4 blood draw 1–2 times a year. So while this patient thought screening the CD4 blood draw for TB was a good idea, he wouldn't want that to delay finding out he had TB, as he would want to know that information as soon as possible. This issue was also raised by clinical providers, and a few in turn mentioned that screening the blood for TB could potentially complement, as opposed to replace, the other TB screening methods given that CD4 blood draws occurs generally every 6–12 months."(South Africa)*

Public funding and supply of the TST was found to improve feasibility in North-Western South Africa, although there were limited data available in support of this factor. A healthcare provider in Kerrigan et al. reported[19]:

> *"It's a challenge, I don't want to lie. It's a big challenge… With recently they told we must do the TST. Yes. And then maybe the challenge… now we don't have a challenge as such because at least they are supplying with the, that Mantoux, the test itself." (South Africa)*

In terms of resource requirements, both the IGRA and TST were found to require trained staff which influenced their feasibility in resource-limited settings. In publicly-funded HIV clinics in the United States, staff reported that having access to more trained staff with TB expertise and knowledge of cultural competency would serve as a facilitator to TB testing[20]. In North West South Africa, concerns were raised about a shortage of trained staff to administer the TST, which was acknowledged to create further delays and inconvenience for consumers[19]. Another issue raised by the clinic staff in this study was that they received guidance that the TST must be used without receiving any training for the procedure:

> *"There were some that were simply unclear about what they should be doing. Many discussed having just received guidance that they were supposed to be using TST now, but they were not sure exactly how to operationalize this new procedure." (South Africa)*

However, some of the clinics in the study had implemented training for their nurses in the TST procedure:

> *"In response to challenges implementing the TST, some providers reported their clinics started adopting innovative practices. For example, in one clinic the management ensured that all nurses were trained on how to place and interpret TST, and the clinic rotates a nurse responsible for TST and initiating IPT daily. By having all nurses trained about the TST and IPT, it allowed for patients to return to have the TST read at any time and it is not dependent on a particular nurse being on duty. In other clinics, patients coming back to have the TST read were not required to wait in line in order to minimize loss to follow-up." (South Africa)*

Another contextual factor that was found to affect test feasibility was the availability of necessary equipment or sample storage facilities to support test procedures. Only 1 study reported on this aspect[20]. Staff from an HIV clinic in the United States reported that having access to sufficient clinic supplies would further facilitate TB screening with IGRA and TST.

3.1.4.5.4 Health Equity Impact of TBI Tests

All four studies indirectly discussed the health equity impacts of TBI tests. One study from Review A focused on the TBST[18], and the 3 studies from Review B focused on the IGRA and TST[19,20,21]. A GRADE-CERQual evaluation based on data from Review A and B found low confidence in the evidence (see Appendix B for more details).

All of the themes describing health equity related to contextual factors around TB testing. The health equity impacts of TBI tests were contributed to by access and transport issues surrounding the tests (e.g., amount of staff in low-resource settings, health environment crowdedness, long journey to the clinic, return visit to the clinic required to read the result, specialized staff training, transportation to the clinic, individual wait time), the scalability of the tests for large scale screening programs, and the suitability of the test in particular subgroups (e.g., children and adolescents, people with HIV, and needle phobic individuals).

In terms of access and transport issues, the return visit required to read the TST result was focused on as a major accessibility issue in 3 studies[19,20,21]. This included at-risk groups in both middle- and high-income countries (South Africa, United States). Kerrigan et al., which focused on an HIV clinic in a middle-income country, documented how many consumers experienced challenges with transportation and wait times in understaffed clinics, serving as a barrier for return visits[19]. Challenges with an adequate number of specialized staff to administer the TST and IGRA were documented in publicly funded services in high- and middle-income countries[19,20].

Only 1 study evaluated the feasibility of TBI tests for use in large scale screening programs. The TBST (Diaskintest) was reported to be convenient and effective for use in mass TB screening with children and adolescents in the Penza region of the Russian Federation[18].

Regarding suitability to subgroups, the IGRA was reported to be less suited for use in needle-phobic individuals[19], and may have reduced performance in persons with HIV compared to the TST[20], reducing accessibility in these groups.

A quote from Pascopella et al. on the use of the IGRA in HIV clinics in the United States[20]:

> *"Many key informants were concerned about IGRA test performance in HIV-infected populations, and had a perception that IGRA produced an unacceptably high frequency of indeterminate results." (United States)*

In other studies, both the TBST and the TST were reported to be suitable for use in children, improving test accessibility in this demographic[18,19].

A quote from Aksenova on the use of the Diaskintest in mass TB screening[18]:

> *"Diaskintest is highly effective in detecting tuberculosis in children and adolescents during a mass examination." (Russian Federation)*

3.1.5 Summary of Work Package A findings from the Systematic Reviews

Overall, there were limited data from the reviews on TBSTs, with only 4 studies identified. Importantly, these studies were conducted in very specific healthcare settings and populations, limiting our confidence in the data quality and the generalisability of these findings. Most of the data available related to TST and IGRA. Despite the limited data available, we identified key factors relating to our four objectives of end user values/preferences, acceptability, feasibility and health equity. Important end-user values affecting test preferences included accuracy, convenience, positive patient experience, cost, and resource requirements. The data highlights the importance of considering people-related factors (e.g. consumer demographics and experiences) and test-factors (e.g. accuracy) for test acceptability. In terms of feasibility, people factors (e.g. need for return visits), test factors (e.g.cost), and contextual factors (e.g. ability to screen alongside other tests), were highlighted. Issues of health equity relating to contextual factors around TB testing, such as access and transport, and the health environment. These findings were used to inform the discussion guide used for Work Package B and for the design of the Discrete Choice Experiment.

3.2 Work Package B-Qualitative user interviews

This activity intends to answer following objectives:

Objective B: To conduct in-depth interviews to obtain data on views and perspectives of TBST compared to existing TBI tests during qualitative interviews with end users and key informants.

Objective C: To evaluate any potential impacts of TBST on health equity

A qualitative interview study was conducted with 20 participants who were individuals living with TB or providers involved in TB testing and /or care of individuals living with TB. Participants were recruited from the research team's global networks across high and low-middle income countries, with support from WHO. Participants were approached through organizations/groups as below:

- TB Treatment Action Group
- TB Civil Societies
- Members of the WHO Guideline Development Group for TBST
- Global Coalition of TB Activists (GCTA)
- Stop TB Partnership
- WHO TBI Task Force

3.2.1 Methods

Recruitment methods

Participants were approached by an email invitation (Appendix D), which was sent via mailing lists or via mutual contacts through the above-mentioned organizations. Those who responded to the initial invitation email were emailed a participant information sheet and consent form to complete and return. Following this, an interview time/date was arranged with the lead interviewer (KL) or objective lead 4 (AC).

A total of 20 semi-structured interviews were conducted with a diverse range of clinicians, laboratory staff, programme officers, and individuals living with latent TB from patient organizations, with the aim of understanding their experiences of using TBI tests, including the IGRA, TST, and TBST.

Inclusion criteria of participants:

- Any participant aged 18 years or over with personal clinical or working experience either as a person living with TB or health professional or other healthcare provider involved in the diagnostic process of TB (e.g. laboratory technician);
- Spoken English at a level sufficient to provide informed consent and partake in the study, or access to an interpreter who was able to provide live translation over Zoom;
- Access to internet or to a device (phone, tablet, laptop etc) with stable internet connection for conducting the interview.

Sampling

Participants were purposively sampled based on their role, and country income level, and recruited based on convenience sampling. The purposive sampling approach was informed by insights on potential inequity concerns noted through the systematic review (work package A) to enable these to be explored further. Due to the location of these participants internationally, interviews were conducted over Zoom video conferencing software.

Discussion guide

The interview topics were informed by the systematic review findings (see systematic review protocol), and respondents were asked to share their experience of using the TBST in comparison to the previously WHO recommended tests for TB infection detection such as the TST and the IGRA[22]. The questions were tailored depending on whether the respondent was an individual living with TB or healthcare provider.

The interview topics included:

1. Current approach to using skin tests as part of the TB diagnostic process;
2. End-user values and preferences with regard to TB testing, from their previous experience with skin-based tests – focusing on ease of use and process for TBST, IGRA, and TST;
3. End-user values and preferences considering the challenges and benefits of skin-based tests, particularly TBST, which test(s) they would prefer and reasons for the preference;
4. Feasibility of implementing TBST in current practice;
5. Overall acceptability of TBST- burden, attitude, understanding, test efficacy;
6. Potential impacts on health equity and policy.

Interview procedures and analysis

Interviews were audio-recorded, then transcribed by an independent third-party professional transcriber, and coded in NVivo 12. Additional field notes were made by the interviewer. At least two members of the research team (KL, AC) reviewed the transcripts and field notes on the topics, discussed the coding framework, revised the framework iteratively and collated them into themes, using an inductive thematic approach.

Owing to the use of Zoom for interviews, we could not triangulate interview data with other evidence commonly collected through ethnographic approaches (such as multiple interviews and informal conversations at the same facility, observations or site visits). However, we supplemented the data gathering with data from the systematic review (work package A) that was conducted prior to these interviews (see systematic review work package). We also obtained respondent validation of the interview data to check the findings and used rich descriptions when analysing the generated themes for trustworthiness.

An ethics exemption was obtained by the lead investigator from the New Zealand Health and Disability Ethics Committee (HDEC) (21/STH/204). Additional ethics review was sought from the Auckland Health Research Ethics Committee (AH23444).

3.2.2 Results

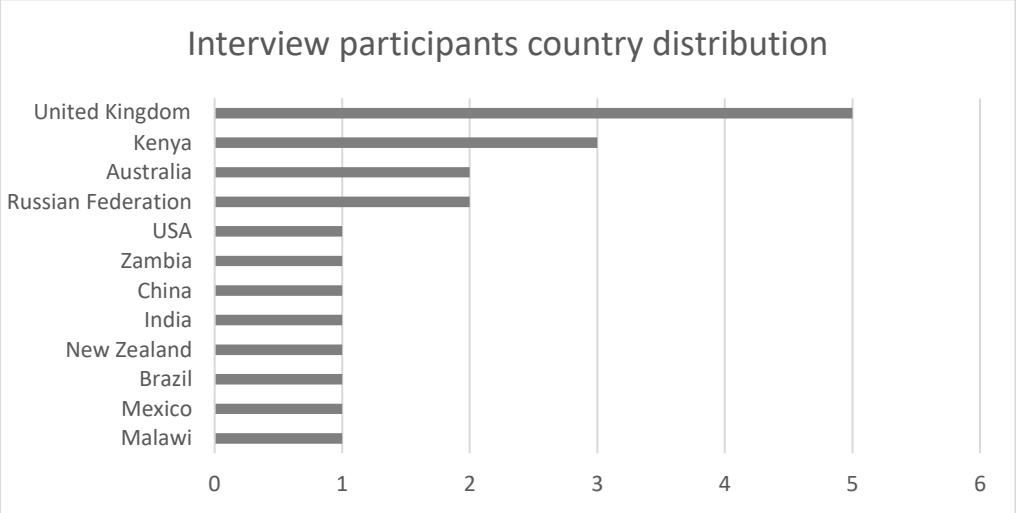
Participant characteristics

Twenty participants took part in the interview study. Participants were 13 TB healthcare providers and 7 TB consumers. Table B1 outlines the participant characteristics. Healthcare providers included program executives and decision-makers, public health practitioners and advocates, physicians, researchers and lab technicians, and a nurse. Participants were from a range of high- and low/middle-income countries, as outlined in Figure B1.

Table B1. Participant Characteristics

Healthcare Provider	**13**
High-Income Country	5
Low/Middle-Income Country	8
Roles	
Program executive or decision-maker	3
Public health practitioner/advocate	2
Physician	4
Researcher or lab technician	3
Nurse	1
Consumer	**7**
High-Income Country	4
Low/Middle-Income Country	3

Figure B1. Participant country distribution (n=20)



3.2.3 Key Findings

The findings from Work Package A were used to inform the coding framework and analyses undertaken in Work Package B – as such, the same subthemes of people, test, and contextual factors were used to explore acceptability and feasibility of the tests, with health equity and end-user values and preferences as separate subthemes. The full coding framework used can be found in Appendix E. As before, the same shading scheme will be used in the figures to illustrate themes and subthemes for acceptability and feasibility: Dark solid grey refers to people factors; square patterns test-factors, and striped contextual factors. The unfilled circles refer to end-user values and preferences, and the lighter grey refers to equity factors.

3.2.3.1 End User Values and Preferences

End-user values and preferences for TBI tests were described in five subthemes that included test accuracy, convenience, positive patient experience, cost, and resource requirements. All subthemes were discussed by both healthcare consumers and providers (see Figure B2).

Figure B2: End-user values and preferences

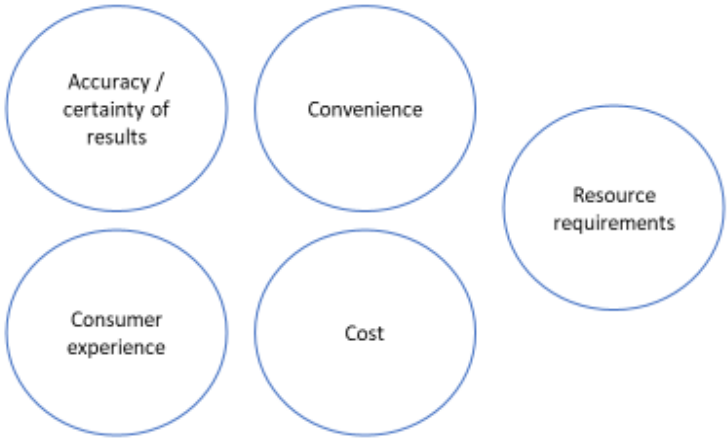


Figure describing the factors influencing end-user values and preferences

Accuracy

Both healthcare consumers and providers were found to value test accuracy. TBSTs and IGRAs were valued for having greater specificity and sensitivity compared to TSTs. End-users associated the TBSTs and IGRAs with many downstream benefits as a result of greater accuracy. These included saving hospital resources and long-term expenses, avoiding unnecessary treatment and psychological stress, and improving rates of diagnosis and treatment evaluation. The TSTs were perceived to have lower accuracy and were associated with downstream challenges by end-users. These included greater expense and psychological stress.

> *"The big issue I think is the false positives with the current tests that we use and that then leads to a whole chain of events which I think is very expensive and stressful for people because if you're told you might have TB or you probably have got TB and then you've got a whole set of tests, if the novel test is more accurate and it comes back and it says no you didn't, that saves a lot of stress and all the hospital resources and scans et cetera."* (ID002, HIC, Consumer)

Convenience

Healthcare consumers and providers valued the convenience of TBI tests. IGRAs were perceived as more convenient than TBSTs and TSTs because they do not require a return visit to read the result. End users were concerned about loss to follow up with skin tests. However, portability was a strength of the skin tests as patients do not necessarily need to visit a healthcare setting to receive the test. Many participants talked about how barriers to seeking TBI tests were due to inconveniences such as a long journey to the clinic, finding transport, and experiencing long delays at healthcare facilities. These convenience factors are avoided with portable tests that can be administered in the community as part of large-scale screening programs, such as TBSTs and TSTs.

> *"Say if you take the sample there and you require him to come after two days, let's look around the issues concerning transport, they would need money. So the element of money should be considered… this you have to go for 48 hours which means the one who goes longer distances then we have to consider transport elements and remember we are trying to think of the catastrophic costs. For these people to come to the hospital, they have spent some money and maybe they have come with their guardians, they have to pay for that. So I don't know how this will be covered."* (ID004, LMIC, Provider)

Positive patient experience

End users were found to value a positive patient experience while receiving a TBI test. This included avoidance of psychological (e.g., anxiety, stigma, stress) and physical consequences (e.g., pain, a welt on arm), and discomfort. While the IGRA was more anxiety-provoking for needle-phobic individuals, the TST was associated with stress from false positives and stigma in some communities (e.g., in rural India). Some participants reported feeling unease from watching a 'live' test result unfold before them with TSTs. Moreover, the TST was found to be uncomfortable. TBSTs were expected to be similar to the TST in terms of comfort and physical consequences, although they would avoid the stress of false positives.

> *"I had [a TST] many years ago and I recall it being quite uncomfortable… not invasive, that's over, that's too strong a word but yeah, quite an uncomfortable test."* (ID001, HIC, Consumer)

Cost

Cost was another aspect of TBI tests that end users valued. Although it was acknowledged that typically governments or healthcare providers would bear the costs of testing, test cost was still discussed by consumers. The TBST and TSTs were viewed as more affordable than the IGRA and thus were deemed as more appropriate for resource-limited settings and community-based screening programmes, as well as beneficial for cost-saving.

> *"Honestly, I love IGRAs. I think they should be the standard of care but they are expensive so trying to be honest and trying to be realistic in countries where money matters, I think this new test can give us a change to improve the programmatic strategies."* (ID009, LMIC, Provider)

Resource requirements

The resource requirements of TBI tests affected end-users preferences towards them, particularly in resource-limited settings. End users discussed how IGRAs require more resources to administer than the TBSTs and TSTs, affecting preferences for use in some settings. IGRAs were noted to require a laboratory, hospital or clinic, phlebotomy and sample storage equipment, highly trained staff, and patient transportation. TBSTs and TSTs require less resources than IGRAs, although trained staff and test equipment are still required, and clinics and patient transport are needed in some instances.

> *"I feel like the good thing about [the TBSTs] is you don't need much lab infrastructure, I feel like the need to draw blood is a barrier for IGRAs."* (ID017, HIC, Provider)

3.2.3.2. Acceptability

People-related factors

There were 7 subthemes relating to people factors for acceptability of TBI tests (Figure B3). These included consumer experience, consumer demographics, perceptions of skin tests, knowledge, familiarity, and provider-consumer communication, which were discussed by both consumers and

providers. Consumers also discussed how useful they perceived the test to be for themselves in terms of diagnosis and management of TBI – whereas this perceived usefulness of the test from a consumer perspective was not discussed by any health provider participants.

Figure B3: Acceptability – people-related factors

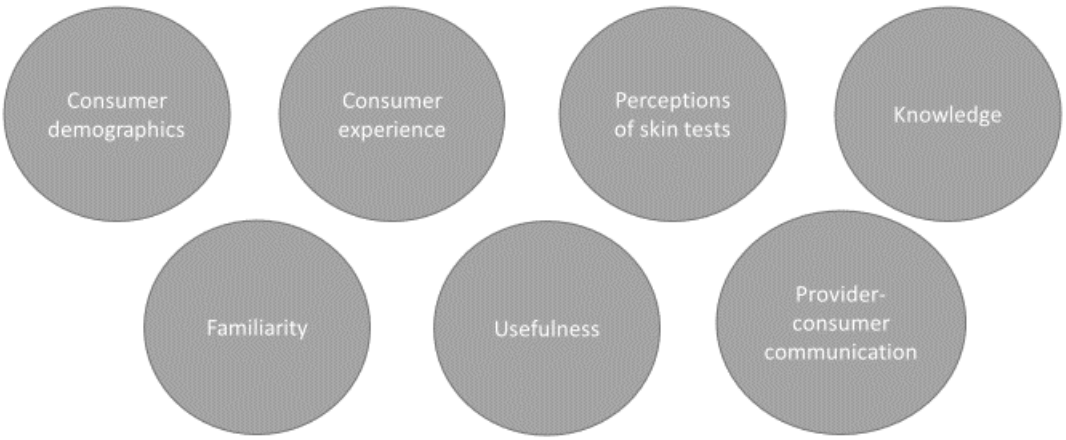


Figure describing the people factors influencing test acceptability

Consumer Experience

Several aspects of the consumer experience were found to affect the acceptability of TBI tests. All 3 TBI tests were reported to involve some discomfort during the procedure, including a slightly sore arm. Although, one consumer reported not finding blood tests uncomfortable.

> *"I had one many years ago and I remember it being quite uncomfortable… not invasive, that's over, that's too strong a word but yeah quite an uncomfortable test."* (ID001, HIC, Consumer)

In terms of the physical consequences of the tests, both the TBST and TST involve a welt on the arm. In rare cases, there may also be significant reactions from the intradermal injection. Moreover, the skin tests were reported to be painful tests at times due to the intradermal injection. Other participants did not perceive any physical consequences of TBST. No physical consequences were reported for the IGRA.

> *"…that's a painful test, you know, in comparison with the IGRA, so that skin test, it's very, it's a painful test, it's intradermal injection, it's very painful."* (ID020, LMIC, Provider)

Mixed feedback was received regarding the psychological consequences of the tests. Some participants anticipated that there would be no psychological consequences of the TBST, while others anticipated that there may be additional anxiety associated with TBST due to these being a new test. For TBST and TST, anxiety may also be felt around the intradermal injection and the uncertainty of what the welt on the skin means. Some consumers reported feeling uneasy from the 'live' reading that took place in front of them, an experience that is avoided with the IGRA.

> *"We also found perspective from the patient that maybe people at the beginning, they worry that it's painful and they scared of injection, but once they (indistinct), they like it very much because you can see physically the test result on your arm… at the beginning, many people worry about it, they worry that it would be unsafe, severe adverse reactions, and*

> *people would die if you injected PPD. Of course it's not the case, there's so many evidence in the world (indistinct) but they still worry, it took us a year to prove that it's safe… It's new things, injecting, people might even think that they are like, when you conduct clinical trial is the mice, they may be considered they are mice for the research."* (ID013, HIC, Provider)

> *"Psychologically, I would say yes, I would say that you've got, you can see and I, I mean I would totally go and Google and I would try to be diagnosing what was going on, and I would imagine that I guess you get some people who wouldn't care, and so I imagine that a fair few people would actually, if they were interested in the outcome of the result, they might actually be concerned about what was going on in their arm and that might cause, psychological distress might be a little bit too strong word but a little bit of unease and uncertainty and apprehension I would say, because you can see what's going on rather than it being a blood test taken away and it's gone, it's in a laboratory, you have no idea what's going on."* (ID001, HIC, Consumer)

Moreover, in some cultures there is a stigma associated with the welt on the arm that a skin test may produce, and which an IGRA test can help to avoid. This stigma was raised as a serious concern in rural Indian communities:

> *"So some extent stigma can also be doing it like, so for these other tests they silently go to the lab and they'll give the sample and come back. Here, this line or this mark is going to be for two days, so have attached, stigma attached to TB, so it could be a thing."* (ID005, LMIC, Provider)

Unique to the TST, the lower accuracy of this test could also prompt unnecessary stress for consumers from false positives:

> *"I think [the TBST are] more beneficial, it's less strain on your mental health… [the existing tests are] probably not as reliable, not efficient (indistinct)… are they up to scratch, the results? There might be something wrong there."* (ID006, HIC, Consumer)

Anxiety was also reported as a consequence of the IGRA for consumers who are particularly uncomfortable with needles:

> *"Some patients may be absolutely, they're completely terrified of having blood test but a skin test might be something they can cope with or who knows, I don't know." (ID012, HIC, Provider)*

Some participants also reported that a hospital environment prompted feelings of anxiety. Although this environment is more commonplace for administration of the IGRA, it is possible that TST and TBST administrations may also take place in this environment.

Another aspect of the consumer experience that affected the acceptability of TBI tests was their degree of invasiveness. The IGRA was perceived to be more invasive than the TBST and TST. Moreover, certain cultures may be less comfortable with the degree of invasiveness that a blood sample requires.

> *"I feel like the need to sort of draw blood is a barrier for IGRA… it's certainly a barrier when you've got to draw blood versus doing it via skin. I mean it's not, I don't, it is non-invasive to a certain degree but there is, it's not like a breath test or peeing on a stick."* (ID017, HIC, Provider)

> *"Any test which is subcutaneous is fairer than intravenous because the blood test makes you draw out blood isn't it?... Africans have some myth about people taking their blood."* (ID008, LMIC, Consumer)

Lastly, the complicated process involved with skin tests was perceived to reduce acceptability in consumers with low health literacy.

> *"First of all the test is a very complex thing, though to the provider it might be easy but from the patient's perspective it is difficult. Like as of my understanding, patient should not have, should not rub on the injected areas and all those certain precautions need to be taken up. So to what extent patient keeps those things into mind is one end and this might require two visits to the test clinic. So from the patient perspective, I feel like it's a bit complicated thing but not too much complicated if proper counselling is to be given to them."* (ID005, LMIC, Provider)

Demographics

Concerns were raised about the suitability of the IGRA and TBST for use in people with darker skin tones. One consumer described challenges they had faced with healthcare providers trying to find their veins for blood tests due to their darker skin tone, which reduces the acceptability of the IGRA in this population. However, concerns were also raised about whether accurate readings could be guaranteed for the TBST for persons with darker skin tones.

> *"Sometimes they can't find my vein so it's difficult to do blood tests… I know in most cases it's easier to get a white person's vein than a black person. So to me, I'll go for the skin one and somebody else, so it's not one size fit all."* (ID007, HIC, Consumer)

> *"Have they looked into the reading of the results with different skin types, different skin colours and can they guarantee that that's accurate across all skin types, would be my other thought?"* (ID001, HIC, Consumer)

All three tests were described as challenging for individuals with a needle phobia, with either a blood draw or an intradermal injection involved:

> *"Some patients also do not like injections. There are a number of patients who does not like injectables, so the moment you are inserting the needle, they feel very psychologically uneasy, so it brought a lot of problems."*
>
> *Researcher: "Is that even with the skin test, that they worry about the needle?"*
>
> *Participant: "Yes."* (ID008, LMIC, Consumer):

Similar concerns were raised about placing intradermal injections in children, an experience that can be particularly challenging as it is more involved than a blood draw. Moreover, children were reported to struggle with the itchiness of the TST:

> *"I suppose with children particularly, it can be quite challenging cos obviously it's a needle and it's not a particularly quick injection either, so you can't just sort of plunge it in and out."* (ID012, HIC, Provider)

> *"Well first of all it is kind of difficult to apply to the children because they move and also they tend to scratch the place that we put the TST."* (ID009, LMIC, Provider)

Perceptions of Skin Tests

Some participants viewed the TBST as an improved version of the TST. This perception tended to be driven by knowledge that the TBST were a more accurate test. Conversely, other participants described skin tests as a basic and dated technology, and questioned the utility of introducing another skin test. Some participants expressed a desire for new kinds of technologies for TB testing.

> *"I think that it's good to have this kind of new test but as a researcher, I would like to see a more advanced one because we know that the barrier with the TST is the injection, that people are afraid. So if there is another way like there's now a technique that you have a tab or something you put on your skin instead of injection, something like that, then it would look less painful."* (ID013, HIC, Provider)

> *"If one is going to persist in using skin tests, it certainly has potential great advantages over the tuberculin skin test. My argument is we should be getting away from skin tests and doing something that makes much more sense… This whole conversation is surreal to me, why people are doing this with a 100 year old technology… If it's better than the TB skin test then the issue is replacing that test with this test, that's, I totally understand that and that's a perfectly reasonable thing. We are trying to get away from skin testing period."* (ID014, HIC, Provider)

Knowledge

Most participants expressed having limited knowledge about the TBST. One participant reported that being able to see the evidence around the TBST would help to improve her understanding and perceptions.

> *"There is a need for a lot of public awareness."* (ID011, LMIC, Consumer)

> *"I would be willing to use it if I see evidence that it was better than what I was using before but I wouldn't know until it's tried out."* (ID007, HIC, Consumer)

Familiarity

The familiarity of skin tests for TB testing was reported to improve the acceptability of TBST. Some participants discussed how they were already used to administering or receiving skin tests, thus using a new skin test would not be a challenge for consumers or providers. Other participants talked about how they were more familiar with the IGRA and thus would prefer to continue using that over the TBST.

> *"For us like, well we are so used to have TST around as a standard of care so having a new TST, it's not that really big challenge or change in the actual procedure."* (ID009, LMIC, Provider)

> *"Everybody is very much used to a skin test, initially PPD, so now Diaskintest, no problems at all."* (ID015, LMIC, Provider)

Provider-Consumer Communication

Both patients and providers felt that the acceptability of TBST could be improved through careful communication with consumers around test use. This could include describing the benefits of TBST over alternatives, and providing counseling to address any consumer concerns around the test (e.g., about the meaning of the skin mark).

> *"For me, it's about perception. If from the initial stages, we market this product in a positive way, I think the perception by the outside, the general population will be good, but if for example, people begin to market in a negative way to say this skin test is also giving a problem in terms of once you are tested it leaves you maybe with a scar, it leaves you with this problem… then you will see that the public perception of the tests will be in the negative… I think we need to do a lot of, engage more community-based interventions in terms of promoting the new skin tests so that the public understanding of the new method is accepted."* (ID003, LMIC, Consumer)

> *"So this test to some extent would be an easy test to convince a patient if proper counseling is given to them because it is just an injection we are giving… if you counsel them properly, I don't see any psychological and physical, I don't see anything except the bulge which will be coming because of the test impact or the test reaction."* (ID005, LMIC, Provider)

> *"Certainly the introduction of novel skin tests required a lot of community education work. People needed to be explained about the usage of these tests and once they would be convinced, then this would abolish any potential psychological problem."* (ID016, LMIC, Provider)

Moreover, endorsement of the TBST by healthcare providers and organizations (e.g., the WHO) was expected to improve acceptability.

> *"I would like to think that if it was something that was being genuinely considered by World Health Organization and healthcare across the world as a suitable test, I would like to that that it's a good one, it's worth carrying on with."* (ID001, HIC, Consumer)

Usefulness

A couple of consumers questioned how useful the TBST would be. Interestingly, this did not come up as a finding from any healthcare provider participant. One consumer reported needing evidence that the TBST are better than the TST and IGRA to think that they are useful. Another consumer felt that how useful the TBI tests are in practice depends on the care pathway that follows a positive result.

> *"If it's not broken, why are we trying to fix it? What's the evidence that's saying what we were doing before doesn't, it's not adequate and why are we doing this? I need to see proof of this."* (ID007, HIC, Consumer)

> *"I would say it was pretty useless because someone, like I got a bit of paper that I was given that said yes, you had exposure to tuberculosis, but there was no follow up from it."* (ID001, HIC, Consumer)

Test Factors

Five subthemes described test factors that contributed to the perceived acceptability of TBI tests. These subthemes were discussed by both consumers and healthcare provider participants, and included test accuracy, reliability, time to receive the result, suitability in specific subgroups, and the ability to differentiate between latent and active TB. There were no subthemes that were discussed only by consumers or providers. These subthemes are described in greater detail below (Figure B4).

Figure B4: Acceptability – Test factors

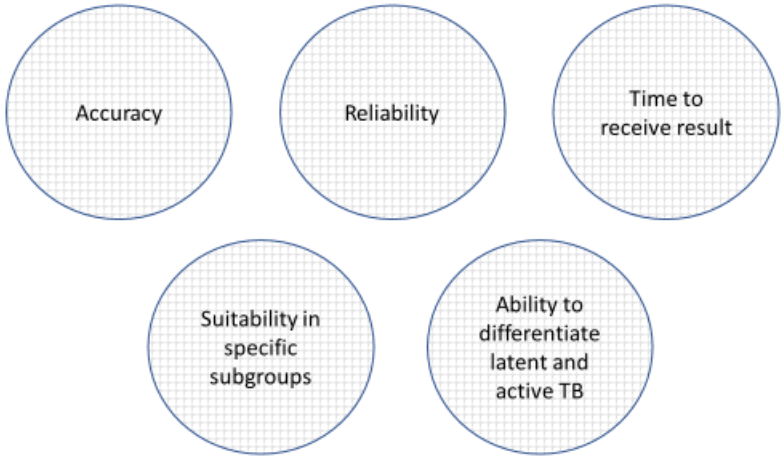


Figure describing the test-related factors influencing test acceptability

Accuracy

A major factor affecting test acceptability was perceived accuracy. TBST and the IGRA were perceived to have greater accuracy than the TST. This was driven by perceptions of greater specificity and sensitivity, and resulted in greater perceived certainty in the results produced by these tests. However, some concerns were raised that similar to the TST, a "boosting" effect may be found with the TBST whereby serial testing involving intradermal injections may increase false-positive results. Moreover, the TBST and TST are more susceptible to human error while reading the results compared to the IGRA.

> *"I guess I don't have a lot of experience with the novel skin tests but I guess my understanding is that it is really, they're sort of equivalent to the IGRA in terms of based on the same kind of antigens so you're essentially getting the equivalent performance in terms of losing the cross-reactivity that you see with the TST with BCG and other non-tuberculosis mycobacterium but in a skin prick form as opposed to blood test… If I was choosing between a TST and a novel skin test, then I would choose the novel skin test due to the improved specificity really."* (ID019, HIC, Provider)

> *"The measurement involves a ballpoint pen with free flowing ink and you sort of make the mark on the person's arm until you start to feel a little bit of resistance and then you stop and then you do that on the other side and you stop, and then you measure the distance between those two stops and that's probably the most accurate way to interpret the induration. But as you can imagine, I mean thank about doing that and think about the issue, how hard that would be and the sort of saying well if it's 9mms they're infected, and if it's*

> *11mms, they're not infected. Really? Seriously? Do you really believe that?"* (ID014, HIC, Provider)

Many participants discussed the downstream benefits of the TBST's greater accuracy in comparison to the TST. These included preventing psychological stress from false positives, improving rates of diagnosis, improving the evaluation of TB treatments, avoiding consumers receiving unnecessary treatments, saving hospital resources, and reducing expenses in the long term. Participants also discussed the downstream challenges of a test with lower accuracy, such as the TST. These included psychological stress for consumers from false-positives and the expense of false-positives.

Reliability

The TST was perceived to be less reliable than the IGRA, and in some cases, the TBST. Most often, lower reliability perceptions were driven by the degree of human error that could take place for the TST comparative to the IGRA, which was perceived to produce a more objective result. Perceptions that TBST produces less false positives than the TST contributed to feeling as though the TBST were a more reliable test.

> *"The reading is not like a laboratory reading saying you've got this amount of whatever, I don't quite know how it works with the blood tests, but it's based on someone actually looking at it and visually grading it, which I would imagine is open to a little bit of personal interpretation and possibly not necessarily getting the accurate range across the like, yeah, if someone different looks at it they might grade it slightly differently."* (ID001, HIC, Consumer)

Time to Receive Result

Some consumers reported being more willing to use the TBST due to the perception that their results would be received more efficiently than the IGRA, which could take one to several weeks to receive a result.

Suitability in Specific Subgroups

Different TBI tests were viewed as more or less suited to particular subgroups. The TBST were believed to be more suited to people with prior BCG vaccination and children than the TST, although they are less suited to children with allergies. One country in which the Diaskintest is in use reported implementing alternative procedures to screen children with severe allergies who receive a positive TBST result, which involved a confirmatory IGRA test.

The IGRA test was discussed as being more suited to use in persons with HIV and BCG vaccination compared to the TST, and more suitable for use in children with allergies compared to the TBST. The TST was judged as more suitable for use in young children than the IGRA due to small veins creating complications with conducting blood draws. Moreover, the TST was discussed as being more suitable for use in groups with prior drug misuse issues whose veins have been damaged, compared to the IGRA:

> *"We are currently running a program in people that use drugs, so we are using TST to help us in order to diagnose latent TB so we can provide them or offering them the preventive therapy… for people that use heroin or injectable drugs because they don't have good veins to draw blood and skin tests is better or is easier to use in those scenarios."* (ID009, LMIC, Provider)

Ability to Differentiate Latent and Active TB

A perceived disadvantage of all 3 TBI tests was that none could differentiate between latent and active TB. Several participants discussed the technical challenges and thus low likelihood of developing a test to achieve that aim:

> *"No the challenges were that it was not really possible to differentiate by the results of the PPD test who is really infected and who really needs preventive treatment… currently all researchers in the field all over the world are looking for some biomarker that could help differentiate latent TB infection from the active TB infection right, yes, but it is very difficult to develop such a biomarker, such a test, because TB infection is very evasive, it can be latent and then it can be active."* (ID015, HIC, Provider)

Contextual Factors

Three subthemes covered contextual factors contributing to TBI test acceptability (Figure B5). Both consumers and healthcare providers discussed how provider perceptions of the test and the infrastructure for test follow-up impact on acceptability. However, only consumer participants discussed how the test environment affects test acceptability – no provider participants mentioned this in their discussion.

Figure B5: Acceptability – Contextual factors

Figure describing the contextual factors influencing test acceptability

Provider Perceptions of the Test

Provider communication with consumers around the TBST was anticipated to affect their acceptability (as discussed in an earlier work package: Acceptability > Provider-consumer communication), as this affected the consumer perceptions of the test. Moreover, endorsement and acceptance by healthcare workers and organizations (e.g., the WHO) was expected to increase the acceptability of the TBST. Providers also discussed how useful they thought the TBST would be for their clinical practice compared to the existing tests and in the context of implementation and acceptability challenges. Several providers felt that the TBST would be useful for providing an alternative option for TB screening in situations where the TST or IGRA were not appropriate. Other providers felt that the TBST would only be useful in situations where it is replacing the TST, because of the resources required. Some providers felt very negatively about TB skin tests due to their limitations in accuracy and reliability, and believed that they would only be helpful in very specific situations (e.g., evaluating the effectiveness of new TB treatments in resource-limited settings).

Infrastructure for Test Follow-Up

The infrastructure available to support test follow-up and the care pathway in place (i.e., the processes in place in the health system to support ongoing communication, follow-up and engagement with the consumer) affected acceptability of the TBI tests and consumer perception of whether the test was worthwhile to take. Participants reported issues such as receiving a positive

TST result in a large-scale screening program but not receiving any clinical follow-up, or government policies such as deporting individuals with a positive TBI test result.

> *"If you are being tested and shows positive, then you have to have the TBI treated before you go and those who travel out of the country and they are subjected to the skin test and it turns positive, then we get them deported, so those are some of the challenges… treatment for TBI is not readily available."* (ID010, LMIC, Provider)

Test Environment

The test environment was not raised as an influencing factor for test acceptability by any health providers. Some consumers felt that hospital environments were anxiety-provoking, as discussed in an earlier section (see Acceptability > Consumer experience). This issue could be applicable to all 3 TBI tests, although a hospital appointment is most often involved when receiving the IGRA compared to the skin tests.

3.2.3.3 Feasibility

People Factors

Only one subtheme described a people factor that impacted feasibility, which was consumer experience. This subtheme was discussed by both healthcare providers and consumers (Figure B6). Whilst provider communication and infrastructure for test follow-up are also potential factors that may be considered as relevant for feasibility considerations, participants only discussed provider-consumer communication as part of acceptability to influence beliefs and uptake of the test. Similarly, whether there was infrastructure in place for follow-up for a positive test result was discussed as part of acceptability, as it influenced consumer perceptions of whether the test was worthwhile getting, rather than in the context of feasibility.

Figure B6: Feasibility – People factors

Figure describing the people factors influencing test feasibility

Consumer Experience

Aspects of the consumer experience that impacted the feasibility of TBI tests included a return visit to read the test results, transport requirements, the timeliness of the results, and the healthcare setting.

A common feasibility concern around the TST and TBST was the need for consumers to return to have their test results read 48-72 hours after test administration as part of a second clinic appointment. Often consumers were lost to follow-up, and issues around access to transportation exacerbated the issue, as some participants reported:

> *"I mean we know a very large percentage of TB skin tests are never read."* (ID014, HIC, Provider)

> *"So we have another issue like when we apply TST we do make an appointment and then we ask them for another appointment at 72 hours in order to know the result but now not always everybody shows up to the second appointment… can be related to transportation or money resources, so maybe they don't have enough resources to move around the city or to go to another meeting… it tends to happen, and when it is we know we can't reapply the TST immediately which we will wait at least well ideally six months but at least three to six months."* (ID009, LMIC, Provider)

A clear advantage for the IGRA in terms of feasibility was the requirement for only one clinic visit to administer the test. This was found to have benefits for healthcare providers in terms of then only needing to allocate resources to follow up positive TBI individuals:

> *"Well honestly I prefer the IGRA. I think it is a better test in so many ways because one single visit."* (ID009, LMIC, Provider)

> *"The good thing about the T-SPOT or the IGRA is that they, once they've had the blood test done that's it and actually if they're not sort of then we can put effort knowing that it's positive or negative. If it's positive then obviously we will be following up and putting more effort into trying to find somebody whereas often with a Mantoux, if they don't come back within the timeframe then it has to be repeated."* (ID012, LMIC, Provider)

Transportation was also a part of the consumer experience for receiving the IGRA, with consumers needing to visit a clinic or hospital to receive the test. This may be more of a feasibility concern in middle/low-income countries compared to high-income countries. This experience is described by a consumer below:

> *"I mean I guess if you're looking slightly wider you have to, like I had to drive to the hospital, I had to park, I had to pay for parking, I had to go down into the appointment, into the occupational (indistinct) work, like screening area."* (ID001, HIC, Consumer)

Another consumer experience factor affecting feasibility was the timeliness of receiving the test result. This was perceived to be better with skin tests than with the IGRA, according to one consumer who preferred receiving a more immediate result a couple of days after administration:

> *"Well it's low cost, the results coming back quicker, you might have to go to the visit but you're not waiting months, waiting weeks or the three weeks, that's quite lengthy to get your results. Waiting two, three days which is good news."* (ID006, HIC, Consumer)

Lastly, the often-busy healthcare environments in which IGRA are typically administered may impair feasibility. Consumers reported needing to spend a lot of time waiting in a crowded or under-staffed hospital clinic to receive the IGRA, which may not be feasible for some consumers. In some cases, consumers speculated that the TBST could be a more efficient process than the IGRA:

> *"If you're going to the clinic and there's no one that can take your blood that day because there's no one there so you go to the other part of the hospital and then you take a ticket and you wait an hour or whatever and I think that could have been done, that sort of thing in*

a couple of minutes. So I think it's not just the cost, I think it's probably more time, my guess would be it's more time efficient as well." (ID002, HIC, Consumer)

Test Factors

Four subthemes described test factors that impacted TBI test feasibility (Figure B7). These were cost, time to receive the result, portability of the test, and ease of administration. All four subthemes were discussed by both consumers and healthcare providers.

Figure B7: Feasibility –Test factors

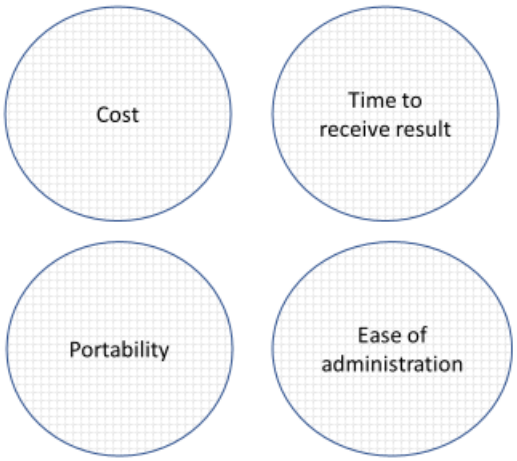


Figure describing the test factors influencing test feasibility

Cost

The TBST and TST were discussed as being of lower cost than the IGRA, thus improving their feasibility in this respect. Some healthcare providers talked about how they would be able to save costs in their current practice with TBST, while others acknowledged that the low cost may increase the willingness of governments to fund the TBST. A couple of providers discussed how the low cost affected their usage intentions towards the TBST. Quotes from providers discussing these aspects are presented below:

"I think in my current practice, since we are now dealing with TB patients as well, it will be an added advantage. I'm saving in costs because we are looking forward to do testing of TB like in children whereby I know it's very difficult for them to submit a sputum and we intend to at least maybe use this tool…" (ID004, LMIC, Provider)

"It depends also, we have to be very clear here that the country's, the service provider, which is the Government in most cases, might want to go for a cheaper option. So here, also the costs come into play, that which one will be cheaper to roll out to a bigger population and therefore it depends on the cost, cost become, so the blood test might be easier, might be a perfect one, but the Government might say oh the cost is a little bit higher so I go for the cheaper commodity." (ID008, LMIC, Consumer)

"The novel skin tests have a very good promising perspective, rollout very easy because it have a low cost and higher specificity in comparison with the IGRA… Generally I guess people

> *believe that the IGRA could have been widely used if we do not think about the cost… We plan to use the novel skin test in future just because the cost, in terms of cost, and also they have a good specificity."* (ID020, LMIC, Provider)

Time to Receive the Result

Although the quick time to receive the result for skin tests relative to the IGRA improved acceptability in some cases, in other cases, the 72 hour wait time for skin tests was described as a disadvantage in terms of feasibility. This is especially the case when consumers need to travel a long distance to reach a clinic, or a remote clinic is visiting rural communities:

> *"I mean, it's considerably better than not doing it without a doubt but I would say that might have limitations in, I've been in teams where I've done clinics and you're not there for 72 hours, so in remote healthcare or people travel considerable distances to receive treatment and testing, they're probably not gonna hang around 72 hours to have the results read."* (ID001, HIC, Consumer)

Portability

Portability was discussed as an advantage of the TBST and TST over the IGRA. The portability of skin tests made them more feasible for use in remote communities or mobile health clinics comparative to the IGRA:

> *"Yeah, I think this is better because if you choose one community you know that needs to be work on it, I think it's going to be better to a small community that we need to diagnose or if you have a case then pulmonary TB around so we can go and do novel test, yeah, that will be the small community."* (ID018, LMIC, Provider)

Moreover, the portability of skin tests was discussed as an advantage for improving test accessibility to people who are unable to or do not want to visit a healthcare setting:

> *"If you were going into say a homeless shelter or somewhere, to take the blood is quite a drawn out thing, pardon the pun, and then it's go to go to the lab and it's got to be processed in a certain time. I mean obviously these tests you have to come back and look at people but you could go to a setting and do that, going to the people rather than people coming to the setting and sometimes I think that's gonna be invaluable… because that's a barrier to a lot of people, it's a barrier and they don't want to go to a healthcare setting."* (ID002, HIC, Consumer)

Ease of Administration

In terms of the ease of administration, the TST was discussed as being relatively easy to use with appropriately trained staff available. Some providers discussed the technique of the TST as being tricky at first, but ease of use came with practice and appropriate training:

> *"I mean, they're fairly easy to use. I mean they're a bit fiddly but they're not, once you get used to it it's just getting the technique… But it's yeah, it's not difficult, once you get the hang of it."* (ID012, HIC, Provider)

> *"We found that we have to retrain our staff at least every six months of the year… with the result sometimes it's incorrect so every, for example in the community they accepting every time they go to a village, I have to spend the first two or three times intensively to check every (indistinct) to make sure that the member have inserted correctly, this is more*

> *challenge… I think from my personal point of view, it's pretty easy to use cos I know how to do it and I do it myself and I train other people… the technique is easy, in our team I mean we use the TST for 10,000 in total and there's only one incident occupation who the needle stick."* (ID013, HIC, Provider)

The TBST were judged as being easy to use due to requiring a similar administration technique to the TST, although a trained staff requirement would also apply:

> *"The novel skin test, the Diaskintest, is performed exactly with the same technique, the same logistic as the PPD test and then it will be assessed after 72 hours of placement of test. So this, well identity of implementation technique makes it very user friendly because there is no need of retraining the staff and all the staff have already been trained."* (ID016, LMIC, Provider)

> *"So you know actually in China so in the past actually especially two years before, we normally use the PPD, always use the traditional one. So two years ago, actually the EC, that's the novel skin test, that gets the license, so we gradually introduced the EC in our field… you know actually for this novel skin test it's very similar for the process and procedure, it's very similar."* (ID020, LMIC, Provider)

> *"I think in terms of the actual administration of the test and the reading of the test, I think, because I think it is very similar to the Mantoux so I don't think they're a particular, well certainly from my point of view, I can't speak for my staff but from my point of view I don't feel any sort of particular concerns about that… I don't think it would be too difficult because we already so, I mean we don't do masses of Mantoux's but we do do them. So we already have clinics set up, so our clinics sort of how they run."* (ID012, HIC, Provider)

The IGRA was viewed as less easy to administer in comparison to the skin tests due to the complexity involved, although this was not heavily discussed:

> *"Ideally if we had an alternative which was a, you know, fingerprick blood test or something that you could do that was much more straightforward that didn't require the same complexity that the IGRA do…" (ID019, HIC, Provider)*

Contextual Factors

Three subthemes covered contextual factors that impacted test feasibility (Figure B8). These were all discussed by both consumers and healthcare providers, and included resource requirements, the COVID-19 pandemic, and the health system setting.

Figure B8: Feasibility – Contextual factors

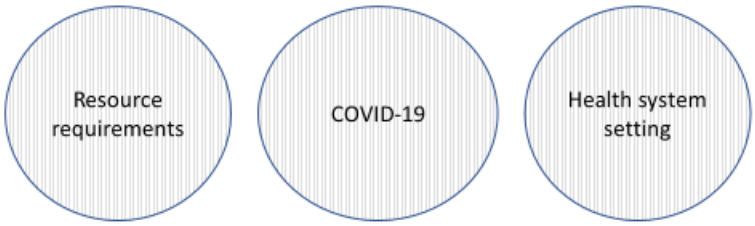


Figure describing the contextual factors influencing test feasibility

Resource Requirements

The TBST, TST, and IGRA all require access to specific test equipment and trained staff – as such, the availability of a trained workforce was considered key for test feasibility. However, the IGRA also requires access to a laboratory for test analysis and a hospital or clinic for the blood draw. As a result, there is also a mandatory transport requirement for the IGRA on the part of the consumer to the hospital or clinic. Patient transport may be a requirement of the TBST and TST in some settings, however not necessarily due to their portability. Overall, the IGRA requires more resources to administer than TBST and the TST, which may reduce its feasibility in more resource-limited settings.

> *"I feel like the good thing about it is you don't need much lab infrastructure, I feel like the need to sort of draw blood is a barrier for IGRA… I think that yeah, the lab aspect, yeah the fact that [the TBST] can be administered in pretty much any kind of setting is good, even though, I don't know about the stability of like what's the names, sometimes they go out of date."* (ID017, HIC, Provider)

COVID-19

Reduced clinic appointments and a shortage of blood test bottles and other test materials reduced the feasibility of all three TBI tests during the COVID-19 pandemic.

> *"I think right now if you were offered a blood test and you were given an appointment, like I think you treat that as a privilege because not all healthcare appointments are easily accessible at all at the moment."* (ID001, HIC, Consumer)

> *"I think they're mainly using the blood test although I don't know what's happened since COVID cos I think it's all sort of grinding to a halt."* (ID002, HIC, Consumer)

Health System Setting

Considerations relating to health system infrastructure was a key aspect of feasibility. The TBST and TST are lower-cost test options that were viewed to be more appropriate for low-resource settings. Moreover, their low cost and portability were discussed as making them more scalable for use in large scale screening programs than the IGRA. However, an advantage of the IGRA was the ability to screen for TBI alongside other tests when used as part of a routine blood draw. Participants also discussed how aspects of the healthcare system would affect the ease of implementing TBST (e.g., government approvals and willingness to fund the tests).

> *"At the time I had the blood taken, I had loads, they look loads of different vials cos they were testing for immunity for all sorts of healthcare related conditions… it was just another*

one of the blood tubes that they happened to take that day, so it didn't bother me at all." (ID001, HIC, Consumer)

"Honestly, I love IGRA. I think they should be the standard of care but they are expensive, so trying to be honest and trying to be realistic in countries where money matters, I think this new test can give us a chance to improve the program." (ID009, LMIC, Provider)

3.2.3.4 Health equity

Factors affecting the health equity impacts of TBI tests were described in five subthemes. These included access and transport issues, infrastructure required, suitability of the test to a low-income country setting, scalability of the test for large scale screening programs, and suitability to subgroups. Whilst infrastructure was also a feasibility issue, some participants discussed infrastructure in relation to equity specifically, where tests requiring less infrastructure to deliver were considered more accessible in resource-limited settings. All five subthemes were discussed by both consumers and healthcare providers (Figure B9).

Figure B9: Equity factors

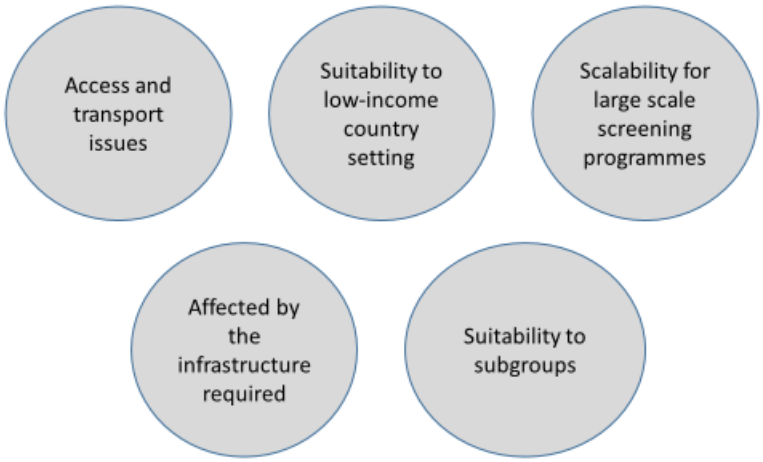


Figure describing the factors influencing test health equity

Access and Transport Issues

The IGRA typically requires a visit to a healthcare facility for test administration, whereas the TBST and the TST are more portable with less resource requirements, thus a visit to a healthcare facility might not always be required. This is a distinct advantage of the skin tests in low-resource settings, as the tests can be made accessible to consumers living a great distance from a healthcare facility, to help them avoid any of the challenges associated with attending a hospital appointment such as accessing transportation, a long journey to the clinic, and an overcrowded clinic environment with a long wait time. Although, in some instances, consumers are still required to travel to receive a TST.

"I think [the TBST] will have an impact because if it's more, if its accuracy resembles IGRA more, people who are at that last mile and unable to afford or access IGRA are gonna be able to have access to a test that is more accurate for them and hopefully more accessible because of the supply issues to TST. And so that will give them yeah will level the playing field

> *a little bit, someone in that area compared to someone who has access to IGRA in a higher resource part of the country."* (ID017, HIC, Provider)

In light of the above-mentioned transportation issues, another factor affecting TBI test accessibility is the requirement for consumers to make a return visit to the clinic 48-72 hours after administration of TBST and the TST. If a long journey or transportation challenges are experienced, this may serve as a major barrier for uptake of TB skin tests. Conversely, the IGRA only requires one clinic visit.

All 3 TBI tests require specialized staff training, which can create challenges in low-resource settings in which there may be a shortage of trained staff. This may increase the wait time for consumers and serve as an additional barrier to seeking TBI testing.

Affected by the Infrastructure Required

The IGRA were viewed as less accessible in low-resource settings due to the infrastructure required to support their use (as outlined in the previous subsection, e.g., a laboratory). On the whole, TBST and the TST were viewed as more accessible in resource-limited settings due to requiring less infrastructure. However, in some cases, such as in low- and lower-middle income countries, even TST were not accessible due to a lack of necessary equipment:

> *"I know that the skin tests in certain facilities, they're being used but I say most of the clinics in urban areas do not have this equipment, like the GeneXpert so they don't have the chest x-rays."* (ID003, LMIC, Provider)

Suitability to a Low-Income Country Setting

The cost and resources required to support the use of TBI tests were discussed as affecting their suitability for use in low-income country settings. As previously discussed, the IGRA was viewed as less suited to low-income country settings due to the higher cost and greater degree of resources required for implementation compared to the TST and the TBST.

> *"I guess the issue of making it more affordable, its bringing the potential of a different technology in a more affordable way to people who have a high burden of the problem."* (ID014, HIC, Provider)

> *"We've got access to phlebotomy and can get the IGRA done and it's quite easy, it sort of feels to me like well what would be the benefit changing. But I see that actually in areas where actually maybe there isn't phlebotomy, you haven't got that sort of access, then having something quite simple to do and actually you could train up, you can train up sort of non-healthcare workers, well you can for phlebotomy as well but in terms of the logistics and stuff I suppose it might be beneficial."* (ID012, HIC, Provider)

Scalability for Large Scale Screening Programs

A distinct advantage of TBST and the TST was their scalability for use in large scale screening programs in hard-to-reach or vulnerable communities. This advantage was driven by their low cost and portability to be taken to locations where there may be limited resources and healthcare facilities to support test administration, which the IGRA would require.

> *"It is very cheap like the PPD test. IGRA tests are very expensive… the Diaskintest is very cheap and that's why we can screen total populations which is huge. And we screen all population every year."* (ID015, LMIC, Provider)

> *"There are local healthcare workers… who roam in these slums and rural areas to give the immunization to children. So how better we capture on these existing health systems is the more important point in getting to the equity of the tests across rural and urban areas."* (ID005, LMIC, Provider)

Suitability to Subgroups

As discussed previously under the Acceptability section, the suitability of the TBI tests varied across different subgroups. Firstly, concerns were raised about the suitability of all 3 tests for use with darker skin tones due to issues with finding veins and viewing the welt size. Moreover, all 3 tests created discomfort for individuals with needle phobia, as they either involved a blood draw or an intradermal injection. However, some participants felt that the intradermal injection may be slightly less anxiety-provoking than the blood draw. TBST were viewed to be suitable for use in children and in persons with a prior BCG vaccination. However, they are less suited to children with allergies, for which the IGRA is best placed. The IGRA was also reportedly better for use in individuals with HIV and the BCG vaccination compared to the TST. The TST was found to have advantages for use in individuals with substance misuse problems that have created vein damage, and very young children for whom there may be difficulty finding veins. However, reservations remain about accuracy of skin-based tests in people with darker skin pigmentation and skin tones; if the TBST was found to be accurate in these populations, it may help increase motivation for use given the difficulties participants described in findings veins in darker skin tones. Overall, the health equity impacts of TBI tests may vary based on the consumer groups in which they are used.

3.2.4 Summary of Work Package B findings from User Interviews

End user values and preferences

Healthcare providers and consumers were found to value TBI test accuracy, convenience, a positive patient experience, low cost and resource requirements, and these factors affected their preferences for TBI tests. The novel antigen-based skin tests (TBSTs) were discussed as having comparable accuracy to the IGRA, however with better affordability and lower resource requirements. In comparison to the TSTs, the TBSTs were viewed to have better accuracy and comparable affordability, resource requirements, patient experience, and convenience. While the portability of TBSTs may improve their convenience (e.g. when used in a community screening programme), like the TST, they require a return visit from the patient to read the test result, and they may result in a similar degree of discomfort and stigma (from the welt in particular communities). Taking these factors into account, preferences for the TBSTs varied based on country income levels and resources of the healthcare setting. The TBSTs were viewed as preferable to the TSTs in resource-limited settings where the TST is already in use. IGRAs remained the preferred test option in high-income countries.

Acceptability

Overall, the results of the user interviews indicate that the TBSTs are likely to be a more acceptable alternative to the TST. TBST were perceived to have greater specificity and sensitivity than TST, thus they may help avoid the negative consequences of false positives (e.g., stress of misdiagnosis, unnecessary treatment, expense, and use of hospital resources). However, the TBST is expected to have many of the same limitations as skin tests in terms of consumer experience (e.g., a return visit required, discomfort, a welt on the arm, stigma, complicated for consumers with low health literacy). The IGRA appears the preferred test option in high-income countries due to its comparable

sensitivity and specificity to the TBST, with the bonus of greater ease of administration and the infrastructure already exists to support IGRA use in these settings. Concerns were raised about the use of skin tests in general; several participants reported that skin tests were a "dated, basic technology with too great a risk of human error". Participants also discussed several strategies for boosting the acceptability of TBST, including careful communication during its implementation with endorsement by healthcare providers and organizations (e.g., the WHO).

Feasibility

Another key finding from the user interviews was that the TBST would be feasible to use in settings where the TST is already in use. This is because these settings have the trained staff, equipment, facilities, and care pathways already in place to support use of the test. Like the TST, TBST are a low-cost, portable test that may be more suited to use in low-resource settings than the IGRA, which is expensive and requires a great degree of resources to implement (e.g., a laboratory). In high-resource settings in which all of the necessary infrastructure is present, the IGRA are a more feasible test option than both skin tests because they do not require a return visit to read the result, a step where consumers are regularly lost to follow-up.

Health equity impacts

In terms of the health equity impacts, the evidence indicates that it is unlikely that TBST would create any new equity issues. Rather, the TBST are likely to improve health equity through the provision of a more accurate, low-cost test for resource-limited settings where the TST is already in use. The ability to have access to a more accurate test with the TBST can improve equity in a low-income setting where traditionally they have had access only to TST or are reliant on symptom screening. Moreover, their portability and low cost make skin tests suited to use in large scale screening programs in vulnerable, hard-to-reach communities. However, it is possible that TBST may not improve health equity in low-resource settings that do not already use the TST (e.g., low-income countries that rely on symptom screening). Barriers to accessing skin tests in these settings would need to be addressed first (e.g., public funding of tests, staff training, equipment), as getting access to the TST itself is a barrier currently to testing uptake and treatment in these low-income settings. Implementing the TBST is therefore unlikely to alleviate these test access issues alone.

Conclusion

In conclusion, TBST were deemed to be more acceptable and equally feasible to the TST, with greater benefits for health equity driven by higher sensitivity and specificity. TBST may be more preferable for use in more resource-limited settings compared to the TST. IGRA appears the preferred test option in high-income countries with the infrastructure to support their use; it is unlikely that widespread use of the IGRA would change in these settings. End-users were found to value attributes such as test accuracy, convenience, a positive patient experience, cost, and resource requirements. These findings were used to inform the design of the Discrete Choice Experiment (Work Package C).

3.3 Work Package C - Discrete choice experiment (DCE)

This activity intends to answer following objective:

Objective D: To explore end-user values and preferences for TB testing by investigating the trade-offs end users are willing to make between test attributes derived from systematic reviews and in-

depth user interviews, and determine the order of preference of different test attributes using a discrete choice experiment.

3.3.1 Aims

The aim of this work package was to conduct a discrete choice experiment (DCE) in order to explore end-user (consumer and provider) values and preferences for the novel TB specific TBST.

3.3.2 Methods

Study design:

A cross-sectional discrete choice experiment survey was conducted online in English. A DCE is a quantitative method used to elicit stated preferences from participants without directly asking them to state their preferred options. DCEs quantify the strength of user preferences and can identify preference heterogeneity across users.

Data collection opened on 19th October 2021, and closed on 10th December, 2021.

Population

Participants were individuals who were either consumers (people living with TBI) or a provider (a health provider involved in the TB care pathway). The inclusion criteria were: participants ≥18 years, with English at a level sufficient to provide informed consent, and to undertake study procedures including having access to the internet. Previous participants were not excluded, nor were they sought to be specifically included in the DCE element. No prior experience of skin testing or TB was required for inclusion, as the respondents were answering from a hypothetical case scenario perspective. Participants responding "yes" to the questions "Are you a provider of TB healthcare?" received the provider DCE, those responding "no" received the non-provider DCE.

3.3.2.1 Design of DCE tool

Attributes and levels

Attributes (rows) refer to the features of the test that may influence the testing uptake decision and the different types of features that may be available (Table C1). Levels refer to the values that attributes take in the DCE design (cells). Table C1 shows the attributes and levels defined for this DCE.

Criteria defined by Hensher et al. were followed to develop the attributes and levels[24] for the DCE. Hensher's criteria are:

(1) All levels and their combinations should be reasonable.

(2) All levels and their combinations should be understandable by respondents and feasibly a choice they could make in reality.

(3) Heterogeneity of the levels should be fully considered in the design to ensure the respondents can make some trade-offs between them.

We used the findings from the systematic review (section A) and user interviews (section B) to determine the attributes and levels using a standard iterative process[25]. The review team discussed the findings with input from the WHO GDG, who identified key attributes that were important for developing the DCE questionnaire. The review team discussed the findings, and reviewed and refined the attributes and levels to meet our criteria. Consistent with the size of DCEs in the published literature[26], we had six attributes with two to six levels for each. The attributes included factors that map to acceptability (waiting time for results, test efficiency, test accuracy (false

positives/ false negatives)); feasibility (cost, ease of test administration); and health equity (ease of integration into existing health systems, for example, with existing medical record or reminder systems).

Choice Tasks

A choice task is where a participant is given a choice between different scenarios based on different attribute and level combinations. We constructed a DCE with choice tasks that included two options of hypothetical TBI test alternatives and an opt-out alternative of choosing neither, which was included to estimate preferences more relevant to real-world choices.

Separate DCE choice tasks were created for respondents who were providers of TB services and for respondents who were consumers. The attributes included in each DCE differed slightly, though some were common across both. The full list of attributes included in both DCEs is shown in Table C1.

3.3.2.2 Experiment and questionnaire design

A full-factorial design of this size using all the attributes and levels resulted in 3 × 3× 3 × 3 × 6 × 2 = 972 possible profiles, which provide 471 906 pairwise choice sets for selection. To make tasks manageable to participants, we used NGENE software to construct 10 pairwise choice sets using a D-optimality algorithm. For piloting, we set priors as zero[23], then used the same software and approach with priors from the pilot to generate a 10-task d-optimal final design. All choice sets were checked for plausibility and dominance. The pilot and final surveys were hosted on an online survey platform (onlinesurveys.ac.uk).

Table C1: List of attributes and levels included in both provider and consumer DCEs

Provider DCE		Consumer DCE	
Attribute	**Levels**	**Description**	**Levels**
Cost to healthcare practice	USD 100	**Time required for test results to come back**	1 day
	USD 50		3 days
	USD 25		7 days
If a follow-up appointment is needed to get test results	In-person follow-up appointment needed	**If a follow-up appointment is needed to get test results**	In-person follow-up appointment needed to get results
	No follow-up appointment required		No follow-up appointment required
Chance of the test wrongly saying you have TB infection when you do not[27,28]	1% chance you will get treatment you don't need	**Chance of the test wrongly saying you have TB infection when you do not[27,28]**	1% chance you will get treatment you don't need
	5% chance you will get treatment you don't need		5% chance you will get treatment you don't need
	10% chance you will get treatment you don't need		10% chance you will get treatment you don't need
Chance of the test wrongly telling you that you do not have TB infection when you do[27,28]	15% chance you will not get treatment when you need it	**Chance of the test wrongly telling you that you do not have TB infection when you do[27,28]**	15% chance you will not get treatment when you need it
	20% chance you will not get treatment when you need it		20% chance you will not get treatment when you need it
	25% chance you will not get treatment when you need it		25% chance you will not get treatment when you need it
Ease of administration and interpretation	Does not require specialist staff or equipment to administer or interpret	**Where you can get the test**	Hospital only
	Requires specialist staff or equipment to administer or interpret		Primary care (e.g. family doctor, GP, pharmacy)
			Community (e.g. workplace or your house)
		How the test is performed	Skin prick and blood drawn
			Skin prick and injection under the skin

NB. Shaded cells denote attributes common across the provider and consumer DCEs.

3.3.2.3 Recruitment

Participants for the DCE were recruited from the same research networks used to recruit the users for the interviews for Work Package B e.g. from a clinical, research and community setting:

- TB Treatment Action Group
- TB Civil Societies
- Members of the WHO Guideline Development Group for TBST
- Global Coalition of TB Activists (GCTA)
- Stop TB Partnership
- WHO TBI Task Force

Potential participants were provided with an online link, which led them to the study information sheet. At the bottom of the information sheet webpage, participants were informed that by pressing "Next" to start the survey, they were consenting to participate.

3.3.2.4 Data analysis

Choice data were analysed in a random utility framework using multinomial logit models, and utility weights estimated for the contribution of each level to respondent utility functions. Dummy coding was used for categorical attributes, and three continuous attributes in each DCE were decided upon at the design stage: the likelihood a test would produce a false positive or negative was modelled as continuous in both provider and non-provider DCEs, while cost was modelled as continuous in the provider DCE, and waiting time in the non-provider DCE. To explore the trade-offs the participants were willing to make between test attributes, we calculated the marginal rates of substitution between attributes, with the denominator of the probability of a false positive result which was the largest in piloting of the two accuracy attributes common across both DCEs. The willingness-to-accept results are interpreted as the amount of false-positive accuracy, measured in percentage points of likelihood that a result is a false positive, that participants are willing to forego or bear for the benefit offered by other attributes.

3.3.2.5 Sample size

A recruitment target of approximately 150-200 provider participants and 150-200 consumers was set for the DCE. This sample size target was confirmed after piloting with ten investigators and colleagues to ensure that the DCE gave sufficient information on important test attributes at a sample level[23]. Diagnostics from the design, specifically the S-error of designs, suggested that statistically significant parameters would be obtained for all attributes in both designs with 45 participants, assuming those in the final sample chose in the same way as the pilot.

3.3.2.6 Data management

After informed consent and self-enrolment on the online survey platform, participants were assigned a nondescript participant number. All data collected from the participants were labelled using only the participant number to protect participant confidentiality. Participant data will not be shared outside of the research team. Data were stored in secure password-protected servers at the University of Auckland and LSHTM, and on the onlinesurveys.ac.uk server.

3.3.3 Results

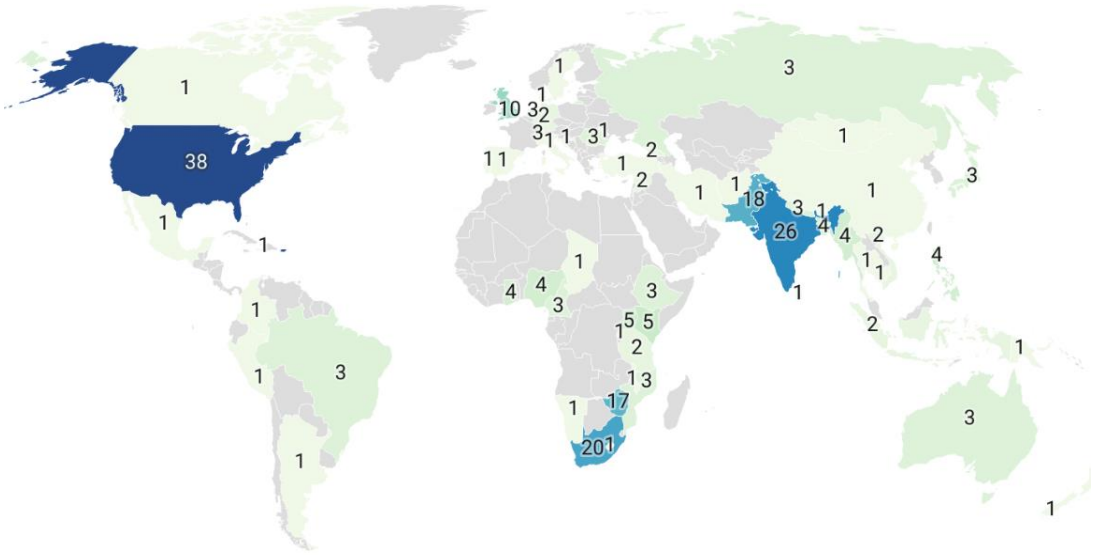
3.3.3.1 Respondent Characteristics

In total 236 people responded to the survey, 186 providers and 48 consumers. Thus, both sub-populations met the minimum possible sample size needed to draw statistical inference. Table 4 shows the characteristics of these respondents, and Figure C1 shows their geographical distribution. Overall, 59% of respondents were female, the majority (56%) were aged 36-55 years, and the main countries in which respondents were based were the India (26%), United States (16%), South Africa (9%), Pakistan (8%) and Zimbabwe (7%).

Table 4: Characteristics of DCE respondents

	Provider (n=186)	Consumer (n=48)
Gender		
Female	112 (60%)	26 (54%)
Male	73 (39%)	22 (46%)
Other	1 (1%)	
Age Category		
18-35	32 (17%)	13 (27%)
36-55	106 (57%)	24 (50%)
56-75	44 (24%)	11 (23%)
75+	4 (2%)	0 (0%)

Figure C1: Location of DCE respondents



3.3.3.2 Choice analysis

Table 4 shows choice model results, analysing DCE data for providers and consumers using multinomial logistic regression models. Where coefficients are significantly greater than zero, an attribute can be interpreted as having a relatively positive impact on participant utility. The magnitude of coefficients can be directly compared within each model. The statistical significance of coefficients (i.e. what is important to people's choices) can be compared across models, but due to possible differences in the scale of choice data across groups, coefficients cannot be directly compared across models.

To aid interpretation, the cost attribute is set at the central level of USD$50, based on a hypothetical cost, whilst the likelihood of false-positive and false negative results is set at 10% for both. In both models, almost all coefficients were of the expected sign suggesting good internal validity of results.

In both providers and consumers, the strongest determinant of choice among test attributes was test accuracy in terms of false positivity and false negativity rates, and preferences for test sensitivity and specificity were not valued differentially among providers (p=0.87) or consumers (p=0.43).

Among consumers, the next most influential attribute was being able to obtain the test from a primary care setting, which was significantly preferred to obtaining a test from a hospital (β=0.427, p=0.02), whilst community testing was weakly preferred to a hospital testing (β=0.395, p=0.07). Consumers displayed no strong preferences for in-person follow up or whether a test required blood to be drawn, but had strong and significant preferences for any kind of test compared to no test, indicated by the large and significant negative coefficient for the opt-out alternative (β=2.55, p<0.01).

Among providers, after test accuracy the next most influential attribute was not requiring an in-person follow-up appointment (β=0.151, p=0.02), followed by not requiring specialist staff or equipment to administer or interpret the test (β=0.144, p=0.01). There was weak evidence that providers preferred more expensive tests, though test cost was the least important determinant of choice (β=0.133, p=0.05). Cost was not an attribute in the consumer DCE.

Table 5: Multinomial logit choice model results

	(1) Consumer		(2) Provider	
	Coefficient	**SE**	**Coefficient**	**SE**
No in person follow-up appointment needed	0.0829	(0.175)	0.151**	(0.0643)
Chance of the test wrongly saying there is TB infection when there is not (false positive)	-0.790***	(0.181)	-0.189**	(0.0819)
Chance of the test wrongly saying there is no TB infection when there is (false negative)	-0.967***	(0.184)	-0.169*	(0.0960)
Waiting time in days for test result	-0.0953*	(0.0523)	(not presented as an attribute)	
Place of the test (Reference: hospital)				
Primary care (e.g. family doctor, GP, pharmacy)	0.427**	(0.179)	(not presented as an attribute)	
Community (e.g. workplace or your house)	0.395*	(0.221)		
How the test is performed (Reference: skin prick and blood drawn)				
Skin prick and injection under the skin	-0.0599	(0.262)	(not presented as an attribute)	
Cost (USD$50)	(not presented as an attribute)		0.133*	(0.068)
Requires specialist staff or equipment to administer or interpret	(not presented as an attribute)		-0.144**	(0.0562)
Opt-out (no test)	-2.548***	(0.612)	-0.419*	(0.234)
Number of respondents	48		186	
Observations	1,467		5,583	
Bayes Information Criteria	1015.7		4096.6	

*** p<0.01, ** p<0.05, * p<0.1

3.3.3.3 Willingness-to-accept analysis

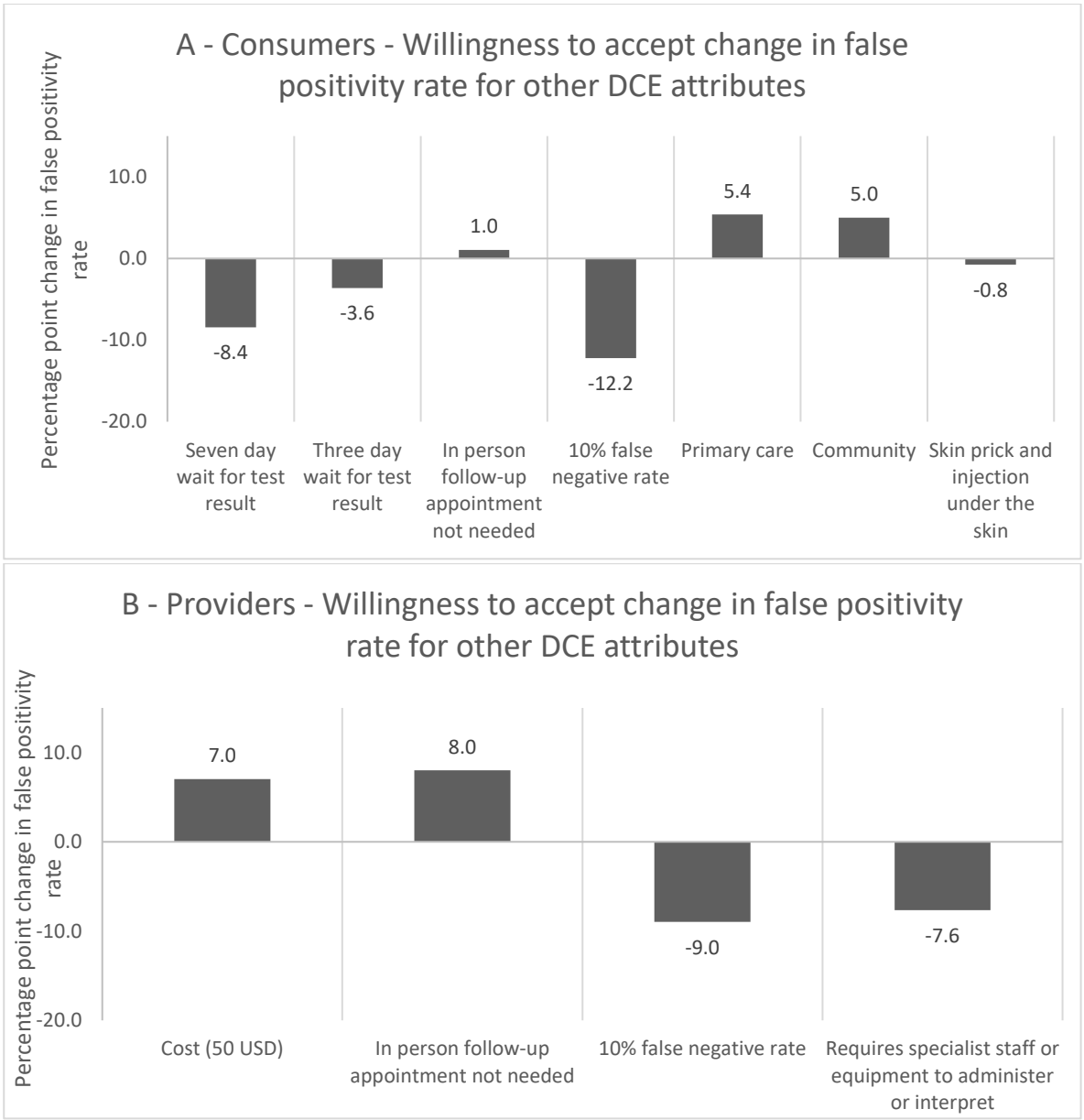
Figure C2 shows participant willingness to accept a change in false positivity rate for different test attributes; these coefficient ratios are comparable across provider and consumer samples. Willingness to accept a loss of accuracy suggests the attribute is favourable and is indicated by bars in the positive region (above zero), while bars in the negative region below zero indicate that a gain in test accuracy would be needed to compensate for an attribute which is less favourable.

Panel A shows that consumers would be willing to accept an 8.4 percentage point greater false positivity rate to avoid a seven-day wait for test results and a 3.6 percentage point greater rate to avoid a three-day wait, and a 5.4 and 5 percentage point greater false positivity rate for a test to be provided in primary care or community settings, respectively. The relatively weak consumer preferences for avoiding in-person follow-up and the type of test are also reflected here.

Panel B shows that providers would be willing to forego 8 percentage points of false positivity rate for a test which did not require an in-person follow up appointment, and 7.6 percentage points for a test which does not require specialist staff or equipment to administer or interpret.

There was some evidence of preference heterogeneity between providers and consumers as avoidance of in-person follow-up was valued eight times more by providers than by consumers. Whilst consumers valued a one percentage point increase in the likelihood of a false negative result slightly more negatively than the same increase in likelihood of a false positive; the opposite was true among providers.

Figure C2: Willingness to accept change in false positivity rate for DCE attributes.



Panel A shows results for consumers, and panel B results for providers. A positive value indicates that respondents are willing to accept a lower likelihood of a false-positive result for that attribute.

Panel A, example interpretation: Consumers would be willing to accept an 8.4 percentage point greater false positivity rate to avoid a seven-day wait for test results.

Panel B, example interpretation: Providers would be willing to forego 8 percentage points of false positivity rate for a test which did not require an in-person follow-up appointment

3.3.4 Summary of Work Package C findings from DCE

This DCE explores end-user values and preferences with regards to TB testing. The DCE was conducted with 236 potential test consumers and providers, identified strong and consistent preferences among both groups for tests that minimize false positive and false negative results, which were identified as the most important test characteristics by both groups. After test accuracy, consumers' strongest preferences were for the location of testing, with testing in primary care strongly preferred to hospital locations. For providers, the next most important attribute was the lack of requiring an in-person follow-up appointment, followed by not needing specialist staff or equipment to interpret or administer the test. Willingness-to-accept analyses indicated that both consumers and providers were willing to bear a substantial reduction in test sensitivity to reduce waiting times for results (8.4 percentage point reduction to avoid a 7-day wait), and to avoid in-person follow-up visits (8 percentage points) or the need for specialist staff (7.6 percentage points).

A strength of this study is the use of a DCE methodology to quantitatively assess respondent preferences for tests. The study achieved reasonable geographic diversity in respondents, and results suggest that tasks were understood well by respondents who were able to trade-off over a number of attributes simultaneously. The design process was also strong, building off the in-depth interview and systematic review evidence to design an initial DCE task for piloting.

This DCE study has several limitations. First, DCEs require choices between hypothetical alternatives which may not reflect real-world choices accurately or comprehensively. Although there is evidence that DCEs predict real-world choices reasonably well[29], there remains the potential for hypothetical bias in these data if respondents chose differently in the online survey, than they would given different alternatives in real-life. Nevertheless, that preferences align with *a priori* expectations is an indication of internal validity.

The study was only able to recruit around 25% of the total target sample size for consumers. Although this was slightly over the smallest sample size needed to get reliable data, as indicated by design diagnostics, this may explain why the consumer sample do not show statistically significant preferences for less important attributes, for example towards requirement for in-person follow-up or how tests are performed. The provider sample was larger, though at 186 was still 14 respondents below the target sample size.

Although there was reasonable geographic diversity in responses, there were insufficient responses to explore preference heterogeneity across different geographical settings. To accurately capture how preferences differ by geography, a country- or region-stratified sample is required.

4. References

1. World Health Organization. Global Tuberculosis Report. Geneva: World Health Organization; 2021. Accessed via: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2021

2. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016. doi:10.1371/journal.pmed.1002152

3. Abajobir AA, Abate KH, Abbafati C, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1260-1344. doi:http://dx.doi.org/10.1016/S0140-6736%2817%2932130-X

4. Comstock GW, Livesay VT, Woolpert SF. The prognosis of a positive tuberculin reaction in childhood and adolescence. Am J Epidemiol 1974; 99: 131–8.

5. World Health Organization. Global Strategy and Targets for Tuberculosis Prevention, Care and Control after 2015. Geneva: World Health Organization 2014.

6. Mitchell EM, Golub J, Portocarrero A. Acceptability of TB screening among at-risk and vulnerable groups. Geneva: World Health Organization 2011.

7. Deck F, Guld J. The WHO tuberculin test. Bull Int Union Tuberc. 1964;34(1):53-70.

8. Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: What is the absolute effect of BCG and non-tuberculous mycobacteria? Int J Tuberc Lung Dis. 2006.

9. Diel R, Goletti D, Ferrara G, Bothamley G, Cirillo D, Kampmann B, Lange C, Losi M, Markova R, Migliori GB, Nienhaus A. Interferon-γ release assays for the diagnosis of latent Mycobacterium tuberculosis infection: a systematic review and meta-analysis. European Respiratory Journal 2011;37(1), pp.88-99.

10. Rangaka MX, Diwakar L, Seldon R, et al. Clinical, Immunological, and Epidemiological Importance of Antituberculosis T Cell Responses in HIV-Infected Africans. Clin Infect Dis. 2007. doi:10.1086/518234

11. Pai M, Denkinger CM, Kik S V, et al. Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. Clin Microbiol Rev. 2014;27(1):3-20. doi:https://dx.doi.org/10.1128/CMR.00034-13

12. Hoff ST, Peter JG, Theron G, Pascoe M, Tingskov PN, Aggerbeck H, et al. Sensitivity of C-Tb: A novel RD-1-specific skin test for the diagnosis of tuberculosis infection. Eur Respir J. 2016. doi:10.1183/13993003.01464-2015

13. Krutikov M, Faust L, Nikolayevskyy V, Hamada Y, Gupta RK, … Rangaka MX. The diagnostic performance of novel skin-based in vivo tests for tuberculosis infection compared to PPD tuberculin skin tests and blood-based in vitro interferon-gamma release assays: A systematic review and meta-analysis. In press.

14. Engel N, Ochodo EA, Karanja PW, Schmidt BM, Janssen R, Steingart KR, Oliver S. User perspectives on NAATs to detect TB and resistance to anti-TB agents: results from a qualitative evidence synthesis (systematic review). Geneva: World Health Organization 2020.

15. Noyes J, Booth A, Cargo M, Flemming K, Harden A, Harris A, Harris J, et al. Chapter 21: Qualitative evidence. In J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane.

16. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan- a web and mobile app for systematic reviews. Systematic Reviews. 2016;5:210.

17. Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, Noyes J, Booth A, Garside R, Rashidian A. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS medicine. 2015 Oct 27;12(10):e1001895.

18. Aksenova V. Summary results of the recombinant tuberculosis allergen (Diaskintest) use in clinical practice. 2021.

19. Kerrigan D, Tudor C, Motlhaoleng K, Lebina L, Qomfu C, Variava E, Chon S, Martinson N, Golub JE. Relevance and acceptability of using the Quantiferon gold test (QGIT) to screen CD4 blood draws for latent TB infection among PLHIV in South Africa: formative qualitative research findings from the TEKO trial. BMC health services research. 2018 Dec;18(1):1-7.

20. Pascopella L, Franks J, Marks SM, Salcedo K, Schmitz K, Colson PW, Hirsch-Moverman Y, Flood J, Sayles J. Opportunities for tuberculosis diagnosis and prevention among persons living with HIV: a cross-sectional study of policies and practices at four large Ryan White Program-funded HIV clinics. PLoS One. 2014 Jul 7;9(7):e101313.

21. Spruijt I, Erkens C, Suurmond J, Huisman E, Koenders M, Kouw P, Toumanian S, Cobelens F, van den Hof S. Implementation of latent tuberculosis infection screening and treatment among newly arriving immigrants in the Netherlands: a mixed methods pilot evaluation. PLoS One. 2019 Jul 1;14(7):e0219252.

22. Forster AS, Rubin G, Emery JD, et al. Measuring patient experience of diagnostic care and acceptability of testing. Diagnosis. 2021.

23. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. The patient. 2015;8(5):373-384.

24. Hensher D, Rose J, Greene W. Applied Choice Analysis 2nd. Vol 1243. 2nd ed. New York: Stern School of Business; 2015.

25. Coast J, Al-Janabi H, Sutton EJ, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health Economics. 2012;21(6):730-741.

26. R Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics. 2019 Feb;37(2):201-226.

27. Krutikov M, Faust L, Nikolayevskyy V, Hamada Y, Gupta RK, Cirillo D, Mateelli A, Korobitsyn A, Denkinger CM, Rangaka MX. 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. Lancet Infect Dis. 2021 Oct 1:S1473-3099(21)00261-9.

28. Kahwati LC, Feltner C, Halpern M, Woodell CL, Boland E, Amick HR, Weber RP, Jonas DE. Primary Care Screening and Treatment for Latent Tuberculosis Infection in Adults: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA. 2016 Sep 6;316(9):970-83.

29. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur J Health Econ. 2018 Nov;19(8):1053-1066.

Supplementary appendix A

Search Terms- Review A

Table S1. Search strategy in Medline and Embase databases (and records retrieved for initial search)

Initial search conducted on 15 May 2019. The search was updated on 20 Oct 2020 using the same search terms but restricting to records published after 15 May 2019.

	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).mp.
3	skin test*.mp. or Skin Tests/
4	(c tb or diaskintest).mp.
5	1 and 2 and 3
6	4 or 5
7	**remove duplicates from 6**

Search Terms- Review B

Table S2. Search strategy in MEDLINE database and records retrieved for an initial search.

The initial search was conducted on 2 August 2021 using "Ovid MEDLINE® Epub Ahead of Print, In Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily, and Ovid MEDLINE® 1946-Present."

	Search Term	Results
1	exp TUBERCULOSIS/ or tuberculosis.mp. or exp MYCOBACTERIUM TUBERCULOSIS/ or tb.mp.	282603
2	TST.mp. or Tuberculin Skin Test.mp. or Tuberculin Test*.mp. or IGRA*.mp. or Interferon gamma release assay*.mp. or Interferon gamma release test*.mp. or QuantiFERION.mp. or QFT-GIT.mp. or QFT.mp. or T-SPOT-TB.mp. or T-SPOT.mp. or Interferon-gamma Release Assays/ or Interferon-gamma Release Assay/ or Interferon gamma Release Assays/ or Interferon gamma Release Tests/ or Interferon-gamma Release Test/ or IGRA/ or IGRA/ or QuantiFERON/ or QFT-GIT/ or QFT/ or T-SPOT-TB/ or T-SPOT/	22125
3	"Patient Acceptance of Health Care"/ or acceptability.mp. or Health Equity/ or equity.mp. or Health Services Accessibility/ or Patient Preference/ or preferences.mp. or Patient Satisfaction/ or Attitude to Health/ or barriers.mp. or challenges.mp. or patient experience*.mp. or "Attitude of Health Personnel"/ or providers experience*.mp.	959837
4	Interviews as Topic/ or interview*.mp. or Interview/ or survey*.mp. or Health Surveys/ or Health Care Surveys/ or "Surveys and Questionnaires"/ or Qualitative Research/ or Focus group discussion*.mp. or Focus Groups/ or "mixed methods".ti. or "mixed methods".ab. or "mixed-methods".ti. or "mixed-methods".ab.	1439956
5	1 and 2 and 3 and 4	50
6	Remove Duplicates from 5	50
7	Limit 6 to yr="2011 -Current"	35
8	Limit 7 to Humans	29

Appendix B

GRADE-CERQual- Summary of Qualitative Findings Table

Summary of review finding	**Studies contributing to the review finding**	**CERQual assessment (confidence in the findings)**	**Explanation of CERQual assessment**
User values and preferences for LTBI tests	Aksenova 2021; Kerrigan 2018; Pascopella 2014; Spruijt 2019	Moderate confidence	Due to moderate concerns about methodological limitations, relevance, and adequacy of the data, and minor concerns about coherence.
Acceptability of LTBI tests	Aksenova 2021; Kerrigan 2018; Pascopella 2014; Spruijt 2019	Moderate confidence	Due to moderate concerns about the adequacy of the data, methodological limitations and relevance, and minor concerns about coherence.
Feasibility of LTBI tests	Kerrigan 2018; Pascopella 2014; Spruijt 2019	Moderate confidence	Due to serious concerns about adequacy of the data, moderate concerns about relevance, and minor concerns about coherence and methodological limitations.
Health equity impacts of LTBI tests	Aksenova 2021; Kerrigan 2018; Pascopella 2014; Spruijt 2019	Low confidence	Due to serious concerns about the adequacy of the data, moderate concerns about the relevance and methodological quality, and minor concerns about coherence.

Appendix C

Coding framework for Work Package A: Systematic Reviews

Contextual factors
COVID-19 pandemic
Healthcare setting
Ability to screen alongside other tests
Aspects of the healthcare system and infrastructure
Provider-related
Resources
Equipment or sample storage requirement
Trained staff requirement
Test environment
Hospital environment
Usefulness of test affected by the care pathway that follows
Equity factors
Access and transport issues
Amount of staff in low resource setting
Health environment crowdedness
Long journey to clinic
Return visit to read result
Specialised staff training
Transportation to clinic
Wait time
Affected by infrastructure required
Scalability for large scale screening programs
Suitability to low-income country setting
Suitability to subgroups
Children
Comorbidities
Needle phobic
Consumer factors
Demographics
Age
Familiarity perceptions
Patient experience
A return visit required to the healthcare setting to read result
Comfort
Discomfort
Health environment crowdedness
Long journey to clinic
Time spent at hospital or clinic
Transport

Perceptions of provider
Skin test perceptions
Test knowledge
Usefulness perceptions
Test factors
Ability to differentiate between latent and active TB
Accuracy
Specificity
Human vs. machine
NSTs, IGRAs are more accurate than TSTs (antigen-specific)
Certainty of results
Cost
Ease of administration
Portability
Reliability
Suitability in subgroups
Children
Comorbidities
Needle phobic
Time to receive result

Appendix D

Email template used for recruitment of participants for Work Package B

Dear _____

We are currently conducting some research on behalf of the World Health Organisation (WHO) into user perspectives of skin-based tests that are used as part of tuberculosis (TB) diagnosis and testing. We are reaching out to you as you have either previously consented to be contacted for research purposes or are currently involved in research for TB. We are seeking your views as someone who is involved in the TB diagnosis pathway, either as a health provider or member of the public.

We are looking for people who are interested in sharing their views and experiences about TB skin testing via a remote video call over 30-45 minutes. Your views will be used to inform future WHO guidelines on TB testing.

If you are interested in taking part and helping inform future TB care, please reply to this email and we will send you some more information.

Many thanks in advance for your help.

Appendix E

Coding framework for Work Package B: WHO TB Interviews

Contextual factors
COVID-19 pandemic
Reduced clinic access
Shortage of blood test bottles or other materials
Healthcare setting
Ability to screen alongside other tests-
Aspects of healthcare system and infrastructure
Public funding of tests
Income of country
Low-cost tests more appropriate for low-resource settings
Scalability for large scale screening programs
Provider-related factors
Communication with patient around the test
Endorsement by others
Provider endorsement
Usefulness perceptions
Resources
Equipment or sample storage requirement
Hospital or clinic environment
Laboratory requirement
Trained staff requirement
Transport requirement
Test environment
Hospital environment
Usefulness of test affected by the care pathway that follows
Equity factors
Access and transport issues
Amount of staff in low resource setting
Greater accessibility from lower cost
Health environment crowdedness
Long journey to clinic
Patient doesn't need to visit healthcare setting
Return visit to read result
Specialised staff training
Transportation to clinic
Wait time
Affected by infrastructure required
Scalability for large scale screening programmes
Portability
To reach vulnerable groups (e.g., homeless)
Suitability to low-income country setting
Cost
Resourcing
Suitability to subgroups
BCG vaccine
Children
Comorbidities

Health workers
Needle phobic
Skin tone
Consumer factors
Demographics
Age
Needle phobic individuals
Skin tones
Familiarity perceptions
Similar experience to other accepted tests
Patient experience
A return visit required to healthcare setting to read the test result
Comfort
Discomfort
Not uncomfortable
Perception that NSTs are more comfortable than IGRA
Sore arm
Health environment crowdedness
Health Literacy
Level of invasiveness
Long journey to clinic
Patient doesn't need to visit healthcare setting
Physical consequences
No physical consequences
Painful
Welt on arm
Psychological consequences
Anxiety
Environment-related
Anxiety from hospital environment
Test-related
NST- Anxiety from seeing welt on arm
No psychological consequences
Stigma
Stress from false positives
Time spent at hospital or clinic
Timeliness of result for patient
Transport
Travel to a healthcare setting required
Unease from 'live' reading in front of patient
Perceptions of provider
Endorsement by others
Provider communication
Provider endorsement
Skin test perceptions
A basic test
A dated technology
An improved version of the TST
Test knowledge
Amount of research evidence

Limited public knowledge
Usefulness perceptions
Test factors
Ability to differentiate between latent and active TB
Accuracy
Downstream benefits of high accuracy
Avoid patients receiving unnecessary treatment
Improve evaluation of TB treatments
Improve rates of diagnosis
Less expensive in the long run
Prevent psychological stress from false positives
Save hospital resources
Downstream challenges of low accuracy
Expense of false positives
Psychological stress from false positives
Sensitivity
Specificity
Concerns about boosting
Human vs. machine
NSTs, IGRAs are more accurate than TSTs (antigen-specific)
Certainty of results
Cost
Ease of administration
Portability
Reliability
Human vs. machine reading of results
Suitability in subgroups
BCG vaccine
Children
Comorbidities
Health workers
Needle phobic
Skin tone
Time to receive result

World Health Organization

9789240056640