HEALTH TECHNOLOGY ASSESSMENT

VOLUME 23 ISSUE 23 MAY 2019
ISSN 1366-5278

Interferon gamma release assays for Diagnostic Evaluation of Active tuberculosis (IDEA): test accuracy study and economic evaluation

Yemisi Takwoingi, Hilary Whitworth, Melanie Rees-Roberts, Amarjit Badhan, Christopher Partlett, Nathan Green, Aime Boakye, Heather Lambie, Luigi Marongiu, Mark Jit, Peter White, Jonathan J Deeks, Onn Min Kon and Ajit Lalvani on behalf of the Interferon Gamma Release Assays for Diagnostic Evaluation of Active Tuberculosis (IDEA) study group

NHS
National Institute for Health Research

DOI 10.3310/hta23230

Interferon gamma release assays for Diagnostic Evaluation of Active tuberculosis (IDEA): test accuracy study and economic evaluation

Yemisi Takwoingi,[1] Hilary Whitworth,[2,3] Melanie Rees-Roberts,[2,4] Amarjit Badhan,[2,4] Christopher Partlett,[1] Nathan Green,[5,6,7,8] Aime Boakye,[2,4] Heather Lambie,[2] Luigi Marongiu,[2] Mark Jit,[8,9] Peter White,[5,6,7,8] Jonathan J Deeks,[1] Onn Min Kon[2,4,10] and Ajit Lalvani[2,4]* on behalf of the Interferon Gamma Release Assays for Diagnostic Evaluation of Active Tuberculosis (IDEA) study group†

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[2]Tuberculosis Research Centre, National Heart and Lung Institute, Imperial College London, London, UK
[3]Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, UK
[4]National Institute for Health Research (NIHR) Health Protection Research Unit in Respiratory Infections, Imperial College London, London, UK
[5]NIHR Health Protection Research Unit in Modelling Methodology, Imperial College London, London, UK
[6]Medical Research Council (MRC) Centre for Outbreak Analysis and Modelling, Imperial College London, London, UK
[7]Department of Infectious Disease Epidemiology, Imperial College London, London, UK
[8]Modelling and Economics Unit, Centre for Infectious Disease Surveillance and Control, Public Health England, London, UK
[9]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
[10]St Mary's Hospital, Imperial College Healthcare Trust, London, UK

*Corresponding author
†For a full list of IDEA study collaborators, please see *Acknowledgements*

Declared competing interests of authors: Ajit Lalvani is the named inventor for several patents underpinning T-cell-based diagnosis including interferon gamma, enzyme-linked immunospot assay, ESAT-6, CFP-10, Rv3615c, Rv3873 and Rv3879c. He has royalty entitlements from the University of Oxford spin-out company (Oxford Immunotec plc), in which he has held a minority share of equity and he is a member of the Efficacy and Mechanism Evaluation Board. Jonathan J Deeks is a member of the Health Technology Assessment (HTA) Commissioning Board and the HTA Efficient Study Designs Board. Onn Min Kon is chairperson of the UK Joint Tuberculosis Committee. Peter White has received research funding from Otsuka SA for a retrospective study of multidrug-resistant tuberculosis treatment in several eastern European countries outside the submitted work. He received grants from the Medical Research Council during the conduct of the study.

Published May 2019
DOI: 10.3310/hta23230

This report should be referenced as follows:

Takwoingi Y, Whitworth H, Rees-Roberts M, Badhan A, Partlett C, Green N, *et al*. Interferon gamma release assays for Diagnostic Evaluation of Active tuberculosis (IDEA): test accuracy study and economic evaluation. *Health Technol Assess* 2019;**23**(23).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 08/106/02. The contractual start date was in March 2011. The draft report began editorial review in May 2016 and was accepted for publication in September 2016. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

Interferon gamma release assays for Diagnostic Evaluation of Active tuberculosis (IDEA): test accuracy study and economic evaluation

Yemisi Takwoingi,[1] Hilary Whitworth,[2,3] Melanie Rees-Roberts,[2,4] Amarjit Badhan,[2,4] Christopher Partlett,[1] Nathan Green,[5,6,7,8] Aime Boakye,[2,4] Heather Lambie,[2] Luigi Marongiu,[2] Mark Jit,[8,9] Peter White,[5,6,7,8] Jonathan J Deeks,[1] Onn Min Kon[2,4,10] and Ajit Lalvani[2,4]* on behalf of the Interferon Gamma Release Assays for Diagnostic Evaluation of Active Tuberculosis (IDEA) study group†

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[2]Tuberculosis Research Centre, National Heart and Lung Institute, Imperial College London, London, UK
[3]Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, UK
[4]National Institute for Health Research (NIHR) Health Protection Research Unit in Respiratory Infections, Imperial College London, London, UK
[5]NIHR Health Protection Research Unit in Modelling Methodology, Imperial College London, London, UK
[6]Medical Research Council (MRC) Centre for Outbreak Analysis and Modelling, Imperial College London, London, UK
[7]Department of Infectious Disease Epidemiology, Imperial College London, London, UK
[8]Modelling and Economics Unit, Centre for Infectious Disease Surveillance and Control, Public Health England, London, UK
[9]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
[10]St Mary's Hospital, Imperial College Healthcare Trust, London, UK

*Corresponding author a.lalvani@imperial.ac.uk
†For a full list of IDEA study collaborators, please see *Acknowledgements*

Background: Interferon gamma release assays (IGRAs) are blood tests recommended for the diagnosis of tuberculosis (TB) infection. There is currently uncertainty about the role and clinical utility of IGRAs in the diagnostic workup of suspected active TB in routine NHS clinical practice.

Objectives: To compare the diagnostic accuracy and cost-effectiveness of T-SPOT.*TB*® (Oxford Immunotec, Abingdon, UK) and QuantiFERON® TB GOLD In-Tube (Cellestis, Carnegie, VIC, Australia) for diagnosis of suspected active TB and to estimate the diagnostic accuracy of second-generation IGRAs.

Design: Prospective within-patient comparative diagnostic accuracy study.

Setting: Secondary care.

Participants: Adults (aged ≥ 16 years) presenting as inpatients or outpatients at 12 NHS hospital trusts in London, Slough, Oxford, Leicester and Birmingham with suspected active TB.

Interventions: The index tests [T-SPOT.*TB* and QuantiFERON GOLD In-Tube (QFT-GIT)] and new enzyme-linked immunospot assays utilising novel *Mycobacterium tuberculosis* antigens (Rv3615c, Rv2654, Rv3879c and Rv3873) were verified against a composite reference standard applied by a panel of clinical experts blinded to IGRA results.

Main outcome measures: Sensitivity, specificity, predictive values and likelihood ratios were calculated to determine diagnostic accuracy. A decision tree model was developed to calculate the incremental costs and incremental health utilities [quality-adjusted life-years (QALYs)] of changing from current practice to using an IGRA as an initial rule-out test.

Results: A total of 363 patients had active TB (culture-confirmed and highly probable TB cases), 439 had no active TB and 43 had an indeterminate final diagnosis. Comparing T-SPOT.*TB* and QFT-GIT, the sensitivities [95% confidence interval (CI)] were 82.3% (95% CI 77.7% to 85.9%) and 67.3% (95% CI 62.1% to 72.2%), respectively, whereas specificities were 82.6% (95% CI 78.6% to 86.1%) and 80.4% (95% CI 76.1% to 84.1%), respectively. T-SPOT.*TB* was more sensitive than QFT-GIT (relative sensitivity 1.22, 95% CI 1.14 to 1.31; $p < 0.001$), but the specificities were similar (relative specificity 1.02, 95% CI 0.97 to 1.08; $p = 0.3$). For both IGRAs the sensitivity was lower and the specificity was higher for human immunodeficiency virus (HIV)-positive than for HIV-negative patients. The most promising novel antigen was Rv3615c. The added value of Rv3615c to T-SPOT.*TB* was a 9% (95% CI 5% to 12%) relative increase in sensitivity at the expense of specificity, which had a relative decrease of 7% (95% CI 4% to 10%).

The use of current IGRA tests for ruling out active TB is unlikely to be considered cost-effective if a QALY was valued at £20,000 or £30,000. For T-SPOT.*TB*, the probability of being cost-effective for a willingness to pay of £20,000/QALY was 26% and 21%, when patients with indeterminate test results were excluded or included, respectively. In comparison, the QFT-GIT probabilities were 8% and 6%. Although the use of IGRAs is cost saving, the health detriment is large owing to delay in diagnosing active TB, leading to prolonged illness. There was substantial between-patient variation in the tests used in the diagnostic pathway.

Limitations: The recruitment target for the HIV co-infected population was not achieved.

Conclusions: Although T-SPOT.*TB* was more sensitive than QFT-GIT for the diagnosis of active TB, the tests are insufficiently sensitive for ruling out active TB in routine clinical practice in the UK. Novel assays offer some promise.

Future work: The novel assays require evaluation in distinct clinical settings and in immunosuppressed patient groups.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and the NIHR Health Protection Research Unit in Respiratory Infections, Imperial College London, London, UK.

Contents

List of tables

List of figures

List of abbreviations

BAL	bronchoalveolar lavage
BCG	bacillus Calmette–Guérin
CD4	cluster of differentiation 4
CFP-10	culture filtrate protein 10
CI	confidence interval
CRF	case report form
CT	computed tomography
DMG	Data Management Group
ELISA	enzyme-linked immunosorbent assay
ELISpot	enzyme-linked immunospot assay
ESAT-6	early secretory antigenic 6 kDa
GEE	generalised estimating equation
HIV	human immunodeficiency virus
HTA	Health Technology Assessment
IDEA	Interferon gamma release assays for Diagnostic Evaluation of Active tuberculosis study
IFN-γ	interferon gamma
IGRA	interferon gamma release assay
LTBI	latent tuberculosis infection
MRI	magnetic resonance imaging
Mtb	*Mycobacterium tuberculosis*
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NPV	negative predictive value
PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PET	positron emission tomography
PI	principal investigator
PPD	purified protein derivative
PPI	patient and public involvement
PPV	positive predictive value
QALY	quality-adjusted life-year
QFT-GIT	QuantiFERON GOLD In-Tube
QFT-Plus	QuantiFERON-TB Gold Plus
R&D	research and development
RPMI	Roswell Park Memorial Institute
SFC	spot-forming cell
SMG	Study Management Group
SSC	Study Steering Committee
STARD	Standards for the Reporting of Diagnostic Accuracy Studies
TB	tuberculosis
TST	tuberculin skin test

Plain English summary

Tuberculosis (TB) is one of the world's most important infectious diseases. In 2014, 1.5 million deaths were caused by the disease – about one death every 25 seconds. Traditional diagnosis of TB is based partly on the tuberculin skin test. Blood tests such as QuantiFERON GOLD In-Tube (QFT-GIT; Cellestis, Carnegie, VIC, Australia) and T-SPOT.*TB*® (Oxford Immunotec, Abingdon, UK) are now available. However, these two tests are not used as part of current NHS practice because of the lack of evidence about how well the tests perform when diagnosing symptomatic (active) TB in routine clinical practice.

The purpose of our study was to compare the ability of QFT-GIT and T-SPOT.*TB* to differentiate people with active TB from those without active TB in a population suspected of the disease. We also assessed new blood tests that are currently being developed for diagnosis of active TB.

We recruited 1074 patients with suspected TB from 14 NHS hospitals in London, Slough, Oxford, Leicester and Birmingham into our study. We found that T-SPOT.*TB* correctly detected more people with active TB than QFT-GIT; T-SPOT.*TB* would miss about 18 people out of every 100, whereas QFT-GIT would miss about 33 people out of every 100 with active TB.

For this reason, neither test is good enough for routine clinical use because the number of people with active TB who are incorrectly diagnosed as not having active TB is unacceptably high. In addition, neither test is good value for money. However, we did find that some of the newer blood tests performed better than T-SPOT.*TB* and their usefulness should be further investigated.

Scientific summary

Background

Interferon gamma release assays (IGRAs) are blood tests recommended for diagnosis of tuberculosis (TB) infection. The two types of commercially available IGRAs are QuantiFERON GOLD In-Tube (QFT-GIT; Cellestis, Carnegie, VIC, Australia), a whole-blood enzyme-linked immunosorbent assay (ELISA), and T-SPOT.*TB*® (Oxford Immunotec, Abingdon, UK), an enzyme-linked immunospot assay (ELISpot). There is currently uncertainty in the role and clinical utility of IGRAs in the diagnostic workup of suspected active TB in routine NHS clinical practice.

Aim

To evaluate and compare the diagnostic accuracy and cost-effectiveness of IGRAs for the diagnosis of active TB.

Objectives

Primary objectives

- To compare the diagnostic performance [sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)] of T-SPOT.*TB* and QFT-GIT for the diagnosis of active pulmonary and extrapulmonary TB in routine clinical practice.
- To develop an evidence-based optimal testing algorithm that defines the role of IGRAs in the diagnostic workup of suspected active TB.
- To deliver the above objectives for a key subgroup: human immunodeficiency virus (HIV) co-infected patients (the highest-risk subgroup of TB).
- To quantify and compare the cost-effectiveness of a range of possible testing strategies against the present testing regime.

Secondary objectives

- To quantify the sensitivity, specificity, and PPV and NPV of T-SPOT.*TB* and QFT-GIT in a number of key patient subgroups, such as patients with pre-existing diabetes mellitus, end-stage renal failure and iatrogenic immunosuppression.
- To quantify the use of second-generation IGRAs compared with existing commercially available assays.

Methods

We used a within-patient (paired) design to compare test accuracy by applying all IGRAs to blood samples from each patient. The final diagnosis of participants was verified using a composite reference standard (based on the Dosanjh's criteria) applied by a panel of clinicians blinded to local (routine) and study IGRA results. The diagnostic accuracy of early secretory antigenic 6 kDa (ESAT-6) and culture filtrate protein 10 (CFP-10) (together these antigens constitute T-SPOT.*TB*) and four new ELISpot-based assays utilising novel *Mycobacterium tuberculosis* antigens (Rv3615c, Rv2654, Rv3879c and Rv3873) were also evaluated, both individually and in test combinations. The test combinations were compared with T-SPOT.*TB*.

Statistical analysis

The main clinical utility of IGRAs in the assessment of suspected active TB is likely to be in their NPV, which enables clinicians to reliably rule out TB from the differential diagnoses. This depends on the sensitivity of the test and the prevalence of active TB in the tested population. To detect a 10% difference in sensitivity between T-SPOT.*TB* and QFT-GIT (assuming a sensitivity of 85% for T-SPOT.*TB* and of 75% for QFT-GIT) at the 5% significance level with 90% power, 855 participants were required, assuming a 40% prevalence of active TB. To allow for missing data, indeterminate index test and reference standard results, withdrawal of consent and possible logistical errors, we aimed to recruit 1012 patients. For the HIV-positive subgroup, we computed sample size based on sensitivities of 85% and 65% for T-SPOT.*TB* and QFT-GIT, respectively. We assumed a 20% prevalence of active TB and so required 390 participants to detect a 20% difference with 80% power.

We estimated sensitivity, specificity, PPV and NPV and likelihood ratios for each test and combinations of tests. For test comparisons, relative test performance was assessed by comparing the sensitivity and specificity of one test relative to those of another test. The comparisons between different IGRAs were done using generalised estimating equation models to exploit the paired nature of the data. Variation in the relative performance of T-SPOT.*TB* and QFT-GIT with HIV infection status and other clinical characteristics was investigated by including one covariate at a time in the models.

Economic evaluation

The economic analyses were based on the main study cohort. The diagnostic tests performed, their costs, and time taken between decision points involving each test were considered in the economic analyses. The analysis was undertaken from a NHS perspective. No discounting was required, as the diagnostic process occurs over a relatively short time period. For preference estimates we followed the National Institute for Health and Care Excellence (NICE) reference case, using quality-of-life weights obtained from the literature. A decision tree model was developed to calculate the incremental costs and incremental health utilities [quality-adjusted life-years (QALYs)] of changing from current practice to using an IGRA as an initial rule-out test. The model was parameterised using the IGRAs for Diagnostic Evaluation of Active tuberculosis (IDEA) study clinical patient records and relevant current literature.

Results

Between 25 November 2011 and 31 August 2013, the IDEA study recruited 1074 adults (aged ≥ 16 years) presenting as inpatients or outpatients at 10 NHS hospital trusts in London, Slough, Oxford, Leicester and Birmingham with suspected active TB. We refer to this group as the main study cohort. Owing to low recruitment of HIV-positive patients, the study was extended to 31 December 2014 to recruit only this group of patients. Two additional hospital trusts were added. A total of 263 HIV-positive patients were recruited from 12 NHS trusts between 25 November 2011 and 19 December 2014. This is the HIV-positive substudy cohort.

In the main cohort, the median age of the 845 patients included in the analyses was 38 (range 16–86) years. Most (59.3%) were male, approximately half (48.2%) were of Indian ethnicity and 135 (16.0%) were HIV positive. There were 88 (10.4%) patients with pre-existing diabetes mellitus, 12 (1.4%) patients with chronic/end-stage renal failure and 105 (12.3%) patients who were on immunosuppressive therapy. In the HIV-positive substudy cohort, the median age of the 201 patients included in the analyses was 43 (range 18–79) years. The majority (67.7%) were male and a substantial number were of black (45.3%) or white (37.8%) ethnicity.

Principal findings of diagnostic accuracy in main study cohort

A total of 363 (43.0%) patients had a diagnosis of active TB (culture-confirmed and highly probable TB cases), whereas active TB was excluded in 439 (52.0%) patients. The remaining 43 (5.1%) patients had an indeterminate final diagnosis and were excluded from all analyses of diagnostic accuracy. The rate of indeterminate IGRA results was higher for QFT-GIT (9.6%) than for T-SPOT.*TB* (7.0%). Indeterminate IGRA

results were excluded from the main analyses. Comparing the two IGRAs, sensitivities were 82.3% [95% confidence interval (CI) 77.7% to 85.9%] and 67.3% (95% CI 62.1% to 72.2%), whereas specificities were 82.6% (95% CI 78.6% to 86.1%) and 80.4% (95% CI 76.1% to 84.1%) for T-SPOT.*TB* and QFT-GIT, respectively. The sensitivity of T-SPOT.*TB* was superior to that of QFT-GIT [relative sensitivity 1.22, 95% CI 1.14 to 1.3; $p < 0.001$), but there was no statistical evidence of a difference in specificity [relative specificity 1.02, 95% CI 0.97 to 1.08; $p = 0.3$]. In sensitivity analyses with indeterminates included as test positives (because only a negative result can rule out TB), conclusions about differences in the sensitivities and specificities of T-SPOT.*TB* and QFT-GIT were unchanged.

The sensitivities of T-SPOT.*TB* and QFT-GIT were lower in patients with HIV co-infection than in the HIV-negative subgroup. Similarly, the two IGRAs had lower sensitivities in patients with diabetes mellitus than in patients without diabetes mellitus. Specificity was higher in HIV-positive patients than in HIV-negative patients, but was lower in patients with diabetes mellitus than in those without diabetes mellitus. Although there appeared to be differences in test performance between subgroups, there was no statistical evidence of an effect of HIV infection status or diabetes mellitus on relative test performance. The findings from these analyses should be taken with caution, as the number of test results in some subgroups was small. Because data were few, subgroup analyses were not possible for the other two key subgroups: those patients with end-stage renal failure and those on immune suppressants.

The most promising novel antigen was RV3615c, with a sensitivity that was higher than that of the other antigens. Test combinations including Rv3615c showed higher sensitivity than T-SPOT.*TB*. The added value of Rv3615c to T-SPOT.*TB* was a 9% (95% CI 5% to 12%) relative increase in sensitivity at the expense of specificity, which showed a relative decrease of 7% (95% CI 4% to 10%). The combination of CFP-10 with Rv3615c (i.e. akin to replacing ESAT-6 in T-SPOT.*TB*) showed a relative increase of 7% (95% CI 3% to 11%) in sensitivity and a relative decrease of 5% (95% CI 2% to 8%) in specificity.

Principal findings of diagnostic accuracy in the HIV-positive substudy cohort

A total of 32 (15.9%) patients had a diagnosis of active TB (culture-confirmed and highly probable TB cases), whereas active TB was excluded in 165 (82.1%) patients. The remaining four (2.0%) patients had an indeterminate final diagnosis and were excluded from all analyses. The indeterminate rate was 19.5% for QFT-GIT and 23.1% for T-SPOT.*TB*. The difference of 3.6% (95% CI –4.5% to 11.6%) was not significant ($p = 0.4$). Excluding indeterminate IGRA results, the sensitivities for T-SPOT.*TB* and QFT-GIT were 62.8% (95% CI 44.1% to 78.3%) and 56.1% (95% CI 38.3% to 72.4%), respectively, and the specificities were 83.4% (95% CI 75.7% to 88.9%) and 91.7% (95% CI 85.4% to 95.4%), respectively. The sensitivity of T-SPOT.*TB* was higher than that of QFT-GIT, with a relative sensitivity of 1.12 (95% CI 0.87 to 1.44). There was no statistical evidence of a difference ($p = 0.4$). In contrast, the specificity of T-SPOT.*TB* was significantly lower ($p = 0.02$) than that of QFT-GIT with a relative specificity of 0.91 (95% CI 0.84 to 0.99).

When indeterminate IGRA results were included in a sensitivity analysis, there was a small increase in the sensitivity of QFT-GIT but a large increase in the sensitivity of T-SPOT.*TB* owing to the higher indeterminate rate for T-SPOT.*TB* (18.2%) among active TB cases than that of QFT-GIT (5.9%). Nevertheless, there was no statistical evidence of a difference in sensitivity ($p = 0.1$) or specificity ($p = 0.1$).

Main findings of economic evaluation

Tuberculosis diagnosis rarely followed the idealised diagnostic pathways and there was considerable individual-level variability. This implies that costs of and time delays in diagnosis may be very different from typical assumptions made in economic analyses. The number and order of diagnostic tests that were performed varied between patients, as well as between final diagnosis categories. For instance, nearly all active TB patients were given a culture test and sputum spear microscopy, whereas approximately 20% and 25% of non-active TB patients were given a culture test and sputum spear microscopy, respectively. The median cost of diagnosis was highest for unconfirmed diagnosis patients (£502) followed by the non-culture-confirmed active TB patients (£476).

The use of current IGRA tests for ruling out active TB would be unlikely to be considered cost-effective if a QALY was to be valued at £20,000 or £30,000. T-SPOT.*TB* performed better than QFT-GIT in the cost-effectiveness analysis. The probability of being cost-effective for a willingness to pay of £20,000/QALY was 26% and 21% for T-SPOT.*TB* when patients with indeterminate test results were excluded or included, respectively. In comparison, the QFT-GIT probabilities were 8% and 6%, respectively.

For the study cohort, the cost saving in these scenarios ranged from £65,120 to £86,850, but the health detriment in QALYs was between –6.50 and –3.58.

Stratifying the main study cohort by HIV infection status, the HIV-negative group of patients had results similar to those from the analyses of the entire cohort. However, cost-effectiveness results were worse for the HIV-positive group, with the probability of being cost-effective at a willingness to pay of £20,000/QALY of approximately 12% and 9% for T-SPOT.*TB* and QFT-GIT, respectively, when patients with indeterminate IGRA results were excluded. When patients with indeterminate IGRA results were included, the probability was 4% for both IGRAs. The HIV-positive group had ranges of cost savings and health detriment of £42,110 to £106,090 and –7.86 to –5.91, respectively.

Although IGRAs are cost saving, the health detriment is large because of delay in diagnosing active TB leading to prolonged illness. Whether there is a net health detriment or gain for the patient cohort as a whole depends on the prevalence of active TB, the performance characteristics of the rule-out test and the length of delay introduced by adding the initial rule-out test.

Conclusions

Implications for health care

Despite the significantly higher sensitivity of T-SPOT.*TB* over QFT-GIT, neither IGRA can be used routinely as a reliable rule-out test for suspected active TB in this patient population in secondary care. Neither IGRA was cost-effective in this setting. However, in patients in whom there is suspicion of TB, but the pre-test probability is low, the NPV of a negative T-SPOT.*TB* result would be correspondingly higher. Hence, it would not be unreasonable to use a negative T-SPOT.*TB* result to weigh the odds in favour of excluding TB from the differential diagnosis as long as the test result is interpreted with an awareness of the limited sensitivity as shown by this study.

The incorporation of novel antigens into T-SPOT.*TB*, in particular Rv3615c, provided high diagnostic sensitivity values coupled with a modest reduction in specificity. Notably, replacing ESAT-6 with Rv3615c also conferred higher sensitivity than T-SPOT.*TB*. This observation is relevant for TB control internationally because one of the leading TB vaccine candidates in clinical trials, H56/IC31 [Statens Serum Institute, Copenhagen, Denmark; Aeras, Rockville, MD, USA; formulated with Valneva's IC31® proprietary adjuvant (Lyon, France)] incorporates ESAT-6. The vaccine is protective in the non-human primate model and is likely to be licensed if it proves to be protective in humans. If rolled out, vaccinated individuals are likely to develop T-cell responses to ESAT-6 that would lead to false-positive IGRA results, akin to the current scenario with bacillus Calmette–Guérin (BCG) vaccination inducing false-positive tuberculin skin test results. Replacing ESAT-6 with Rv3615c may be a potential solution because a CFP-10- and Rv3615c-based IGRA would have significantly higher sensitivity than existing IGRAs and specificity would not be compromised in H56/IC31-vaccinated individuals.

Recommendations for research

The second-generation IGRAs evaluated in the IDEA study do not need to be re-evaluated in a UK routine practice setting because this study provided an equally rigorous evaluation of these novel assays, as it did for conventional IGRAs. However, the novel assays require evaluation in distinct clinical settings with much lower or much higher prevalence of active TB and in immunosuppressed subgroups. A new generation of QFT-GIT, QuantiFERON-TB Gold Plus (QFT®-Plus; Qiagen GmbH, Hilden, Germany), was recently launched.

Although the QFT-GIT has been replaced by the QFT-GIT-Plus since our study was conducted, its diagnostic accuracy does not appear to be significantly better than QFT-GIT and there is no evidence it is as sensitive as T-SPOT.*TB*. A comparative accuracy study of the novel assays and QFT-Plus may be needed.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research (NIHR) and the NIHR Health Protection Research Unit in Respiratory Infections, Imperial College London, London, UK.

Chapter 1 Introduction

Background

In 2014, globally, an estimated 9.6 million cases of tuberculosis (TB) and 1.5 million deaths caused by the disease were reported to the World Health Organization.[1] Co-infection with TB and human immunodeficiency virus (HIV) accounts for a significant proportion of cases globally (12%) and together these infections are the biggest infectious causes of death. Health inequalities are exacerbated by the burden that TB places on the most vulnerable and poor around the world. Over the past 20 years, worldwide TB incidence and mortality have declined. However, despite this, the prevalence of TB still remains unacceptably high. Furthermore, there is not yet evidence of a reduction in the number of cases in England, particularly in major cities such as London and Birmingham.

In England, 6520 cases of TB were reported in 2014, with an incidence rate of 12.0 per 100,000.[2] Of these, 2572 cases were in London, where the incidence rate was 30.1 per 100,000.[2] Other urban high-incidence areas include Leicester, Birmingham, Luton, Manchester and Coventry. The majority of cases in England (72%) occurred in new entrants who were born outside the UK.[2]

Tuberculosis is caused by active infection with *Mycobacterium tuberculosis* (Mtb). Upon initial infection with Mtb, a state of asymptomatic dormancy typically occurs. In most infected individuals this results in a prolonged and perhaps lifelong latent TB infection (LTBI). In others, however, a breakdown of immune control and activation of infection results in symptomatic disease, which can be fatal without treatment. Co-infection with HIV dramatically increases the likelihood of progression from latent to active disease.

Active TB can manifest with pulmonary or extrapulmonary phenotypes. In its active pulmonary form, TB is highly contagious. The diagnosis of active TB is central to preventing the spread of disease and thus controlling the TB epidemic.[3] However, the slow speed and poor sensitivity of existing diagnostic tools often lead to delays diagnosis and treatment of the disease.[4]

Diagnosis of tuberculosis

Conventional methods of diagnosing active TB rely primarily on the identification of Mtb bacilli, as well as imaging affected areas. Smear microscopy is quick and inexpensive, but typically lacks sensitivity. Although cell culture is considered the 'gold standard' for diagnosis of active TB because of its higher sensitivity, it is often slow, taking up to 6 or 8 weeks to give a result. Polymerase chain reaction (PCR)-based tests for Mtb, such as the new GeneXpert Mtb/RIF® (Sunnyvale, Cepheid, CA, USA), can be quick and sensitive, but they are typically expensive and require a high standard of infrastructure. Imaging techniques such as chest radiography and computed tomography (CT) are quick and usually sensitive, but are not specific. Furthermore, they can be expensive, and, again, require a high standard of infrastructure (for example in the case of CT). All of these tests tend to be less accurate in cases of active TB with HIV co-infection, and also in cases of extrapulmonary TB.

Currently available tests for LTBI include the tuberculin skin test (TST) and interferon gamma release assays (IGRAs). The TST measures the in vivo delayed-type hypersensitivity response to intradermal inoculation of a crude mixture of mycobacterial antigens. Because this mixture contains antigens also present in bacillus Calmette–Guérin (BCG), the test can be confounded by prior BCG vaccination. IGRAs, on the other hand, detect ex vivo interferon gamma (IFN-γ) release from T cells (lymphocytes that play a key role in cell-mediated immunity) in response to Mtb-specific antigens, early secretory antigenic 6 kDa (ESAT-6) and culture filtrate protein 10 (CFP-10). These antigens are absent from the BCG vaccine and most environmental mycobacteria, and thus IGRAs tend to be more specific than the TST.[5] The two types of commercially available IGRAs are

QuantiFERON GOLD In-Tube (QFT-GIT; Cellestis, Carnegie, VIC, Australia), a whole-blood enzyme-linked immunosorbent assay (ELISA), and T-SPOT.*TB*® (Oxford Immunotec, Abingdon, UK), an enzyme-linked immunospot assay (ELISpot); both IGRAs utilise peripheral blood mononuclear cells (PBMCs). These tests have been recommended by the UK, European and North American guidelines for diagnosis of LTBI.[6]

Although typically used in the diagnosis of LTBI, TSTs and IGRAs actually detect Mtb infection in its entirety (i.e. active or latent). The role of the tests within published guidelines for diagnostic evaluation of suspected active TB to date has been limited because of their low specificity for the disease: they could never confirm a diagnosis of active TB because they cannot differentiate latent and active infection. However, because Mtb infection is a pre-requisite for TB disease, reliable determination of infection status could accelerate diagnostic assessment by enabling rapid exclusion of TB (within 24 hours) when the result is negative.

In order for the IGRA or TST to reliably rule out a diagnosis of Mtb infection and thus TB disease, the sensitivity of the test must be very high (> 95%). The sensitivity and specificity of IGRAs compared with the TST in active TB have been examined in a number of studies, varying in size and quality.[7] IGRAs are typically more specific than the TST for diagnosing Mtb infection and T-SPOT.*TB* is more sensitive than the TST for diagnosing TB. However, the diagnostic accuracy of T-SPOT.*TB* and QFT-GIT has not been compared directly head to head in suspected active TB in the UK, nor comprehensively assessed in immunosuppressed patients. Therefore, there is uncertainty in the role and clinical utility of IGRAs in the diagnostic workup of suspected TB, as well as their cost-effectiveness in UK NHS practice.

Aim and study objectives

Aim

To evaluate and compare the diagnostic accuracy and cost-effectiveness of IGRAs with conventional testing for diagnosis of active TB. Specifically, the study aimed to determine the sensitivity, specificity, positive predictive values (PPVs) and negative predictive values (NPVs), and likelihood ratios of T-SPOT.*TB* and QFT-GIT for diagnosis of active TB in routine NHS clinical practice. Second-generation IGRAs were also evaluated.

Study objectives

Primary objectives

1. To compare the diagnostic accuracy of T-SPOT.*TB* and QFT-GIT for the diagnosis of active pulmonary and extrapulmonary TB in routine clinical practice.
2. To develop an evidence-based optimal testing algorithm that defines the role of IGRAs in the diagnostic workup of suspected active TB.
3. To deliver the objectives above for a key subgroup: HIV co-infected patients (the highest-risk subgroup of TB).
4. To quantify and compare the cost-effectiveness of a range of possible testing strategies against the present testing regime.

Secondary objectives

1. To quantify the sensitivity, specificity, and positive and NPVs of T-SPOT.*TB* and QFT-GIT in a number of key patient subgroups, such as patients with pre-existing diabetes mellitus, end-stage renal failure and iatrogenic immunosuppression.
2. To quantify the use of second-generation IGRAs compared with existing commercially available assays.

Chapter 2 Methods

This chapter describes the study design and methods for the evaluation of the diagnostic accuracy of IGRAs in active TB. Our report adheres to the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) guideline,[8] as shown in *Appendix 1*. Methods for the health economic evaluation are described separately in *Chapter 6*.

Overview of the study design

This prospective multicentre study comparing the accuracy of IGRAs was conducted in routine clinical practice in the UK. Adults presenting with suspected active TB at NHS outpatient or inpatient services to participating hospitals in London, Slough, Oxford, Leicester and Birmingham were recruited. We used a within-patient design to compare test accuracy by performing all IGRAs on blood samples from each patient with the presence or absence of active TB verified using the reference standard. This design minimises between-patient variability while also allowing estimation of the accuracy of combinations of IGRAs. Blood samples for IGRA testing were collected from patients at baseline and follow-up (2 and 6 months). If necessary, and when available, TST results were used as part of the composite reference standard for verifying the final diagnosis of patients. The TST results were obtained from routine clinical care and so the availability of TST results reflects local practice in participating hospitals.

Participants

Inclusion criteria

Adults (aged $\geq$ 16 years) presenting with suspected (pulmonary or extrapulmonary) active TB to NHS outpatient or inpatient services were included. To replicate clinical practice, patients with a previous diagnosis of TB and/or history of TB treatment were recruited. However, they were excluded from the analyses on the basis that, when evaluating patients with suspected active TB, the clinician should not perform an IGRA because any immunological biomarker would remain positive and thus affect test accuracy. The study population was expected to be representative of the national TB burden in terms of ethnic mix and range of comorbidities. A key subgroup was HIV-positive patients.

Exclusion criteria

Participants aged < 16 years or those unable to give informed consent were excluded.

Setting

Patients were recruited at the point of diagnostic workup from 14 hospitals in 10 NHS trusts in the UK.

Recruitment process

An overview of the recruitment process is shown in *Figure 1*. Potential participants presenting to participating NHS centres were referred to a TB research nurse by the attending clinician. The nurse then screened participants to ensure that they were eligible for the study according to the inclusion/exclusion criteria stated above. Each potential participant was provided with an information sheet and a verbal description of the study. Participants were included in the study if they were willing and informed consent was obtained.

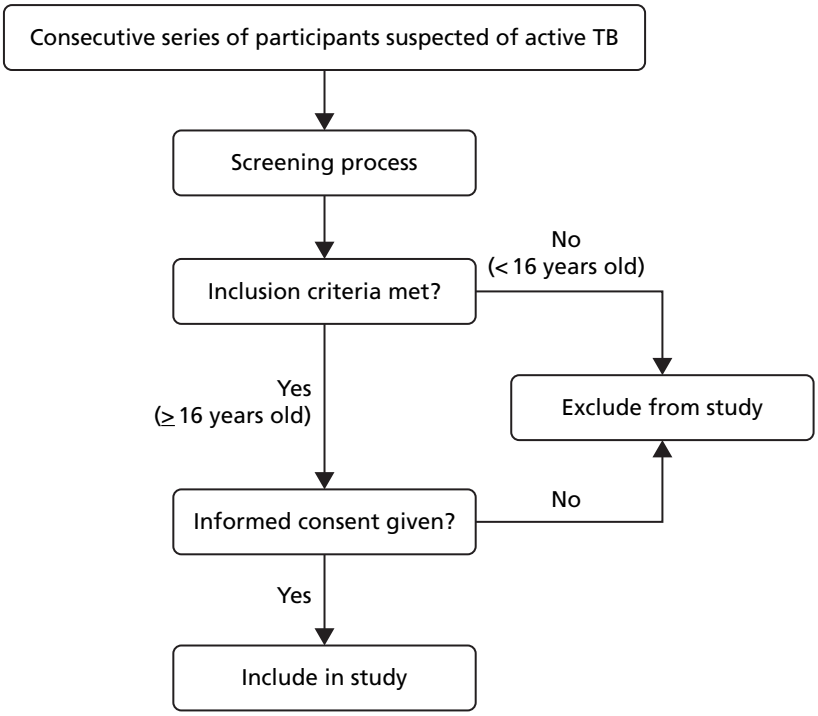

FIGURE 1 Overview of recruitment process.

Data collection and management

Follow-up

Participants were seen by research nurses for follow-up visits at 2 and 6 months after recruitment. Follow-up visits were scheduled to be carried out at the same time as the patient's routine clinic appointments. At these visits blood was collected for IGRA testing and data on patient diagnosis were obtained. If patients were no longer being seen as part of their routine NHS care, they were not required to attend study follow-up visits. In such cases, information about a patient's diagnosis was obtained from their medical notes. In addition to following up patients at 2 and 6 months, a review of patient records was performed up to 1 year post recruitment, if required, in order to obtain a final diagnosis.

Data management

A case report form (CRF) was used to collect patient data at each recruiting hospital after a patient consented to participate in the IGRAs for Diagnostic Evaluation of Active tuberculosis (IDEA) study. Following receipt of the CRFs, data were entered into an electronic password-protected database. The results of study IGRAs and all other study laboratory tests were entered into a separate secure database. This laboratory database was accessible only to specific laboratory staff. Thus, the Study Management Group (SMG) responsible for day-to-day management of the IDEA study was blinded to the IGRA results. Furthermore, NHS clinicians responsible for routine care and diagnosis of patients involved in the study were also blinded to study IGRA results. To enable preparation of study reports for meetings of the independent oversight committee [Study Steering Committee (SSC)], the study statistician was granted access to the clinical database and excerpts of the laboratory database at specific periods during the study.

Sample collection

Blood samples

Blood sampling was done at three time points: at baseline and at follow-up at 2 and 6 months. Only blood samples collected at baseline were used in the assessment of the index tests and for the reference standard.

Samples taken at follow-up were used for further IGRA testing (but not analysed as part of the IDEA study) and also stored for future research as consented by patients. Blood-taking (venepuncture) procedures were carried out in accordance with local trust venepuncture guidelines. The baseline sample was obtained no later than 14 days after the start of treatment for TB or within 7 days of consent, whichever was earlier, depending on the patient's diagnosis. For patients with a diagnosis of active TB, sarcoidosis or other non-TB diagnosis, follow-up study blood samples were taken 2 and 6 months after the start of treatment. For patients with a diagnosis of latent TB, when possible and if the patient returned to the same clinical team, blood samples were taken 3 and 6 months after the initiation of treatment (if treatment was indicated and given).

A 35-ml blood sample was taken at each time point. Blood was collected into heparinised collection tubes and QFT-GIT collection tubes for T-SPOT.*TB* and QFT-GIT assays, respectively. Furthermore, heparinised blood was collected for performance of second-generation IGRAs and to store plasma and PBMCs for future research. In addition, blood was collected into uncoated collection tubes in order to store serum for future research.

Blood samples were transported on the same day as sample collection either by a member of the research team or by courier in the appropriate United Nations-type approved packaging to the TB Research Centre for testing. All samples were processed within 6 hours of blood collection for London-based sites and within 8 hours for outer London sites. Excess PBMCs were stored in a liquid nitrogen tank, and serum and plasma were stored in a –80 °C freezer at the TB Research Centre (led by Professor Ajit Lalvani) at Imperial College London (St. Mary's Hospital).

Diagnostic bronchoscopy samples

In patients with sputum smear-negative pulmonary TB, diagnostic bronchoscopies were performed and bronchoalveolar lavage (BAL) was obtained as part of routine clinical care. When surplus BAL samples were available and not required for diagnostic procedures, aliquots were cryopreserved and stored in the research biorepository for subsequent testing by IGRA. The bronchoscopic procedure, along with collection of surplus BAL samples, were applicable only for patients recruited at St. Mary's Hospital, Imperial College Healthcare NHS Trust, in accordance with set clinical practice guidelines. The BAL sample consent was covered under the consent form as 'tissue samples'. However, patients were informed if a surplus BAL sample was kept for the IDEA study.

Index tests

Interferon gamma release assays

Two types of commercially available IGRAs, QFT-GIT and T-SPOT.*TB*, were evaluated. In addition, a new ELISpot-based assay utilising novel antigens (Rv3615c, Rv2654, Rv3879c and Rv3873) was evaluated. The performance of each antigen was evaluated individually and in combinations that included either ESAT-6 and CFP-10, the two antigens that constitute T-SPOT.*TB*, or both. IGRA testing is not standard practice for HIV-negative patients suspected of having TB in the hospitals of our consortium and is not currently recommended for HIV-positive patients suspected of having TB. However, if IGRAs were used locally at participating hospitals as part of the routine diagnostic workup of patients, we recorded the tests done but we did not analyse the test results as study results for the IDEA study. Thus, only from IGRAs performed in our research laboratory specifically for this study were recorded and assessed. Laboratory staff performing the IGRAs and recording the test results were blinded to clinical information and reference standard results.

Analysis of T-SPOT.*TB* and novel antigens

Peripheral blood mononuclear cells were isolated from heparinised whole blood using the Ficoll-Paque™ density centrifugation method (GE Healthcare Bio-Science, Uppsala, Sweden), as described by Whitworth *et al.*[6] In brief, whole blood was diluted in Roswell Park Memorial Institute (RPMI) 1640 medium (Sigma-Aldrich, Dorset, UK) and layered onto Ficoll-Paque™ Plus at a ratio of 2 : 1 in 50-ml Falcon® centrifuge tubes

(Corning Science Mexico S.A. de C.V., Reynosa, USA). The tubes were centrifuged for 20 minutes at 18–25 °C and the cloudy PBMC layer was aspirated into fresh RPMI 1640 medium. Cells were washed with fresh RPMI 1640 medium and counted using trypan blue stain for use with the Countess® Automated Cell counter (Life Technologies, Eugene, OR, USA). T-SPOT.*TB* was applied to the freshly isolated PBMCs as per the manufacturer's instructions[9] and as described by Whitworth *et al.*[6]

Cells were resuspended in AIM-V® Serum-Free Medium (Gibco by Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA) at a concentration of 2.5 million cells/ml and 250,000 cells per well were incubated overnight (18 hours) at 37 °C with Mtb-specific antigens (ESAT-6, CFP-10, Rv3615c, Rv2654, Rv3879c and Rv3873) individually, and positive (phytohaemagglutinin) and negative (RPMI 1640 medium) controls in a 96-well plate, pre-coated with IFN-γ-specific monoclonal capture antibodies (included in T-SPOT.*TB* kit). Thus, a total of eight wells were used per patient and samples from 12 patients were included on a plate. Overnight incubation of the cells with antigens allows for IFN-γ secretion from activated Mtb-specific effector T cells present in the culture. Secreted IFN-γ binded to the pre-coated IFN-γ-specific monoclonal capture antibodies on the membrane of each well. After incubation, wells were washed with phosphate-buffered saline and an alkaline phosphatase-conjugated secondary IFN-γ-specific monoclonal antibody was added to bind to any captured IFN-γ. For a visible representation of the spots on the membrane, an alkaline phosphatase chromogen substrate was added.

Each spot formed on the membrane signifies IFN-γ release by a single activated Mtb antigen-specific T cell. Spot-forming cells (SFCs) are expected to be detected in positive control wells and absent in negative control wells. SFCs in TB antigen-stimulated wells indicate infection.

Spot-forming cells were counted using an automated ELISpot plate reader (AID ELISpot read system ELRIFL04; Advanced Imaging Devices GmbH, Strasbourg, Germany), with saturation level set at 60%. For individual antigens, test results were classified as negative, positive, borderline (equivocal) or indeterminate (invalid) by subtracting the spot count in the negative control well from the spot count in each panel, according to the algorithm illustrated in *Figure 2* (based on the package insert for T-SPOT.*TB*).[9] Panels A, B, C, D, E and F correspond to ESAT-6, CFP-10, Rv3615c, Rv2654c, Rv3879c and Rv3873, respectively.

For T-SPOT.*TB* as well as other antigen combinations, if the positive control spot count was < 20, the result was deemed indeterminate, unless the response to one of the Mtb antigens was positive (or borderline), in which case the test result for the combination was deemed positive (or borderline). Thus, we applied an 'OR' rule (at least one antigen spot count deemed positive) for antigen combinations. For example, for T-SPOT.*TB*, a result was positive if the negative-control spot count was ≤ 10 and either 'panel A minus negative control' or 'panel B minus negative control' was ≥ 8 spots. This implies the T-SPOT.*TB* result was negative if both 'panel A minus negative control' and 'panel B minus negative control' were negative (≤ 4 spots). In the IDEA study, borderline test results (5–7 spots) were considered as positives.

Analysis of QuantiFERON GOLD In-Tube

The QFT-GIT assay was performed in two stages as per the manufacturer's instructions[10] and as described in Whitworth *et al.*[6] First, whole blood was collected from each participant into three QFT-GIT tubes containing a negative control, mitogen-positive control and Mtb antigens [ESAT-6, CFP-10 and TB7.7 (also known as Rv2654c, a possible PhiRv2 prophage protein) combined], as provided by the manufacturer.[10] The tubes were incubated at 37 °C for 16–24 hours to allow IFN-γ secretion from antigen-specific effector T cells into the extracellular fluid (plasma). After incubation, tubes were centrifuged and 150 µl of plasma was collected and stored in a 96-well plate for up to 4 weeks at 2–8 °C prior to performing the remainder of the assay.

To perform the ELISA step, 50 µl of plasma from each of the QFT-GIT tubes (i.e. containing a mitogen control, negative control and Mtb antigens) was transferred to wells of another 96-well plate pre-coated with IFN-γ-specific monoclonal capture antibodies and incubated with a conjugate [an IFN-γ-specific antibody conjugated to horseradish peroxidase (included in T-SPOT.*TB* kit)] for 2 hours at room temperature (22 °C). Plasma samples in each well were mixed thoroughly using a microplate shaker (PMS-1000i Microplate Shaker;

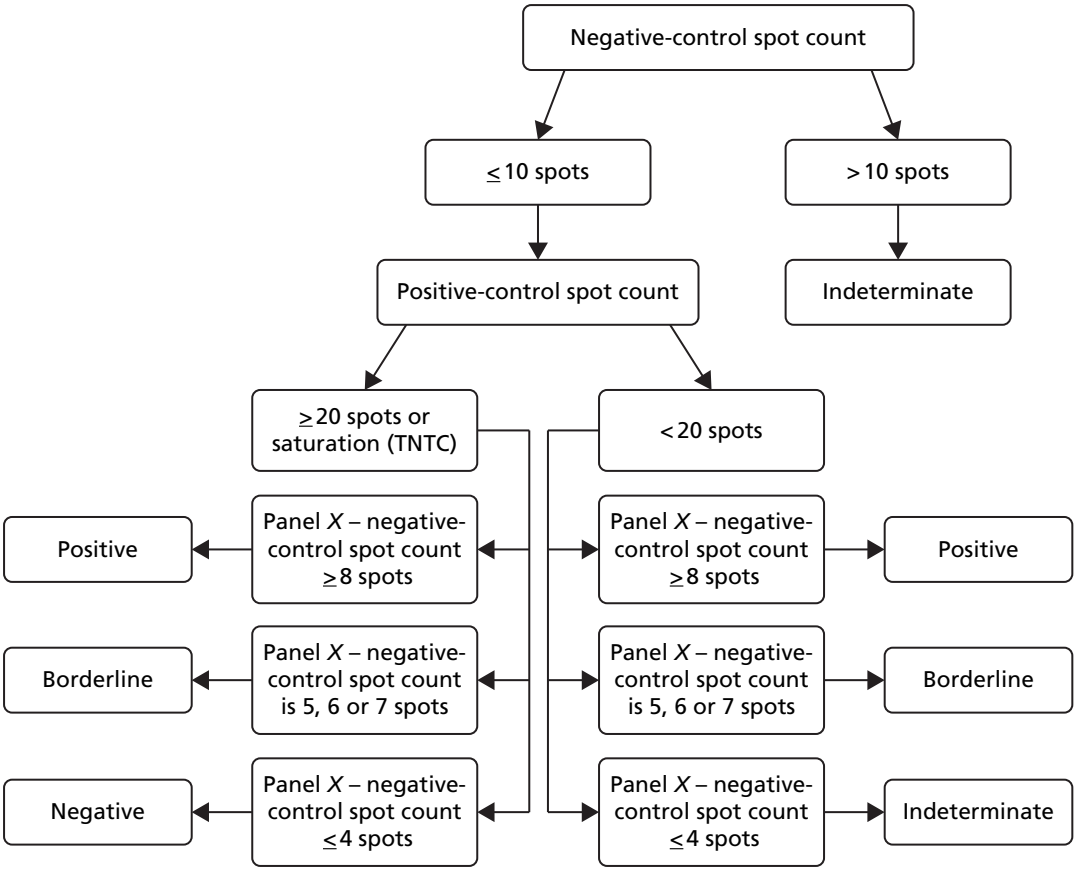

FIGURE 2 Algorithm for defining test positivity of antigens. Panel *X* indicates one of the six panels, A to F, which correspond to the following antigens: A, ESAT-6; B, CFP-10; C, Rv3615c; D, Rv2654; E, Rv3879c; and F, Rv3873. TNTC, too numerous to count.

Grant Instruments Ltd, Shepreth, UK) for 1 minute to ensure that any IFN-γ was evenly distributed throughout the sample. Secreted IFN-γ in the plasma will be sandwiched between the two antibodies.

After incubation and thorough washing with detergent, a photosensitive chromogen substrate solution (3,3′,5,5′-tetramethylbenzidine; included in QFT-GIT kit) was added, which converted the sample to a detectable form (blue colour signal). The reaction was stopped with a substrate stopping solution (sulphuric acid; included in QFT-GIT kit). The intensity of the colour is directly proportional to the levels of IFN-γ present in the plasma after activation of TB-specific T cells by Mtb antigens. Colour should develop in the positive control well and not in the negative control well. Colour in the TB antigen well indicates infection.

The optical density of each well was measured using a microplate reader (Elx800 Absorbance Reader; BioTek, Carnegie, VIC, Australia) with a 450-nm filter and a 620- to 650-nm reference filter. The concentration (IU/ml) of IFN-γ for the plasma sample from each of the three tubes (negative, mitogen and Mtb antigen) was determined against a series of standard concentrations (the standard curve). The test result (negative, positive or indeterminate) was calculated from the concentration values using a US Food and Drug Administration-approved algorithm (*Figure 3*) run on QuantiFERON-TB Gold In-Tube Analysis Software version 2.62 (Cellestis) in accordance with the manufacturer's instructions.[10]

Tuberculin skin test

A TST was performed as part of routine clinical care. Each recruiting centre has its own policy for TST use based on National Institute for Health and Care Excellence (NICE) guidance.[11] Patients eligible for the TST, as defined by local or NICE guidance, received a single intradermal injection of two tuberculin units or 0.1 ml of unlicensed tuberculin Mantoux test [Tuberculin PPD RT23 SSI (Statens Serum Institut, Copenhagen, Denmark); this purified protein derivative (PPD) was used for the main site; however, other sites may have

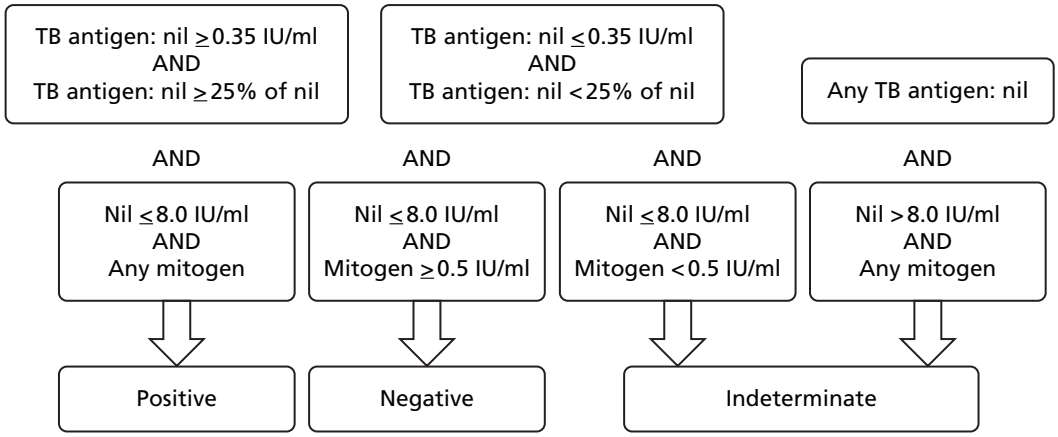

FIGURE 3 Determination of the test results for QFT-GIT. Reprinted from *Methods*, Vol. 61, Whitworth HS, Scott M, Connell DW, Dongés B, Lalvani A. IGRAs – the gateway to T cell based TB diagnosis, pp. 52–62. © 2013, with permission from Elsevier.[6]

used other PPDs]. The test was administered by trained specialist TB nurses who were entitled to administer medicines within the trust under a Patient Group Directive. The degree of skin induration was measured 48–96 hours later. Test results were obtained from centres and we determined test positivity using three thresholds: ≥ 5 mm, ≥ 10 mm[3,12] and a stratified threshold based on BCG vaccination status (≥ 6 mm for unvaccinated and ≥ 15 mm for vaccinated participants).[11] Patients were considered BCG vaccinated if they reported they had been vaccinated and/or had a BCG scar.

Reference standard

Participating hospitals followed the minimum set of tests defined within the NICE guideline[11] for diagnosing active TB and in accordance with local routine practice. However, the *final* diagnosis of participants was verified using the composite reference standard defined in *Appendix 2*. The reference standard was applied by a panel of clinicians blinded to local (routine) and study IGRA results. The role of the clinical panel was to assess anonymised patient clinical data without knowledge of the IGRA results in order to confirm the diagnostic category of all study participants. The composition of the panel and the assessment process is described in *Composition of the clinical panel* and *Assessment of final diagnosis*.

Composition of the clinical panel

Clinical panel members were appointed from among the principal investigators (PIs) and co-investigators by the chief investigators. The panel included:

- an independent chairperson
- a chest physician
- a HIV physician with knowledge of HIV/TB infection
- an infectious disease physician.

All physicians had extensive TB expertise.

Structure of case panel meetings

Meetings were arranged and minutes were recorded by the study co-ordinator or a designated member of the study team. The panel meetings were held quarterly and the schedule was decided by the SMG depending on the number of data available for review. At each meeting, priority was given to assessment of indeterminate (category 3) cases. Additional meetings were arranged to ensure that all patients in categories 2 and 3, and some in category 4, were reviewed (see *Assessment of final diagnosis*). All panel members had to attend the meeting in person.

Assessment of final diagnosis

Diagnosis data (based on the Dosanjh categorisation[3] outlined in *Appendix 2*) received from recruiting centres were reviewed as follows:

- Category 1: all culture-confirmed cases were not reviewed, but were signed off at the end of the clinical panel meeting by the chief or co-investigators.
- Category 2: all probable cases were reviewed.
- Category 3: all indeterminate cases were reviewed.
- Category 4: non-active TB cases with a confirmed alternative diagnosis were not reviewed by the panel. Complicated category 4 patients were reviewed by the panel.

Reviewing cases in this way ensured consistent final diagnosis categorisation.

Data available to the panel

For each patient who was reviewed, the following information was presented to the panel:

- patient demographics
- TB symptoms, previous TB information, TB exposure history, current medication, patient medical history, follow-up data, HIV infection status and relevant clinical information and travel data
- relevant clinical correspondence and test results during diagnosis and follow-up (excluding results of routine and study IGRAs) such as culture, smear, PCR, TST, bronchoscopy, biopsy and/or radiological reports.

Documentation

Each panel member reviewed a patient's documents and completed a form with the following information:

1. diagnosis category (based on the Dosanjh criteria)
2. body site of disease (only if final diagnosis was active TB)
3. method of diagnosis: included culture, PCR, imaging, smear microscopy, histology, clinical features, response to treatment, multiple and other. For multiple and other, details were to be specified.

Confirmation of diagnosis

Final diagnosis decisions were made by a majority vote, with the chairperson having a casting vote, if necessary. Final diagnoses were recorded by the study co-ordinator. If necessary, when a panel member had treated a patient being reviewed, the member was asked to provide information (without disclosing local or study IGRA results) and the member was excluded from final decision-making (their vote was replaced with a vote from the chairperson) for the patient.

Outcomes

Sensitivity, specificity, PPVs and NPVs and likelihood ratios for each test and combinations of tests were calculated to determine their diagnostic accuracy and clinical utility. The likely primary clinical utility of immune-based testing is to exclude TB. Thus, when interpreting the analyses and drawing conclusions, the focus was primarily on the sensitivities and NPVs. For test comparisons, relative test performance was assessed by comparing the sensitivity and specificity of one test with those of another test. These results were presented as relative sensitivities and relative specificities.

Statistical analyses

Sample size calculation

As stated in *Outcomes*, the primary clinical utility of IGRA results in the assessment of suspected active TB is likely to be in their NPV, which may enable clinicians to reliably rule out TB from the differential diagnoses.

This, in turn, depends on the sensitivity of the test and the prevalence of active TB in the tested population. In a meta-analysis, the average sensitivity of T-SPOT.*TB* and QFT-GIT was 90% [95% confidence interval (CI) 86% to 93%] and 70% (95% CI 63% to 78%), respectively.[7] However, the estimates were mainly based on small studies and most studies included only patients without HIV infection. Furthermore, the estimates were not based on head-to-head comparative accuracy studies. Two large studies ($n = 194$ active TB cases diagnosed from $n = 389$ TB suspects;[3] $n = 216$ active TB cases diagnosed from $n = 413$ TB suspects[13]) gave more robust estimates for T-SPOT.*TB* of 85.1% (95% CI 79.2% to 89.9%) and 85.2% (95% CI 76.1% to 91.9%), respectively. The latter study compared T-SPOT.*TB* and QFT-GIT, and gave an estimate of 78.1% (95% CI 70.7% to 84.3%) for QFT-GIT.[13]

Given the available evidence, we powered the IDEA study to detect a conservatively estimated 10% difference in sensitivity between T-SPOT.*TB* and QFT-GIT, assuming a sensitivity of 85% for T-SPOT.*TB* and of 75% for QFT-GIT. To detect this difference at the 5% significance level (two-tailed) with 90% power, 855 patients were required (each receiving both tests), assuming a 40% prevalence of active TB in the study population. This calculation was done using a method that accounts for the paired nature of the data (based on McNemar's test).[14] The method requires knowledge of the probability of positive T-SPOT.*TB* and positive QFT-GIT results among cases of active TB (concordance probability). A positive correlation, such as may be expected between both blood tests, would give a lower sample size than assuming independence or a negative correlation of test errors. However, as no pilot data were available to inform the choice of the concordance probability, we chose to be conservative and so assumed independence. To allow for missing data, indeterminate index test and reference standard results, withdrawal of consent and possible logistical errors, we aimed to recruit 1012 participants.

According to published evidence, the sensitivity of QFT-GIT decreases in HIV-positive subgroups, whereas that of T-SPOT.*TB* is unaffected in some studies and decreased in others.[15] Therefore, we computed sample size for the HIV-positive subgroup based on sensitivities of 85% and 65% for T-SPOT.*TB* and QFT-GIT, respectively. Assuming a 50% prevalence of active TB among these participants, we thus required 156 patients to detect a 20% difference between the IGRAs at the 5% significance level with 80% power. We aimed to recruit 200 patients for similar reasons to those outlined above.

Revision of sample size calculation and study extension for recruitment of HIV-positive participants

During the study recruitment period, the proportion of HIV-positive patients with a final diagnosis of active TB was found to be substantially lower (20% rather than 50%) than originally anticipated when the study was designed. This was attributed to a decrease in TB incidence in this population in recent years.[16] In order to answer a key objective regarding the utility of IGRAs in HIV-positive patients, the SSC and funder supported an extension of recruitment of HIV-positive participants to ensure that the study was adequately powered.

The sample size calculation was revised to take into account the reduced prevalence of active TB in this population. Given a prevalence of 20%, 390 HIV-positive participants will be required to detect a 20% difference between the sensitivity of T-SPOT.*TB* and QFT-GIT with 80% power at the 5% significance level. Thus, the sample size was increased from 156 to 390. Ethics approval was sought and an extension of 12 months was granted on 5 November 2014 to recruit and follow up only HIV-positive participants. Given the purposive recruitment of additional HIV-positive participants, the results are presented separately for the main cohort (including HIV-positive participants recruited during the first phase of the IDEA study prior to the extension period) in *Chapters 3* and *4*, and for the entire HIV-positive cohort in *Chapter 5*.

Data analysis

Sensitivity, specificity, PPV, NPV and likelihood ratios for each test and combination of tests were calculated to determine their diagnostic accuracy and clinical utility in the main cohort and within the key subgroups outlined in the study objectives. For all proportions, 95% CIs were calculated using the Wilson method.[17,18] CIs for positive and negative likelihood ratios were calculated using the method by Simel *et al.*[19] Separate analyses of the complete cohort of HIV-positive participants were also performed.

Patients classified as having culture-confirmed (category 1) or highly probable (category 2) active TB and those without active TB (category 4) were included in analyses of diagnostic accuracy (see definition of categories in *Appendix 2*). However, despite not being included in the analysis, the proportion of patients classified as clinically indeterminate (category 3) was reported. While borderline T-SPOT.*TB* results were included as test positives in primary analyses of the main study cohort, we examined the impact of this by excluding borderline test results in sensitivity analyses. For the primary analyses, patients with indeterminate IGRA results were excluded from the analyses. In clinical practice, if an IGRA was used as a rule-out test for active TB, then an indeterminate result would have the same implications as a positive result, that is, it could not rule out a diagnosis of the disease (and thus TB would remain a differential diagnosis). The impact of including indeterminate IGRA results as test positives (i.e. true positives and false positives depending on final diagnosis of active TB or no active TB) was assessed in sensitivity analyses. Sensitivity analyses were also conducted to investigate the impact of excluding category 2 patients on the sensitivity of IGRAs.

Comparisons between different IGRAs were performed using generalised estimating equation (GEE) models to exploit the paired nature of the data. Our analysis was based on all available data (i.e. it included patients who did not have a complete set of index test results). Separate GEE models were fitted for those with active TB (Dosanjh categories 1 and 2) and those with no active TB (category 4) to determine differences in sensitivity and specificity, respectively. The outcome variable in the GEE model was IGRA result (positive vs. negative) and the explanatory variable was type of IGRA, for example T-SPOT.*TB* versus QFT-GIT. The natural outputs from these models are odds ratios. For example, comparing the sensitivity of T-SPOT.*TB* to that of QFT-GIT, the odds ratio is the odds of a positive T-SPOT.*TB* result compared with the odds of a positive QFT-GIT result in patients with active TB. For the comparison of specificities, the odds ratio is the odds of a negative T-SPOT.*TB* result compared with the odds of a negative QFT-GIT result in non-TB patients. As odds ratios do not have an intuitive interpretation, we computed ratios of sensitivities (relative sensitivity) and ratios of specificities (relative specificity) using a function (*nlcom*) that computes point estimates and CIs for non-linear combinations of parameter estimates post estimation of the models. The CIs are computed using the delta method.

Variation in the relative performance of T-SPOT.*TB* and QFT-GIT with HIV infection status and other clinical characteristics was investigated by including one covariate at a time in the GEE models.[20] To assess the effect of a covariate on relative test performance, an interaction term for test type and the covariate was included in the model. In addition to these characteristics pre-specified in the protocol, we also investigated the effect of smoking, because this has been associated with a twofold increase in the risk of developing active TB.[21] We included both inpatients and outpatients in all our analyses of diagnostic accuracy to allow for generalisability of these tests as an initial test in any case of possible active TB. However, because disease severity and spectrum can influence the diagnostic performance of a test, we also investigated the effect of clinical setting (inpatients vs. outpatients) to determine if our approach was tenable. We were interested in exploring the effect of vitamin D, as it is an important cofactor for the intracellular killing of TB. It is associated with an increased resistance to TB infection, and with the phenotype of active TB. The value of vitamin D supplementation in active TB to improve disease outcome is unclear. However, we were unable to evaluate the effect of vitamin D status on IGRA performance because of variation between centres in the definition of vitamin D status.

In the subset of patients presenting with a TST result as part of their clinical diagnostic workup, the performance of the TST used in sequence with IGRAs was evaluated. The performance of this combination was assessed using logistic regression models constructed in the same form as Bayesian updating (post-test odds = pre-test odds × likelihood ratio) by including the log of the pre-test odds of prevalence (a constant term of known value) as an offset in the model. A linear predictor was then used to estimate log-likelihood ratios, rather than log-odds ratios, and bootstrap methods were used to obtain valid CIs.[22] Model parameterisations from Knottnerus[23] were used to compute likelihood ratios for the additional diagnostic value of each test in a testing sequence. Non-parametric, bias-adjusted CIs for parameter estimates from 1000 bootstrap samples were computed.

We performed all analyses using Stata®, version 13.0 (StataCorp LP, College Station, TX, USA).

Patient and public involvement

Ms Nisha Karnani was our patient and public representative for the duration of the study. She was consulted at key points during the study and was invited to SSC meetings and the IDEA study presentation at the end of the study.

Study oversight and management arrangements

Study Management Group

The SMG included the chief investigators, study co-ordinator, lead research nurse and a post-doctoral research associate. The day-to-day management of the study was carried out by the study co-ordinator, with close support provided by the chief investigators and other members of the SMG. The SMG met monthly to discuss study progress and oversight.

Data Management Group

The Data Management Group (DMG) consisted of members of the statistical team and members of the SMG. The group met regularly to review data on recruitment and the prevalence of TB and HIV in the study cohort. The DMG reported to the SSC (see *Study Steering Committee*).

Study Steering Committee

Independent oversight was provided by the SSC. The committee included an independent chairperson (Professor Khalid Khan), three other independent members (Dr Stephen Gordon, Dr James Grey and Dr Johannes B Reitsma) and a patient and public involvement (PPI) representative (Ms Nisha Karnani).

Ethics arrangements and regulatory approvals

Ethics approval for this study

This study received ethics approval from the London – Camden Kings Cross Research Ethics Committee (REC) (reference number 11/H0722/8). The research study was submitted for site-specific assessment at each participating NHS trust. The chief investigators required a copy of the research and development (R&D) approval letter before accepting participants into the study. The study was conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Declaration of Helsinki (1964) and later revisions.[24]

Consent and study withdrawal

Consent to enter the study was sought from each participant only after a full explanation had been given, an information sheet offered and time allowed for consideration. Signed participant consent was then obtained. The right of the participant to refuse to participate without giving reasons was respected. After a participant was entered into the study the clinician remained free to give any treatment that he or she considered necessary, or to refer onto an appropriate health-care professional, at any stage if it was judged to be in the best interest of the participant. Reasons for such decisions were recorded. In such cases, participants remained in the study for the purposes of follow-up and data analysis. All participants were free to withdraw at any time from the study without giving reasons. Assurance was provided by the person taking consent that withdrawal will not affect the patient's care. In accordance with good clinical practice guidance, participants who withdrew were not required to give a reason for withdrawal. Data were collected on participant's final diagnosis unless consent was withdrawn for any data to be used.

Confidentiality

The chief investigator and all members of the research team abided by the Data Protection Act[25] and preserved the confidentiality of participants involved in the study. Participants were allocated a unique identifying code (anonymised) on recruitment, with no personal identifiers recorded on any sample or data.

Indemnity

The Imperial College London, as sponsor of this study, holds negligent and non-negligent harm insurance policies that applied to this study. These were arranged through the Joint Research Office.

Protocol amendments

Between April 2011 and February 2015, the protocol underwent seven amendments as detailed in *Appendix 3*, *Table 53*. Six of the seven amendments were deemed substantial amendments that required ethics approval, whereas one was a minor amendment.

Chapter 3 Participant characteristics

The results presented in this chapter are based on all participants recruited prior to the study extension, including those with HIV infection. Thus, the chapter excludes HIV-positive patients recruited during the study extension period.

Recruitment of participants into main study cohort

A total of 1074 participants, including 177 (16.5%) who were HIV positive, were recruited from 10 NHS trusts into the main study between 25 November 2011 and 31 August 2013. The number of patients recruited at each centre is shown in *Table 1*. Over half of the study participants were recruited from two trusts: Imperial College Healthcare NHS Trust (26.4%) and London North West Healthcare NHS Trust (27.8%).

The flow of patients through the study is shown in *Figure 4*. Of the 1074 patients recruited, 845 were included in the analyses. Reasons for exclusion are shown in *Figure 4*. Forty patients from Frimley Health NHS Foundation Trust were excluded because patients all had diagnoses of confirmed or highly probable TB (categories 1 and 2) due to an error of implementation of recruitment criteria at this site (i.e. the natural spectrum of patients with suspected TB was not being recruited). The decision to exclude these patients was approved by the SSC following a SSC meeting during which the diagnosis of patients recruited at each centre was reviewed.

Table 2 shows the number of patients assigned to each diagnostic category. There were 43 (5.1%) patients with a clinically indeterminate (category 3) diagnosis. Of the remaining 802 patients, there were 363 (45.3%) cases of active TB [based on those with culture-confirmed (category 1) and highly probable (category 2) TB] and 439 (54.7%) in whom active TB was excluded (categories 4A to 4D). Of the 439 non-active TB cases, 117 (26.7%) were category 4D.

Baseline characteristics of participants

The main demographic characteristics of the 845 patients are given in *Table 3*. Most patients (64.4%) were recruited from an outpatient setting, and the remaining were recruited from an inpatient setting. The median age of patients was 38 (range 16–86) years and most (59.3%) of the patients were male.

TABLE 1 Recruitment by centre

Hospital trust	Patients recruited, *n* (%)
Imperial College Healthcare NHS Trust	283 (26.4)
Heart of England NHS Foundation Trust	106 (9.9)
Chelsea and Westminster Hospital NHS Foundation Trust	52 (4.8)
Royal Free London NHS Foundation Trust	59 (5.5)
St George's Healthcare NHS Trust	61 (5.7)
Frimley Health NHS Foundation Trust	44 (4.1)
University Hospitals of Leicester NHS Trust	116 (10.8)
London North West Healthcare NHS Trust	299 (27.8)
Oxford University Hospitals NHS Trust	4 (0.4)
Sandwell and West Birmingham Hospitals NHS Trust	50 (4.7)
Total	1074 (100)

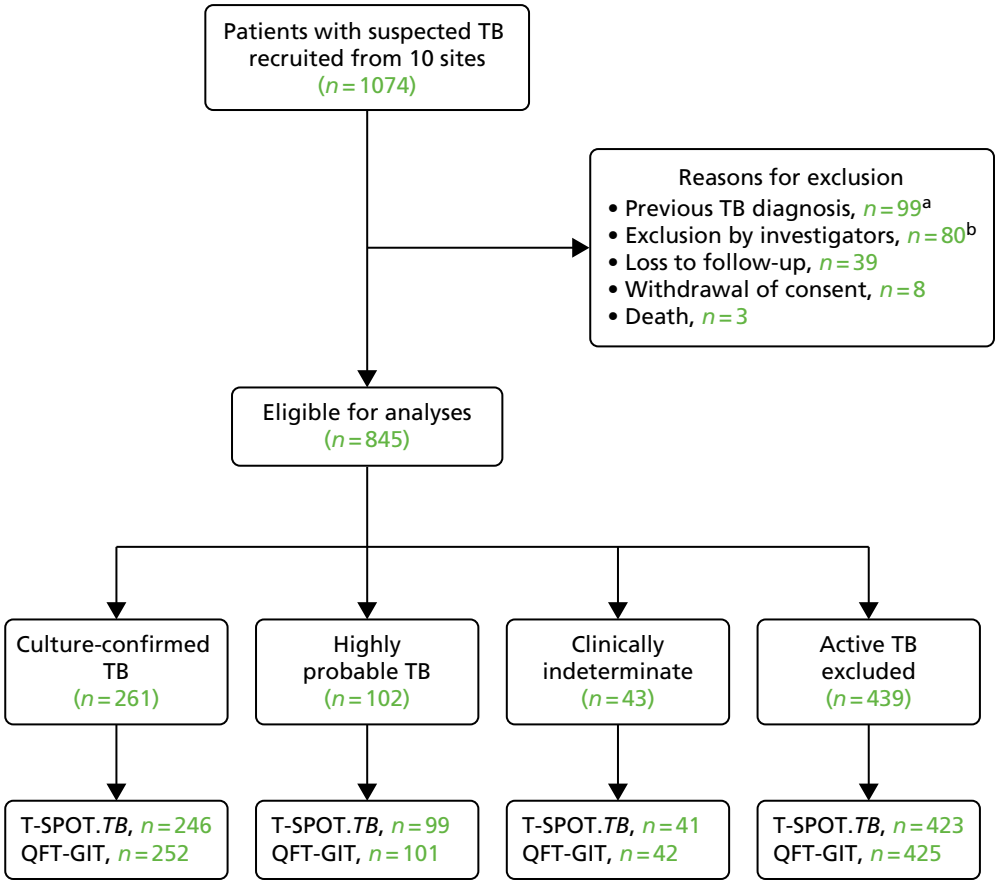

FIGURE 4 Study flow diagram of patients with suspected active TB. The final four boxes show the number of patients with available IGRA results. a, Patients with previously diagnosed TB were excluded from analyses because IGRA results cannot be reliably interpreted in previously treated patients. The decision to exclude was taken by the expert diagnostic panel and study management group in consultation with the independent Study Steering Committee before unblinding of IGRA and next-generation IGRA results. b, On advice from the Study Steering Committee, and following consultation between the study management group and data management groups, 80 patients were excluded from analyses. Patients recruited from Frimley Health NHS Foundation Trust ($n = 40$) were excluded because they all had diagnoses of confirmed or highly probable TB (categories 1 and 2) due to an error of implementation of the recruitment criteria at this site, i.e. the natural spectrum of patients with *suspected* TB was not being recruited. A further subset of patients ($n = 27$) were, on review, considered by the expert diagnostic panel to be ineligible (before unblinding IGRA results) on the basis that they were being investigated for TB (due to an incidental abnormal chest X-ray, known contact with active TB, or screening for anti-TNF treatment), but did not present with symptoms or signs suggestive of TB. An additional 13 patients were excluded due to invalid consent forms. Republished with permission of Elsevier Science and Technology Journals, from Clinical utility of existing and second-generation interferon-γ release assays for diagnostic evaluation of tuberculosis: an observational cohort study, *The Lancet Infectious Diseases*, Whitworth HS, Badhan A, Boakye AA, Takwoingi Y, Rees-Roberts M, Partlett C, *et al.*, vol. 19, pp. 193–202, copyright 2019;[27] permission conveyed through Copyright Clearance Center, Inc.

Almost half (48.2%) of the study population were of Indian origin. Altogether, 75 countries of birth were represented in the study; 16 of the countries had at least 10 participants, with India (26.7%) and the UK (22.8%) accounting for almost half of the study population. The complete list of countries is given in *Appendix 4*, *Table 54*.

Table 4 shows the clinical characteristics of the patients. Of the 845 patients, 135 (16.0%) were HIV positive. Out of the 135 patients, there were two (1.5%) with a clinically indeterminate diagnosis. Of the remaining 133 patients, 25 (18.8%) were active TB cases and 108 (81.2%) were non-active TB cases.

Of the 845 patients, over half had other comorbidities: 300 (35.5%) patients had a single comorbidity, 127 (15.0%) had multiple comorbidities and the remaining 418 (49.5%) had none. There were 88

TABLE 2 Reference standard results according to Dosanjh categories

Diagnostic category	Criteria	Number of patients
1: Culture-confirmed TB	Microbiological culture of Mtb AND suggestive clinical and radiological findings	261
2: Highly probable TB	Clinical and radiological features highly suggestive of TB and unlikely to be caused by other disease AND a decision to treat made by a clinician AND appropriate response to therapy AND histology supportive (if available)	102
3: Clinically indeterminate	Final diagnosis of TB neither highly probable nor reliably excluded	43
4: Active TB excluded		
4A: inactive TB	Stable CXR changes AND TST positive[a] (if done) AND bacteriologically negative (if done) AND no clinical evidence of active disease	7
4B: one or more risk factors for TB exposure,[b] TST positive[a]	TST positive[a] AND bacteriologically negative (if done) AND no clinical evidence of active disease	48
4C: one or more risk factors for TB exposure,[b] TST negative	History of TB exposure AND TST negative (if done)	267
4D: no risk factors for TB exposure,[b] TST negative	No history of TB exposure AND TST negative (if done)	117
Total		845

CXR, chest radiography.
a A TST using the Mantoux test with a threshold of ≥ 15 mm considered positive.
b Risk factors for TB exposure: recent exposure to active TB patient, born in country of high prevalence or belonging to an ethnic group with a high prevalence of TB (incidence > 100/100,000[26]).
Note
Diagnostic categories adapted from Dosanjh *et al.*[3]

TABLE 3 Demographics

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Clinical setting, *n* (%)					
Inpatient	90 (34.5)	30 (29.4)	11 (25.6)	170 (38.7)	301 (35.6)
Outpatient	171 (65.5)	72 (70.6)	32 (74.4)	269 (61.3)	544 (64.4)
Age (years), median (range)	32 (16–81)	36 (18–76)	38 (16–79)	44 (17–86)	38 (16–86)
Male, *n* (%)	177 (67.8)	53 (52.0)	21 (48.8)	250 (56.9)	501 (59.3)
Ethnic origin, *n* (%)					
Asian	16 (6.1)	6 (5.9)	5 (11.6)	14 (3.2)	41 (4.9)
Black	50 (19.2)	22 (21.6)	10 (23.3)	102 (23.2)	184 (21.8)
Hispanic	1 (0.4)	0 (0.0)	0 (0.0)	7 (1.6)	8 (0.9)
Indian subcontinent	167 (64.0)	61 (59.8)	16 (37.2)	168 (38.3)	412 (48.8)
Middle Eastern	4 (1.5)	0 (0.0)	0 (0.0)	12 (2.7)	16 (1.9)
Mixed	1 (0.4)	4 (3.9)	0 (0.0)	8 (1.8)	13 (1.5)
White	22 (8.4)	9 (8.8)	12 (27.9)	126 (28.7)	169 (20.0)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.5)	2 (0.2)

continued

TABLE 3 Demographics (*continued*)

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Years in the UK, median (range)	4.9 (0.1–52.9)	6.1 (0.3–59.7)	10.5 (0.4–56.9)	13.2 (0.0–60.3)	8.3 (0.0–60.3)
Profession, *n* (%)[a]					
Paid employment	130 (49.8)	52 (51.0)	21 (48.8)	214 (48.7)	417 (49.3)
Unpaid employment	62 (23.8)	24 (23.5)	16 (37.2)	164 (37.4)	266 (31.5)
Student	50 (19.2)	13 (12.7)	3 (7.0)	26 (5.9)	92 (10.9)
Health-care/laboratory worker	16 (6.1)	9 (8.8)	2 (4.7)	24 (5.5)	51 (6.0)
Social/prison worker	1 (0.4)	1 (1.0)	0 (0.0)	2 (0.5)	4 (0.5)
Sex worker	0 (0.0)	1 (1.0)	0 (0.0)	2 (0.5)	3 (0.4)
Unknown	2 (0.8)	2 (2.0)	1 (2.3)	7 (1.6)	12 (1.4)

a Some patients had more than one profession.

Note

Republished with permission of Elsevier Science and Technology Journals, from Clinical utility of existing and second-generation interferon-γ release assays for diagnostic evaluation of tuberculosis: an observational cohort study, *The Lancet Infectious Diseases*, Whitworth HS, Badhan A, Boakye AA, Takwoingi Y, Rees-Roberts M, Partlett C, *et al.*, vol. 19, pp. 193–202, copyright 2019;[27] permission conveyed through Copyright Clearance Center, Inc.

TABLE 4 Clinical characteristics

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Height (m), median (range)	1.7 (1.4–2.0)	1.7 (1.5–1.9)	1.6 (1.5–1.8)	1.7 (1.3–2.0)	1.7 (1.3–2.0)
Weight (kg), median (range)	63 (35–127)	64 (40–116)	71 (37–110)	68 (38–157)	65 (35–157)
BMI (kg/m^2), median (range)	22 (14–48)	22 (16–42)	24 (13–45)	24 (15–47)	23 (13–48)
BCG vaccinated, *n* (%)	194 (74.3)	79 (77.5)	36 (83.7)	340 (77.4)	649 (76.8)
BCG scar visible, *n* (%)					
Yes	172 (65.9)	72 (70.6)	29 (67.4)	283 (64.5)	556 (65.8)
No	12 (4.6)	3 (2.9)	3 (7.0)	19 (4.3)	37 (4.4)
Unsure	16 (6.1)	8 (7.8)	6 (14.0)	44 (10.0)	74 (8.8)
Missing	61 (23.4)	19 (18.6)	5 (11.6)	93 (21.2)	178 (21.1)
Known TB contact, *n* (%)	70 (26.8)	25 (24.5)	12 (27.9)	83 (18.9)	190 (22.5)
HIV positive, *n* (%)	13 (5.0)	12 (11.8)	2 (4.7)	108 (24.6)	135 (16.0)
Other pre-existing conditions/comorbidities, *n* (%)[a]					
None	169 (64.8)	61 (59.8)	19 (44.2)	169 (38.5)	418 (49.5)
Diabetes	22 (8.4)	5 (4.9)	8 (18.6)	53 (12.1)	88 (10.4)
Hepatitis B	5 (1.9)	1 (1.0)	0 (0.0)	5 (1.1)	11 (1.3)
Hepatitis C	1 (0.4)	1 (1.0)	0 (0.0)	10 (2.3)	12 (1.4)

TABLE 4 Clinical characteristics (*continued*)

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Chronic/end-stage renal failure	5 (1.9)	1 (1.0)	2 (4.7)	4 (0.9)	12 (1.4)
Cancer	1 (0.4)	1 (1.0)	0 (0.0)	12 (2.7)	14 (1.7)
Organ transplantation	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.5)	2 (0.2)
Asthma	12 (4.6)	5 (4.9)	4 (9.3)	50 (11.4)	71 (8.4)
Sarcoidosis	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Other	74 (28.4)	37 (36.3)	20 (46.5)	228 (51.9)	359 (42.5)
Vitamin D deficiency,[b] *n* (%)					
Deficient	106 (40.6)	35 (34.3)	12 (27.9)	59 (13.4)	212 (25.1)
Insufficient	49 (18.8)	14 (13.7)	7 (16.3)	65 (14.8)	135 (16.0)
Normal	13 (5.0)	8 (7.8)	5 (11.6)	34 (7.7)	60 (7.1)
Not known	93 (35.6)	45 (44.1)	19 (44.2)	281 (64.0)	438 (51.8)

BMI, body mass index.
a Some patients had multiple comorbidities and so the numbers do not add up to 845.
b For the definition of the thresholds used for categorising vitamin D status, see *Appendix 5*.

TABLE 5 Medication history

Medication	Dosanjh category, *n* (%)[a]				Total, *n* (%)
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
None	63 (24.1)	35 (34.3)	13 (30.2)	203 (46.2)	314 (37.2)
Chemotherapy	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)	1 (0.1)
Corticosteroids ≥ 15 mg/day	20 (7.7)	5 (4.9)	5 (11.6)	20 (4.6)	50 (5.9)
Corticosteroids < 15 mg/day	13 (5.0)	7 (6.9)	1 (2.3)	19 (4.3)	40 (4.7)
Corticosteroids unknown	1 (0.4)	1 (1.0)	0 (0.0)	0 (0.0)	2 (0.2)
Ciclosporin, tacrolimus or everolimus	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)	1 (0.1)
Other immune suppressants	0 (0.0)	0 (0.0)	0 (0.0)	5 (1.1)	5 (0.6)
Methotrexate	1 (0.4)	0 (0.0)	0 (0.0)	5 (1.1)	6 (0.7)
Other	191 (73.2)	64 (62.7)	30 (69.8)	233 (53.1)	518 (61.3)
Unknown	1 (0.4)	0 (0.0)	0 (0.0)	1 (0.2)	2 (0.2)

a Some patients had multiple medications and so the numbers do not add up to 845.

(10.4%) patients with pre-existing diabetes mellitus, 12 (1.4%) patients with chronic/end-stage renal failure and 105 (12.4%) patients were on immunosuppressive therapy (*Table 5*). These were the three key subgroups that we had planned to investigate in the subgroup analyses. The thresholds used to categorise vitamin D status varied between hospital trusts, as detailed in *Appendix 5*, *Table 55*. Although vitamin D measurements were missing for a large number of patients (49.1%), when the results were available, many patients were categorised as either vitamin D deficient (26.5%) or insufficient (16.9%), with few (7.5%) having normal results (see *Table 4*).

The social history of participants is outlined in *Table 6*. Smoking history was missing for two patients; about two-thirds of the patients had never smoked, and the remaining patients were current or ex-smokers. Most (58.6%) patients also had no history of alcohol use. Almost all patients (97.6%) had no history of homelessness and a few patients (27/845, 3.2%) had a history of imprisonment.

TABLE 6 Social history

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Smoking history, *n* (%)					
Never smoked	181 (69.3)	81 (79.4)	26 (60.5)	248 (56.5)	536 (63.4)
Ex-smoker	31 (11.9)	8 (7.8)	6 (14.0)	93 (21.2)	138 (16.3)
Current smoker	49 (18.8)	13 (12.7)	11 (25.6)	96 (21.9)	169 (20.0)
Unknown	0	0	0	2 (0.5)	2 (0.2)
Pack years if current smoker, *n* (%)					
≤ 10	11 (22.4)	6 (46.2)	4 (36.4)	27 (28.1)	48 (28.4)
11–20	1 (2.0)	2 (15.4)	1 (9.1)	6 (6.3)	10 (5.9)
21–50	1 (2.0)	1 (7.7)	0	7 (7.3)	9 (5.3)
> 51	1 (2.0)	0	0	1 (1.0)	2 (1.2)
Unknown	35 (71.4)	4 (30.8)	6 (54.5)	55 (57.3)	100 (59.2)
History of alcohol use, *n* (%)					
Non-drinker	163 (62.5)	80 (78.4)	27 (62.8)	225 (51.3)	495 (58.6)
Ex-drinker	10 (3.8)	2 (2.0)	1 (2.3)	35 (8.0)	48 (5.7)
Current drinker	88 (33.7)	20 (19.6)	15 (34.9)	175 (39.9)	298 (35.3)
Unknown	0	0	0	4 (0.9)	4 (0.5)
Units/week if current drinker, median (range)	4 (0–250)	5 (1–35)	2 (0–140)	5 (0–210)	4 (0–250)
History of alcohol misuse, *n* (%)	9 (3.4)	0	1 (2.3)	20 (4.6)	30 (3.6)
History of recreational drug use, *n* (%)					
Non-user	21 (8.0)	10 (9.8)	1 (2.3)	18 (4.1)	50 (5.9)
Ex-user	2 (0.8)	0	1 (2.3)	5 (1.1)	8 (0.9)
Current user	5 (1.9)	3 (2.9)	1 (2.3)	13 (3.0)	22 (2.6)
Unknown	233 (89.3)	89 (87.3)	40 (93.0)	403 (91.8)	765 (90.5)

TABLE 6 Social history (*continued*)

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
History of homelessness, *n* (%)					
None	256 (98.1)	101 (99.0)	43 (100.0)	425 (96.8)	825 (97.6)
Previously homeless	2 (0.8)	1 (1.0)	0	11 (2.5)	14 (1.7)
Currently homeless	3 (1.1)	0	0	3 (0.7)	6 (0.7)
Years homeless if currently or previously homeless, median (range)	4 (0–6)	12	–	6 (0–24)	6 (0–24)
History of imprisonment, *n* (%)	4 (1.5)	2 (2.0)	0	21 (4.8)	27 (3.2)

Note
The percentages are column percentages for each characteristic.

Table 7 summarises the frequency of presenting symptoms for 827 patients. The main symptoms recorded were cough, fever, night sweats, weight loss, haemoptysis and lethargy. Patients generally presented with multiple symptoms, but a cough was often present (576/827, 69.6%). The median number of symptoms was four (range 1–10).

TABLE 7 Symptoms at presentation

Symptom	Diagnosis as per reference standard[1]				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Cough, *n* (%)	174 (68.0)	53 (53.5)	23 (53.5)	326 (76.0)	576 (69.6)
Fever, *n* (%)	126 (49.2)	49 (49.5)	14 (32.6)	195 (45.5)	384 (46.4)
Night sweats, *n* (%)	129 (50.4)	53 (53.5)	20 (46.5)	215 (50.1)	417 (50.4)
Weight loss, *n* (%)	154 (60.2)	54 (54.5)	21 (48.8)	211 (49.2)	440 (53.2)
Haemoptysis, *n* (%)	31 (12.1)	8 (8.0)	3 (7.0)	65 (15.2)	107 (12.9)
Lethargy, *n* (%)	133 (52.0)	56 (56.6)	23 (53.5)	222 (51.7)	434 (52.5)
Other, *n* (%)	163 (63.7)	59 (59.46)	25 (58.1)	202 (47.1)	449 (54.3)
Number of symptoms, median (range)	4 (1–10)	4 (1–8)	3 (1–7)	3 (1–10)	4 (1–10)

Eighteen participants were recruited on the basis of abnormal clinical signs rather than symptoms. The percentages are column percentages for each symptom.

Note
Republished with permission of Elsevier Science and Technology Journals, from Clinical utility of existing and second-generation interferon-γ release assays for diagnostic evaluation of tuberculosis: an observational cohort study, *The Lancet Infectious Diseases*, Whitworth HS, Badhan A, Boakye AA, Takwoingi Y, Rees-Roberts M, Partlett C, *et al.*, vol. 19, pp. 193–202, copyright 2019;[27] permission conveyed through Copyright Clearance Center, Inc.

Final diagnosis

Table 8 shows the diagnostic tests performed during the diagnostic workup of patients. Chest radiography and culture were often performed (in 89.4% and 86.5% of patients, respectively), but cerebrospinal fluid testing and magnetic resonance imaging (MRI) were uncommon (in 3.6% and 12.0%, respectively). The number of T-SPOT.*TB*, QFT-GIT and TST tests performed as part of routine care at each centre is shown in *Appendix 6, Table 56*. However, for the purpose of the IDEA study, IGRAs were not used in determining the final diagnosis of patients. TSTs were performed in only 336 patients across the nine centres and results were available for 322 patients. Most of these 336 patients were recruited at London North West Healthcare NHS Trust (57%) and Imperial College Healthcare NHS Trust (26%).

The final diagnosis of active TB patients is detailed in *Table 9*. Of the 363 patients with active TB, 237 (65.3%) had smear-negative TB. Forty-five (12.4%) active TB cases had both pulmonary and extrapulmonary TB. Approximately half (189/363, 52.1%) had only extrapulmonary TB. This occurred more often among those diagnosed as having highly probable TB (75/102, 73.5%) relative to culture-confirmed cases (114/261, 43.7%). In 129 (35.5%) active TB cases, patients had only pulmonary TB. In contrast to extrapulmonary TB, pulmonary TB was more common among culture-confirmed cases (110/261, 42.15%) than in highly probable TB cases (19/102, 18.6%). The most common sites of TB infection were the lungs (174/363, 47.9%) and lymph nodes (154/363, 42.4%). The drug sensitivity profile shows that of the 351 culture tests performed, 239 (65.8%) were fully sensitive and 22 (6.3%) were drug resistant.

Table 10 shows the final diagnosis of non-active TB patients. A patient may have multiple conditions. Of the seven conditions listed in the table, pneumonia was the most frequent diagnosis, with 104 of 439 (23.7%) patients having the condition. A higher proportion of inpatients were diagnosed with cancer (14.1%) or pneumonia (38.8%) than outpatients (4.5% and 14.1%, respectively). In contrast, a higher proportion of outpatients were diagnosed with chest infections, latent TB infection and sarcoidosis.

TABLE 8 Diagnostic tests performed during diagnostic workup

Test	Dosanjh category, *n* (%)				Total, *n* (%)
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
BAL investigation	51 (19.5)	20 (19.6)	7 (16.3)	108 (24.6)	186 (22.0)
CXR	231 (88.5)	95 (93.1)	38 (88.4)	391 (89.1)	755 (89.3)
CSF investigation	7 (2.7)	6 (5.9)	3 (7.0)	14 (3.2)	30 (3.6)
CT	142 (54.4)	69 (67.6)	26 (60.5)	273 (62.2)	510 (60.4)
Culture	261 (100)	90 (88.2)	29 (67.4)	351 (80.0)	731 (86.5)
Histology or biopsy	72 (27.6)	42 (41.2)	15 (34.9)	101 (23.0)	230 (27.2)
MRI	29 (11.1)	17 (16.7)	8 (18.6)	47 (10.7)	101 (12.0)
PCR	85 (32.6)	20 (19.6)	6 (14.0)	66 (15.0)	177 (20.9)
Smear test	232 (88.9)	75 (73.5)	24 (55.8)	335 (76.3)	666 (78.8)

CSF, cerebrospinal fluid; CXR, chest radiography.

TABLE 9 Final diagnosis of patients with active TB

Characteristic	Category of TB, *n* (%)		Total, *n* (%)
	1	2	
All TB	261 (71.9)	102 (28.1)	363 (100)
Smear-positive TB	67 (25.7)	3 (2.9)	70 (19.3)
Smear-negative TB	165 (63.2)	72 (70.6)	237 (65.3)
Pulmonary TB	110 (42.1)	19 (18.6)	129 (35.5)
Extrapulmonary TB	114 (43.7)	75 (73.5)	189 (52.1)
Pulmonary and extrapulmonary TB	37 (14.2)	8 (7.8)	45 (12.4)
Site of infection[a]			
Abdomen	6 (2.3)	3 (2.9)	9 (2.5)
Bones	5 (1.9)	0	5 (1.4)
Brain	2 (0.8)	4 (3.9)	6 (1.7)
Chest wall	1 (0.4)	1 (1.0)	2 (0.6)
Lungs	147 (56.3)	27 (26.5)	174 (47.9)
Lymph node	105 (40.2)	49 (48.0)	154 (42.4)
Miliary TB (disseminated)	11 (4.2)	0	11 (3.0)
Pericardium	4 (1.5)	2 (2.0)	6 (1.7)
Pleura	15 (5.7)	11 (10.8)	26 (7.2)
Spine	10 (3.8)	6 (5.9)	16 (4.4)
Other	15 (5.7)	16 (15.7)	31 (8.5)
Drug sensitivity profile[b]			
Fully sensitive	239 (91.6)	0	239 (68.1)
Drug resistant	21 (8.1)	0	21 (6.0)
MDR	1 (0.4)	0	1 (0.3)
Not tested	0	90 (100)	90 (25.6)

MDR, multidrug resistant.
a Some patients had TB at multiple sites and so percentages do not add up to 100%.
b Available where culture tests were performed during diagnostic work up ($n = 351$).

Note
Republished with permission of Elsevier Science and Technology Journals, from Clinical utility of existing and second-generation interferon-γ release assays for diagnostic evaluation of tuberculosis: an observational cohort study, *The Lancet Infectious Diseases*, Whitworth HS, Badhan A, Boakye AA, Takwoingi Y, Rees-Roberts M, Partlett C, *et al.*, vol. 19, pp. 193–202, copyright 2019;[27] permission conveyed through Copyright Clearance Center, Inc.

TABLE 10 Final diagnosis of patients without active TB

Diagnosis	Non-active TB patients, *n* (%)		
	Inpatients (*n* = 170)	Outpatients (*n* = 269)	Total, *n* (%) (*N* = 439)
Cancer	24 (14.1)	12 (4.5)	36 (8.2)
Chest infection	1 (0.6)	15 (5.6)	16 (3.6)
Lower respiratory tract infection	10 (5.9)	13 (4.8)	23 (5.2)
Pneumonia	66 (38.8)	38 (14.1)	104 (23.7)
Sarcoidosis	5 (2.9)	33 (12.3)	38 (8.7)
Upper respiratory tract infection	0 (0.0)	13 (3.0)	13 (3.0)
Other	70 (41.2)	153 (56.9)	223 (50.8)

Note
Some patients had multiple diagnoses and so percentages represent the number of patients with that particular diagnosis out of the total number.

Chapter 4 Diagnostic accuracy results

Overview

Estimates of the accuracy of IGRAs presented in this chapter are for the main cohort of patients (including HIV-positive patients recruited prior to the extension period). Estimates of the accuracy of T-SPOT.*TB* and QFT-GIT are presented individually, followed by comparisons of test accuracy (T-SPOT.*TB* vs. QFT-GIT). The results of subgroup analyses are then presented for T-SPOT.*TB* and QFT-GIT. Finally, test accuracy estimates are provided for ESAT-6, CFP-10 and second-generation IGRAs (Rv3615c, Rv2654, Rv3879c and Rv3873) individually and in combinations; the diagnostic accuracy of the test combinations were then compared with that of T-SPOT.*TB*. For completeness, we also briefly address the accuracy of the TST.

Completeness of interferon gamma release assay results

Of the 845 patients, T-SPOT.*TB* and QFT-GIT results were available for 809 (96%) and 820 (97%) patients, respectively. All 809 patients with T-SPOT.*TB* results also had results for the second-generation IGRAs (Rv3615c, Rv2654, Rv3879c and Rv3873). *Table 11* shows the results for T-SPOT.*TB* and QFT-GIT by diagnostic category. For 805 patients, results were available for both IGRAs; reasons for missing T-SPOT.*TB* and QFT-GIT results are shown in *Table 12*. The cross-classified results of the two tests are given in *Appendix 7* (see *Table 57*) for the 805 patients.

TABLE 11 Results for T-SPOT.*TB* and QFT-GIT by diagnostic category

	Dosanjh category								
Index test result	1	2	3	4A	4B	4C	4D	4A–D	Total
T-SPOT.*TB*									
Positive	185	68	15	0	18	27	6	51	319
Negative	33	25	23	6	26	200	87	319	400
Borderline	16	1	0	0	1	14	1	16	33
Indeterminate	12	5	3	1	1	18	17	37	57
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
Median SFCs ESAT-6 (range)	13 (0–387)	13 (0–492)	0 (0–325)	0 (0–1)	2 (0–274)	0 (0–210)	0 (0–147)	0 (0–274)	1 (0–492)
Median SFCs CFP-10 (range)	17 (0–465)	13 (0–437)	1 (0–315)	0 (0–3)	1 (0–160)	0 (0–166)	0 (0–148)	0 (0–166)	1 (0–465)
QFT-GIT									
Positive	163	57	14	0	19	49	6	74	308
Negative	68	39	22	7	25	187	85	304	433
Indeterminate	21	5	6	0	2	26	19	47	79
Missing	9	1	1	0	2	5	7	14	25
Total	261	102	43	7	48	267	117	439	845
Median IFN-γ levels (range)	0.69 (0–10)	0.64 (0–10)	0.15 (0–6.69)	0.02 (0–0.34)	0.15 (0–10)	0.03 (0–10)	0.01 (0–7.58)	0.03 (0–10)	0.12 (0–10)

TABLE 12 Reasons for missing IGRA results

Reason	Test, *n*	
	QFT-GIT	T-SPOT.*TB*
No sample could be taken	1	0
Sample destroyed for laboratory reasons	2	11
Sample unsuitable for testing	6	8
Unable to obtain sample from patient	16	17
Total	25	36

Note
The same reason applied to both test results for 20 participants.

Diagnostic accuracy of T-SPOT.*TB* and QuantiFERON GOLD In-Tube

Table 13 shows the cross-tabulation of T-SPOT.*TB* and QFT-GIT results against active TB status (i.e. active TB or not). The proportion of indeterminate test results was 7.0% (57/809) for T-SPOT.*TB* and 9.6% (79/820) for QFT-GIT. The difference between the two proportions was 2.6% (95% CI –0.1% to 5.3%; $p = 0.06$). Based on all culture-confirmed and highly probable active TB cases and excluding indeterminate IGRA results, sensitivity was 82.3% (95% CI 77.8% to 86.1%) for T-SPOT.*TB* and 67.3% (95% CI 62.0% to 72.1%) for QFT-GIT. Among those in whom active TB was excluded, specificity was 82.6% (95% CI 78.6% to 86.1%) for T-SPOT.*TB* and 80.4% (95% CI 76.1% to 84.1%) for QFT-GIT (*Table 14*). The PPVs for T-SPOT.*TB* and QFT-GIT were 80.1% (95% CI 75.5% to 84.0%) and 74.8% (95% CI 69.6% to 79.5%), respectively, and the NPVs were 84.6% (95% CI 80.6% to 87.9%) and 74.0% (95% CI 69.5% to 78.0%), respectively. For T-SPOT.*TB*, 4.1% (33/809) of the test results were borderline (see *Table 13*). When these borderline results were excluded from the T-SPOT.*TB* analysis, sensitivity (95% CI) was 81.4% (76.6% to 85.3%), specificity (95% CI) was 86.2% (82.3% to 89.4%), PPV (95% CI) was 83.2% (78.6% to 87.0%) and NPV (95% CI) was 84.6% (80.6% to 87.9%). The full results of the analysis are shown in *Appendix 7*, *Table 58*.

Using only culture-confirmed active TB cases, the sensitivities of T-SPOT.*TB* and QFT-GIT increased slightly to 85.9% (95% CI 80.9% to 89.8%) and 70.6% (95% CI 64.4% to 76.1%), respectively. The sensitivity of T-SPOT.*TB* was 7.0% lower in patients diagnosed with pulmonary TB than in those with extrapulmonary TB [77.0% (95% CI 68.4% to 83.8%) vs. 84.0% (95% CI 77.9% to 88.7%)]. In contrast, although the difference was small (2.3%), the sensitivity of QFT-GIT was higher in patients who had pulmonary TB (68.4%, 95% CI 59.4% to 76.2%) than in those with extrapulmonary TB (66.1%, 95% CI 58.7% to 72.8%). Using only category 4D patients without active TB gave much higher specificities than using all patients without active TB (see *Table 14*).

In the sensitivity analyses with indeterminate test results included as test positives, the sensitivity of T-SPOT.*TB* was 83.2% (95% CI 78.9% to 86.8%) and 69.7% (95% CI 64.7% to 74.2%) for QFT-GIT. The specificity of T-SPOT.*TB* was 75.4% (95% CI 71.1% to 79.3%) and 71.5% (95% CI 67.1% to 75.6%) for QFT-GIT. Full results are provided in *Appendix 7*, *Table 59*.

Sensitivity, specificity and predictive values are presented as percentages. For TSPOT.*TB*, there were 69 test positives out of 94 highly probable TB cases, with sensitivity (95% CI) of 73.4% (63.7% to 81.3%). For QFT-GIT, there were 57 test positives out of 96 highly probable TB cases, with sensitivity (95% CI) of 59.4% (49.4% to 68.7%). Note that in the primary analyses of T-SPOT.*TB*, borderline test results were included as test positives. Sensitivity analyses were performed with borderline T-SPOT.*TB* results excluded and the results are shown in *Appendix 7*, *Table 58*. Indeterminate IGRA results were excluded from all analyses. See *Appendix 7* for sensitivity analyses using all IGRA results with indeterminates included as test positives.

TABLE 13 Cross-tabulation of T-SPOT.*TB* and QFT-GIT against final diagnosis[27]

		T-SPOT.*TB*, *n*											
		Active TB positive (categories 1 and 2)						Active TB negative (category 4)					
		Positive	Negative	Borderline	Indeterminate	Missing	Total	Positive	Negative	Borderline	Indeterminate	Missing	Total
QFT-GIT	Positive	187	13	6	9	5	220	37	30	3	3	1	74
	Negative	49	41	8	7	2	107	12	250	12	26	4	304
	Indeterminate	16	4	3	1	2	26	2	36	1	8	0	47
	Missing	1	0	0	0	9	10	0	3	0	0	11	14
	Total	253	58	17	17	18	363	51	319	16	37	16	439

Note
Republished with permission of Elsevier Science and Technology Journals, from Clinical utility of existing and second-generation interferon-γ release assays for diagnostic evaluation of tuberculosis: an observational cohort study, *The Lancet Infectious Diseases*, Whitworth HS, Badhan A, Boakye AA, Takwoingi Y, Rees-Roberts M, Partlett C, *et al.*, vol. 19, pp. 193–202, copyright 2019;[27] permission conveyed through Copyright Clearance Center, Inc.

TABLE 14 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT for diagnosis of active TB

	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	270/328	82.3 (77.8 to 86.1)	220/327	67.3 (62.0 to 72.1)
Culture-positive TB	201/234	85.9 (80.9 to 89.8)	163/231	70.6 (64.4 to 76.1)
Culture-negative TB	59/83	71.1 (60.6 to 79.7)	48/84	57.1 (46.5 to 67.2)
Smear-positive TB	50/60	83.3 (72.0 to 90.7)	42/56	75.0 (62.3 to 84.5)
Smear-negative TB	179/216	82.9 (77.3 to 87.3)	148/222	67.7 (60.2 to 72.5)
Pulmonary TB	87/113	77.0 (68.4 to 83.8)	78/114	68.4 (59.4 to 76.2)
Extrapulmonary TB	147/175	84.0 (77.9 to 88.7)	113/171	66.1 (58.7 to 72.8)
Specificity for a diagnosis of active TB				
Active TB excluded	319/386	82.6 (78.6 to 86.1)	304/378	80.4 (76.1 to 84.1)
Active TB excluded, TST negative, no risk factors for LTBI	87/94	92.3 (85.4 to 96.4)	85/91	93.4 (86.4 to 96.9)
Predictive values				
PPV	270/337	80.1 (75.5 to 84.0)	220/294	74.8 (69.6 to 79.5)
NPV	319/377	84.6 (80.6 to 87.9)	304/411	74.0 (69.5 to 78.0)
Likelihood ratios				
Positive likelihood ratio	–	4.74 (3.79 to 5.93)	–	3.44 (2.76 to 4.27)
Negative likelihood ratio	–	0.21 (0.17 to 0.27)	–	0.41 (0.35 to 0.48)

Comparison of diagnostic accuracy of T-SPOT.*TB* and QuantiFERON GOLD In-Tube

Excluding indeterminate IGRA results, there were 714 T-SPOT.*TB* results and 705 QFT-GIT results. Based on analyses using GEE models, the sensitivity of T-SPOT.*TB* was superior to that of QFT-GIT (relative sensitivity 1.22, 95% CI 1.14 to 1.31; $p < 0.001$), but there was no statistical evidence of a difference in specificity (relative specificity 1.02, 95% CI 0.97 to 1.08; $p = 0.3$) (*Table 15*). Excluding the 33 borderline T-SPOT.*TB* results (see *Appendix 7*, *Table 60*), the results for relative sensitivity were similar to those from the main analysis but there was statistical evidence of a difference in specificity; the relative sensitivity (95% CI) was 1.20 (1.12 to 1.29; $p < 0.001$) and the relative specificity (95% CI) was 1.07 (1.02 to 1.12; $p = 0.004$).

TABLE 15 Comparison of diagnostic accuracy of T-SPOT.*TB* and QFT-GIT

Test	Number of test results[a]	Sensitivity (95% CI)	Number of test results[b]	Specificity (95% CI)
T-SPOT.*TB*	328	82.2 (77.7 to 85.9)	386	83.0 (78.9 to 86.4)
QFT-GIT	327	67.3 (62.1 to 72.2)	378	81.0 (76.8 to 84.6)
Ratio (95% CI);[c] *p*-value	–	1.22 (1.14 to 1.31); < 0.001	–	1.02 (0.97 to 1.08); 0.3

a Number of test results among those with active TB.
b Number of test results among those without active TB.
c The ratio of the sensitivity (or specificity) of T-SPOT.*TB* to that of QFT-GIT. The natural outputs from GEE models are odds ratios. Ratios of sensitivities (relative sensitivity) and ratios of specificities (relative specificity) were computed post estimation of the models. CIs were obtained using the delta method.

Note
Sensitivity and specificity values are presented as percentages.

When all IGRA results were analysed and indeterminate IGRA results were included as test positives in a sensitivity analysis, the analysis included 768 T-SPOT.*TB* results and 778 QFT-GIT results. Similar to the primary analysis, there was statistical evidence of a difference in sensitivity (relative sensitivity 1.19, 95% CI 1.12 to 1.28; $p < 0.001$) and no evidence of a difference in specificity (relative specificity 1.05 95% CI 0.98 to 1.13; $p = 0.1$) (see *Appendix 7*, *Table 61*).

Subgroup analyses for T-SPOT.*TB* and QuantiFERON GOLD In-Tube

Human immunodeficiency virus-positive and -negative patients

Human immunodeficiency virus co-infected patients are the highest-risk subgroup for TB. In this section we briefly present results for the 135 HIV-positive patients in the main cohort and address the primary objectives related to this subgroup in *Chapter 5* using all HIV-positive patients recruited into the IDEA study. *Appendix 8* (see *Table 62*) shows the results for T-SPOT.*TB* and QFT-GIT against active TB status. *Appendix 8* (see *Table 63*) shows the cross-tabulation of T-SPOT.*TB* and QFT-GIT results in HIV-positive patients. For 134 patients, results were available for both T-SPOT.*TB* and QFT-GIT.

The number of test results for active TB and non-active TB patients who were available for the analyses of test performance in HIV-positive patients was small (see *Appendix 8*, *Table 62*). Using culture-confirmed and highly probable active TB cases and excluding indeterminate IGRA results, sensitivity was 63.2% (95% CI 41.0% to 80.9%) for T-SPOT.*TB* and 56.5% (95% CI 36.8% to 74.4%) for QFT-GIT. Among all non-active TB patients, specificity was 89.9% (95% CI 81.3% to 94.8%) for T-SPOT.*TB* and 92.0% (95% CI 84.3% to 96.1%) for QFT-GIT (*Table 16*). For T-SPOT.*TB* and QFT-GIT, the PPVs were 60.0%

TABLE 16 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in HIV-positive patients

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	12/19	63.2 (41.0 to 80.9)	13/23	56.5 (36.8 to 74.4)
Culture-positive TB	7/11	63.6 (35.4 to 84.8)	8/13	61.5 (35.5 to 82.3)
Culture-negative TB	5/8	62.5 (30.6 to 86.3)	5/9	55.6 (26.7 to 81.1)
Smear-positive TB	1/4	25.0 (4.6 to 69.9)	3/5	60.0 (23.1 to 88.2)
Smear-negative TB	7/11	63.6 (35.4 to 84.8)	7/14	50.0 (26.8 to 73.2)
Pulmonary TB	3/5	60.0 (23.1 to 88.2)	5/8	62.5 (30.6 to 86.3)
Extrapulmonary TB	7/10	70.0 (39.7 to 89.2)	6/12	50.0 (25.4 to 74.6)
Specificity for a diagnosis of active TB				
Active TB excluded	71/79	89.9 (81.3 to 94.8)	80/87	92.0 (84.3 to 96.1)
Active TB excluded, TST negative, no risk factors for LTBI	28/29	96.6 (82.8 to 99.4)	36/38	94.7 (82.7 to 98.5)
Predictive values				
PPV	12/20	60.0 (38.7 to 78.1)	13/20	65.0 (43.3 to 81.9)
NPV	71/78	91.0 (82.6 to 95.6)	80/90	88.9 (80.7 to 93.9)
Likelihood ratios				
Positive likelihood ratio	–	6.24 (2.97 to 13.1)	–	7.03 (3.17 to 15.6)
Negative likelihood ratio	–	0.41 (0.22 to 0.74)	–	0.47 (0.30 to 0.76)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were excluded from these analyses. See *Appendix 8*, *Table 64*, for the sensitivity analyses with indeterminates included as test positives.

(95% CI 38.7% to 78.1%) and 65.0% (95% CI 43.3% to 81.9%), respectively, and the NPVs were 91.0% (95% CI 82.6% to 95.6%) and 88.9% (95% CI 80.7% to 93.9%), respectively.

For HIV-negative patients (*Table 17*), results were generally similar to those of the entire cohort that included both HIV-negative and -positive patients (see *Table 14*). For both subgroups, inclusion of indeterminate IGRA results in sensitivity analyses led to higher sensitivities, lower specificities and no change in NPVs, which is as expected because of the inclusion of indeterminates as test positives (see *Appendix 8*, *Tables 64* and *65*).

Other key patient subgroups

For our secondary objective of quantifying test accuracy in three key patient subgroups (patients with pre-existing diabetes mellitus, end-stage renal failure and iatrogenic immunosuppression), numbers of participants in the subgroups were small but, nonetheless, we evaluated the performance of IGRAs in patients with diabetes (see *Table 4*). The results for T-SPOT.*TB* and QFT-GIT against active TB status are shown in *Appendix 9*, *Table 66*. Owing to the limited number of data for the TST, only the performance of T-SPOT.*TB* and QFT-GIT is presented below (see *Table 18*). *Table 67* in *Appendix 9* shows the cross-tabulation of T-SPOT.*TB* and QFT-GIT results in the 88 patients with diabetes mellitus. For 86 patients, the results were available for both T-SPOT.*TB* and QFT-GIT.

TABLE 17 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in HIV-negative patients

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	258/309	83.5 (79.0 to 87.2)	207/304	68.1 (62.7 to 73.1)
Culture-positive TB	194/223	87.0 (81.9 to 90.8)	155/218	71.1 (64.8 to 76.7)
Culture-negative TB	54/75	72.0 (61.0 to 80.9)	43/75	57.3 (46.1 to 67.9)
Smear-positive TB	49/56	87.5 (76.4 to 93.8)	39/51	76.5 (63.2 to 86.0)
Smear-negative TB	172/205	83.9 (78.3 to 88.3)	141/208	67.8 (61.2 to 73.8)
Pulmonary TB	84/108	77.8 (69.1 to 84.6)	73/106	68.9 (59.5 to 76.9)
Extrapulmonary TB	140/165	84.8 (78.6 to 89.5)	107/159	67.3 (59.7 to 74.1)
Specificity for a diagnosis of active TB				
Active TB excluded	248/307	80.8 (76.0 to 84.8)	224/291	77.0 (71.8 to 81.4)
Active TB excluded, TST negative, no risk factors for LTBI	59/65	90.8 (81.3 to 95.7)	49/53	92.5 (82.1 to 97.0)
Predictive values				
PPV	258/317	81.4 (76.7 to 85.3)	207/274	75.6 (70.1 to 80.3)
NPV	248/299	82.9 (78.3 to 86.8)	224/321	69.8 (64.6 to 74.6)
Likelihood ratios				
Positive likelihood ratio	–	4.35 (3.44 to 5.49)	–	2.96 (2.37 to 3.70)
Negative likelihood ratio	–	0.20 (0.16 to 0.26)	–	0.42 (0.35 to 0.49)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were excluded from these analyses. See *Appendix 8*, *Table 65*, for sensitivity analyses with indeterminates included as test positives.

Excluding indeterminate IGRA results, sensitivity was 68.0% (95% CI 48.4% to 82.8%) for T-SPOT.*TB* and 55.6% (95% CI 37.3% to 72.4%) for QFT-GIT among patients with culture-confirmed and highly probable active TB. Based on all non-active TB patients, specificity was 77.6% (95% CI 64.1% to 87.0%) for T-SPOT.*TB* and 78.7% (95% CI 65.1% to 88.0%) for QFT-GIT (*Table 18*). The PPVs were 60.7% (95% CI 42.4% to 76.4%) and 60.0% (95% CI 40.7% to 76.6%), and the NPVs were 82.6% (95% CI 69.3% to 90.9%) and 75.5% (95% CI 61.9% to 85.4%) for T-SPOT.*TB* and QFT-GIT, respectively. These analyses were based on limited data and so should be interpreted with caution. *Appendix 9* (see *Table 68*) shows the performance of both tests when indeterminate IGRA results were included as test positives in sensitivity analyses.

*Variation in relative performance of T-SPOT.*TB *and QuantiFERON-TB Gold Plus*

We explored the effect of HIV co-infection, diabetes mellitus, smoking and clinical setting on the relative test performance of T-SPOT.*TB* and QFT-GIT. Each covariate was investigated in a separate regression model. The sensitivities of both tests were superior in HIV-negative patients than in HIV-positive patients. However, there was no statistical evidence of an effect of HIV infection status on relative sensitivity ($p = 0.2$). In contrast, specificities were superior in HIV-positive patients than in HIV-negative patients, but there was no statistical evidence of an effect of HIV infection status on relative specificity ($p = 0.2$).

Although the sensitivities and specificities of both tests were higher in those without diabetes mellitus than in those with diabetes mellitus, there was no statistical evidence of an effect on relative test performance (*Table 19*). Although the *p*-value from the Wald test of the interaction between test type and smoking

TABLE 18 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in patients with diabetes mellitus

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	17/25	68.0 (48.4 to 82.8)	15/27	55.6 (37.3 to 72.4)
Culture-positive TB	14/21	66.7 (45.4 to 82.8)	12/22	54.5 (34.7 to 73.1)
Culture-negative TB	3/4	75.0 (30.1 to 95.4)	3/5	60.0 (23.1 to 88.2)
Smear-positive TB	7/9	77.8 (45.3 to 93.7)	6/9	66.7 (35.4 to 87.9)
Smear-negative TB	9/14	64.3 (38.8 to 83.7)	7/16	43.8 (23.1 to 66.8)
Pulmonary TB	4/8	50.0 (21.5 to 78.5)	5/9	55.6 (26.7 to 81.1)
Extrapulmonary TB	10/13	76.9 (49.7 to 91.8)	8/14	57.1 (32.6 to 78.6)
Specificity for a diagnosis of active TB				
Active TB excluded	38/49	77.6 (64.1 to 87.0)	37/47	78.7 (65.1 to 88.0)
Active TB excluded, TST negative, no risk factors for LTBI	6/7	85.7 (48.7 to 97.4)	4/5	80.0 (37.6 to 96.4)
Predictive values				
PPV	17/28	60.7 (42.4 to 76.4)	15/25	60.0 (40.7 to 76.6)
NPV	38/46	82.6 (69.3 to 90.9)	37/49	75.5 (61.9 to 85.4)
Likelihood ratios				
Positive likelihood ratio	–	3.03 (1.69 to 5.44)	–	2.61 (1.37 to 4.98)
Negative likelihood ratio	–	0.41 (0.23 to 0.75)	–	0.57 (0.36 to 0.88)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were excluded from these analyses. See *Appendix 9*, *Table 68* for sensitivity analyses with indeterminates included as test positives.

TABLE 19 Effect of comorbidities, smoking and clinical setting on relative test performance of T-SPOT.*TB* and QFT-GIT

			Sensitivity (95% CI)					Specificity (95% CI)		
Covariate	N_{TD}	N_{QD}	T-SPOT.*TB*	QFT-GIT	*p*-value	N_{TND}	N_{QND}	T-SPOT.*TB*	QFT-GIT	*p*-value
None	328	327	82.2 (77.7 to 85.9)	67.3 (62.1 to 72.2)	N/A	386	378	83.0 (78.9 to 86.4)	81.0 (76.8 to 84.6)	N/A
HIV infection status										
Negative	309	304	83.5 (79.0 to 87.3)	68.2 (62.8 to 73.2)	–	307	291	80.8 (76.0 to 84.8)	77.5 (72.4 to 81.9)	–
Positive	19	23	60.1 (38.0 to 78.7)	56.1 (36.1 to 74.35)	0.2	79	87	90.0 (81.2 to 94.7)	92.3 (84.7 to 96.3)	0.2
Diabetes										
No	303	300	83.3 (78.7 to 87.1)	68.3 (62.9 to 73.3)	–	337	331	83.8 (79.6 to 87.3)	81.2 (76.7 to 85.0)	–
Yes	25	27	67.8 (47.8 to 82.9)	55.6 (36.9 to 72.8)	0.5	49	47	76.9 (63.2 to 86.6)	79.8 (66.0 to 88.8)	0.4
Smoking status[a]										
Never smoked	238	239	83.8 (78.6 to 87.9)	65.3 (59.1 to 71.1)	–	220	216	83.0 (77.5 to 87.3)	78.0 (72.1 to 83.0)	–
Previous smoke	34	29	77.0 (60.2 to 88.1)	82.2 (64.5 to 92.2)	0.01	82	77	81.0 (71.1 to 88.0)	80.4 (70.2 to 87.8)	0.4
Current smoke	56	59	78.5 (66.0 to 87.3)	68.3 (55.6 to 78.8)	0.2	82	83	84.3 (75.0 to 90.6)	88.8 (80.2 to 94.0)	0.06
Clinical setting										
Inpatient	104	96	77.7 (69.0 to 84.6)	64.3 (54.5 to 73.1)	–	140	132	86.8 (80.2 to 91.5)	90.0 (83.6 to 94.1)	–
Outpatient	224	231	84.2 (78.8 to 88.4)	68.7 (62.4 to 74.3)	0.4	246	246	80.6 (75.2 to 85.0)	76.0 (70.3 to 80.9)	0.09

N/A, not applicable; N_{QD}, number of QFT-GIT results in active TB cases; N_{QND}, number of QFT-GIT results in non-active TB patients; N_{TD}, number of T-SPOT.*TB* results in active TB cases; N_{TND}, number of T-SPOT.*TB* results in non-active TB patients.

a The global effect of smoking status on the relative performance of T-SPOT.*TB* and QFT-GIT was not significant for sensitivity ($p = 0.7$) or for specificity ($p = 0.4$).

Note

Sensitivity and specificity are presented as percentages. The *p*-values are from the Wald tests of the interaction between test type and the covariate.

gave a *p*-value of 0.01 for the difference in relative sensitivity between the group that had never smoked and the group of previous smokers, the global test of whether or not relative test performance varied across smoking subgroups was not statistically significant for sensitivity ($p = 0.7$) or for specificity ($p = 0.4$). The sensitivities of T-SPOT.*TB* and QFT-GIT were lower in inpatients than in outpatients; the converse was the case for their specificities.

Diagnostic accuracy of second-generation interferon gamma release assays

The diagnostic accuracy of ESAT-6, CFP-10 and four second-generation IGRAs (Rv3615c, Rv2654, Rv3879c and Rv3873) is presented in the following sections as individual tests and as test combinations. The accuracy of different test combinations was compared with that of T-SPOT.*TB*.

Individual antigens

The results of the six antigens cross-tabulated against active TB status are shown in *Table 20* and their diagnostic accuracy in *Table 21*. Based on all culture-confirmed and highly probable active TB cases and excluding indeterminate IGRA results, the three antigens with the highest sensitivities were Rv3615c, ESAT-6 and CFP-10. The sensitivities were 78.0% (95% CI 73.1% to 82.1%), 69.4% (95% CI 64.1% to 74.1%) and 71.6% (95% CI 66.5% to 76.2%), and the specificities were 82.7% (95% CI 78.7% to 86.1%), 87.8% (95% CI 84.2% to 90.7%) and 86.3% (95% CI 82.5% to 89.4%) for Rv3615c, ESAT-6 and CFP-10, respectively. The performance of the remaining three antigens was very poor with sensitivities of 34.2% (95% CI 29.2% to 39.5%), 37.4% (95% CI 33.3% to 42.9%) and 38.2% (95% CI 33.1% to 43.7%) for Rv3873, Rv2654 and Rv3879c, respectively. However, their specificities were high: 95.1% (95% CI 92.4% to 96.8%), 91.7% (95% CI 88.5% to 94.0%) and 93.3% (95% CI 90.3% to 95.4%) for Rv3873, Rv2654 and Rv3879c, respectively. *Table 21* also shows the performance of the tests in different clinical groups. For each antigen, sensitivity was higher in those with culture-confirmed active TB than in those with culture-negative active TB.

TABLE 20 Results of second-generation IGRAs by diagnostic category

	Dosanjh category, *n*								
IGRA result	1	2	3	4A	4B	4C	4D	4A–D	Total, *n*
Rv3615c									
Positive	171	53	13	0	13	35	6	54	291
Negative	38	33	24	6	28	200	86	320	415
Borderline	21	6	1	0	5	6	2	13	41
Indeterminate	16	7	3	1	0	18	17	36	62
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
SFCs, median (range)	25 (0–642)	19 (0–450)	1 (0–102)	0 (0–4)	1 (0–134)	0 (0–493)	0 (0–406)	0 (0–493)	2 (0–642)
Rv3873									
Positive	59	23	3	0	2	11	3	16	101
Negative	146	64	34	6	43	225	92	366	610
Borderline	22	5	0	0	0	3	0	3	30
Indeterminate	19	7	4	1	1	20	16	38	68
Missing	15	3	2	0	2	8	6	16	36

continued

TABLE 20 Results of second-generation IGRAs by diagnostic category (*continued*)

	Dosanjh category, *n*								
IGRA result	1	2	3	4A	4B	4C	4D	4A–D	Total, *n*
Total	261	102	43	7	48	267	117	439	845
SFCs, median (range)	2 (0–178)	1 (0–181)	0 (0–145)	0 (0–0)	0 (0–30)	0 (0–64)	0 (0–71)	0 (0–71)	0 (0–181)
Rv2654									
Positive	74	27	6	0	5	13	5	23	130
Negative	138	61	29	6	36	222	88	352	580
Borderline	14	4	2	0	4	4	1	9	29
Indeterminate	20	7	4	1	1	20	17	39	70
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
SFCs, median (range)	2 (0–260)	1 (0–395)	1 (0–55)	0 (0–1)	1 (0–44)	0 (0–370)	0 (0–57)	0 (0–370)	0 (0–395)
Rv3879c									
Positive	66	30	4	0	3	13	3	19	119
Negative	138	59	32	6	41	222	90	359	588
Borderline	23	3	1	0	1	4	2	7	34
Indeterminate	19	7	4	1	1	20	16	38	68
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
SFCs, median (range)	2 (0–241)	1 (0–225)	0 (0–154)	0 (0–3)	0 (0–42)	0 (0–83)	0 (0–58)	0 (0–83)	0 (0–241)
ESAT-6									
Positive	149	54	9	0	14	16	5	35	247
Negative	63	36	27	6	29	217	87	339	465
Borderline	18	3	1	0	2	8	2	12	34
Indeterminate	16	6	4	1	1	18	17	37	63
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
SFCs, median (range)	14 (0–642)	13 (0–492)	0 (0–325)	0 (0–1)	2 (0–274)	0 (0–210)	0 (0–147)	0 (0–274)	1 (0–642)
CFP-10									
Positive	157	57	11	0	14	24	4	42	267
Negative	57	35	26	6	29	208	90	333	451
Borderline	17	1	1	0	2	9	0	11	30
Indeterminate	15	6	3	1	1	18	17	37	61
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
SFCs, median (range)	18 (0–642)	13 (0–437)	1 (0–315)	0 (0–3)	1 (0–160)	0 (0–166)	0 (0–148)	0 (0–166)	1 (0–642)

TABLE 21 Diagnostic accuracy of individual second-generation IGRAs, ESAT-6 and CFP-10

	Antigen											
	Rv3615c		Rv3879c		Rv3873		Rv2654		ESAT-6		CFP-10	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*
Sensitivity for a diagnosis of active TB												
All TB	251/322	78.0 (73.1 to 82.1)	122/319	38.2 (33.1 to 43.7)	224/323	69.4 (64.1 to 74.1)	224/323	69.4 (64.1 to 74.1)	224/323	69.4 (64.1 to 74.1)	232/324	71.6 (66.5 to 76.2)
Culture-positive TB	192/230	83.5 (78.1 to 87.7)	89/227	39.2 (33.1 to 45.7)	167/230	72.6 (66.5 to 78.0)	167/230	72.6 (66.5 to 78.0)	167/230	72.6 (66.5 to 78.0)	174/231	75.3 (69.4 to 80.4)
Culture-negative TB	51/82	62.2 (51.4 to 71.9)	28/82	34.1 (24.8 to 44.9)	48/83	57.8 (47.1 to 67.9)	48/83	57.8 (47.1 to 67.9)	48/83	57.8 (47.1 to 67.9)	50/82	61.0 (50.2 to 70.8)
Smear-positive TB	48/58	82.8 (71.1 to 90.4)	22/57	38.6 (27.1 to 51.6)	39/58	67.2 (54.4 to 77.9)	39/58	67.2 (54.4 to 77.9)	39/58	67.2 (54.4 to 77.9)	43/59	72.9 (60.4 to 82.6)
Smear-negative TB	162/212	76.4 (70.3 to 81.6)	86/212	40.6 (34.2 to 47.3)	151/213	70.9 (64.5 to 76.6)	151/213	70.9 (64.5 to 76.6)	151/213	70.9 (64.5 to 76.6)	155/215	72.1 (65.7 to 77.7)
Pulmonary TB	85/111	76.6 (67.9 to 83.5)	35/111	31.5 (23.6 to 40.7)	71/111	64.0 (54.7 to 72.3)	71/111	64.0 (54.7 to 72.3)	71/111	64.0 (54.7 to 72.3)	75/113	66.4 (57.3 to 74.4)
Extrapulmonary TB	133/172	77.3 (70.5 to 83.0)	71/170	41.8 (34.6 to 49.3)	124/173	71.7 (64.5 to 77.9)	124/173	71.7 (64.5 to 77.9)	124/173	71.7 (64.5 to 77.9)	127/172	73.8 (66.8 to 79.8)
Specificity for a diagnosis of active TB												
Active TB excluded	320/387	82.7 (78.7 to 86.1)	359/385	93.3 (90.3 to 95.4)	339/386	87.8 (84.2 to 90.7)	339/386	87.8 (84.2 to 90.7)	339/386	87.8 (84.2 to 90.7)	333/386	86.3 (82.5 to 89.4)
Active TB excluded, TST negative, no risk factors for LTBI	86/94	91.5 (84.1 to 95.6)	90/95	94.7 (88.3 to 97.7)	87/94	92.6 (85.4 to 96.3)	87/94	92.6 (85.4 to 96.3)	87/94	92.6 (85.4 to 96.3)	90/94	95.7 (89.6 to 98.3)

continued

TABLE 21 Diagnostic accuracy of individual second-generation IGRAs, ESAT-6 and CFP-10 (*continued*)

	Antigen											
	Rv3615c		Rv3879c		Rv3873		Rv2654		ESAT-6		CFP-10	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*
Predictive values												
PPV	251/318	78.9 (74.1 to 83.1)	122/148	82.4 (75.5 to 87.7)	224/271	82.7 (77.7 to 86.7)	224/271	82.7 (77.7 to 86.7)	224/271	82.7 (77.7 to 86.7)	232/285	82.4 (77.3 to 86.8)
NPV	320/391	81.8 (77.7 to 85.4)	359/556	64.6 (60.5 to 68.4)	339/492	77.4 (73.3 to 81.1)	339/492	77.4 (73.3 to 81.1)	339/492	77.4 (73.3 to 81.1)	333/425	78.4 (74.2 to 82.0)
Likelihood ratios												
Positive likelihood ratio		4.50 (3.59 to 5.64)		5.66 (3.81 to 8.42)		5.70 (4.32 to 7.52)		5.70 (4.32 to 7.52)		5.70 (4.32 to 7.52)		5.22 (4.02 to 6.76)
Negative likelihood ratio		0.27 (0.22 to 0.33)		0.66 (0.61 to 0.73)		0.35 (0.30 to 0.41)		0.35 (0.30 to 0.41)		0.35 (0.30 to 0.41)		0.33 (0.28 to 0.39)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were excluded from these analyses. See *Appendix 10*, *Table 69*, for sensitivity analyses with indeterminates included as test positives.

The results of the sensitivity analyses in which indeterminate test results were included as test positives are given in *Appendix 10*, *Table 69*. The results show slight increases in sensitivity with reductions in specificities.

Combinations of antigens

Various combinations of the antigens are shown cross-tabulated against diagnostic category in *Table 22*. *Table 23* shows the diagnostic accuracy of the combinations. The sensitivity of the four-antigen (CFP-10, ESAT-6, Rv3615c and Rv3879c) and three-antigen (CFP-10, ESAT-6 and Rv3615c) combinations containing both CFP-10 and ESAT-6 were identical irrespective of whether all active TB cases or different subgroups were analysed, as shown in *Table 23*. For this three-antigen combination, the sensitivity was 89.9% (95% CI 86.2% to 92.7%) among all active TB cases and the specificity was 76.5% (95% CI 72.0% to

TABLE 22 Results of combinations of IGRAs against diagnostic category

	Dosanjh category, *n*								
IGRA combination result	1	2	3	4A	4B	4C	4D	4A–D	Total, *n*
CFP-10 + Rv3615c									
Positive	197	66	15	0	19	44	7	70	348
Negative	15	23	23	6	22	188	86	302	363
Borderline	22	4	1	0	5	9	1	15	42
Indeterminate	12	6	2	1	0	18	17	36	56
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
CFP-10 + Rv3615c + Rv3879c									
Positive	197	66	15	0	19	49	8	76	354
Negative	14	22	23	6	22	184	84	296	355
Borderline	23	5	1	0	5	8	3	16	45
Indeterminate	12	6	2	1	0	18	16	35	55
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
ESAT-6 + CFP-10 + Rv3615c									
Positive	203	70	15	0	20	46	8	74	362
Negative	13	20	23	6	22	184	84	296	352
Borderline	18	4	1	0	4	11	2	17	40
Indeterminate	12	5	2	1	0	18	17	36	55
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845
ESAT-6 + CFP-10 + Rv3615c + Rv3879c									
Positive	203	70	16	0	20	49	9	78	367
Negative	13	20	22	6	22	180	82	290	345
Borderline	18	4	1	0	4	12	4	20	43
Indeterminate	12	5	2	1	0	18	16	35	54
Missing	15	3	2	0	2	8	6	16	36
Total	261	102	43	7	48	267	117	439	845

TABLE 23 Diagnostic accuracy of combinations of second-generation IGRAs

	Antigen combination							
	ESAT-6 + CFP-10 + Rv3615c + Rv3879c		ESAT-6 + CFP-10 + Rv3615		CFP-10 + Rv3615c + Rv3879c		CFP-10 + Rv3615c	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB								
All TB	295/328	89.9 (86.2 to 92.7)	295/328	89.9 (86.2 to 92.7)	291/327	89.0 (85.1 to 91.9)	289/327	88.4 (84.5 to 91.4)
Culture-positive TB	221/234	94.4 (90.7 to 96.7)	221/234	94.4 (90.7 to 96.7)	220/234	94.0 (90.2 to 96.4)	219/234	93.6 (89.7 to 96.1)
Culture-negative TB	63/83	75.9 (65.7 to 83.8)	63/83	75.9 (65.7 to 83.8)	61/82	74.4 (64.0 to 82.6)	60/82	73.2 (62.7 to 81.6)
Smear-positive TB	57/60	95.0 (86.3 to 98.3)	57/60	94.9 (86.1 to 98.3)	57/60	95.0 (86.3 to 98.3)	57/60	95.0 (86.3 to 98.3)
Smear-negative TB	192/216	88.9 (84.0 to 92.4)	192/216	88.9 (84.0 to 92.4)	189/215	87.9 (82.9 to 91.6)	187/215	87.0 (81.8 to 90.8)
Pulmonary TB	101/113	89.4 (82.4 to 93.8)	101/113	89.4 (82.4 to 93.8)	101/113	89.4 (82.4 to 93.8)	100/113	88.5 (81.3 to 93.2)
Extra pulmonary TB	156/175	89.1 (83.7 to 92.9)	156/175	89.1 (83.7 to 92.9)	152/174	87.4 (81.6 to 91.5)	151/174	86.8 (80.9 to 91)
Specificity for a diagnosis of active TB								
Active TB excluded	290/388	74.7 (70.2 to 78.8)	296/387	76.5 (72.0 to 80.4)	296/388	76.3 (71.8 to 80.3)	302/387	78.0 (73.7 to 81.9)
Active TB excluded, TST negative, no risk factors for LTBI	82/95	86.3 (78.0 to 91.8)	84/94	89.4 (81.5 to 94.1)	84/95	88.4 (80.4 to 93.4)	86/94	91.5 (84.1 to 95.6)
Predictive values								
PPV	295/393	75.1 (70.6 to 79.1)	295/386	76.4 (71.9 to 80.4)	291/383	76.0 (71.5 to 80.0)	289/374	77.3 (72.8 to 81.2)
NPV	290/323	89.8 (86.0 to 92.6)	296/329	90.0 (86.3 to 92.8)	296/332	89.2 (85.4 to 92.1)	302/340	88.8 (85.0 to 91.8)
Likelihood ratios								
Positive likelihood ratio		3.56 (2.99 to 4.24)		3.83 (3.18 to 4.59)		3.75 (3.13 to 4.51)		4.02 (3.32 to 4.88)
Negative likelihood ratio		0.14 (0.10 to 0.19)		0.13 (0.10 to 0.18)		0.14 (0.11 to 0.20)		0.15 (0.11 to 0.20)

80.4%) in all non-active TB patients. Another three-antigen combination (CFP-10, Rv3615c and Rv3879c) gave a similar sensitivity and specificity of 89.0% (95% CI 85.1% to 91.9%) and 76.3% (95% CI 71.8% to 80.3%). Reducing this combination to a two-antigen combination of CFP-10 and Rv3615c gave a sensitivity of 88.4% (84.5% to 91.4%) and a specificity of 78.0% (95% CI 73.7% to 81.9%). See *Appendix 10*, *Table 70*, for sensitivity analyses with borderline test results excluded as test positives. For the results of the sensitivity analyses including indeterminate test results, see *Appendix 10*, *Table 71*.

Sensitivity, specificity and predictive values are presented as percentages. For both the four-antigen (ESAT-6, CFP-10, Rv3615c and Rv3879c) and three-antigen (ESAT, CFP-10, Rv3615), there were 74 test positives out of 94 highly probable TB cases, with a sensitivity (95% CI) of 78.7% (69.4% to 85.8%). For the other three-antigen (CFP-10, Rv3615c, Rv3879c), there were 71 test positives out of 93 highly probable TB cases, with a sensitivity (95% CI) of 76.3% (66.8% to 83.8%). For the two-antigen combination of CFP-10 and Rv3615, there were 70 test positives out of 93 highly probable TB cases, with a sensitivity (95% CI) of 75.3% (65.6% to 82.9%). Note that in the primary analyses of combinations of next-generation IGRAs, borderline test results were included as test positives. Sensitivity analyses were performed with borderline IGRA results excluded and the results are shown in *Appendix 10*, *Table 70*. Indeterminate IGRA results were excluded from these analyses. See *Appendix 10*, *Table 71*, for the sensitivity analyses with indeterminates included as test positives.

*Comparison of novel antigen combinations with T-SPOT.*TB

The results of individual antigens classified against those of T-SPOT.*TB* show few discordant results between Rv3615c and T-SPOT.*TB*, unlike those of the other three novel antigens (Rv3873, Rv2654 and Rv3879c), which are not components of T-SPOT.*TB* (*Table 24*).

TABLE 24 Cross-tabulation of individual IGRA results against T-SPOT.*TB* results

T-SPOT.*TB*, *n*						
IGRA result	**Positive**	**Negative**	**Borderline**	**Indeterminate**	**Missing**	**Total**
Rv3615c						
Positive	247	36	8	0	0	291
Negative	48	351	16	0	0	415
Borderline	18	13	8	2	0	41
Indeterminate	6	0	1	55	0	62
Missing	0	0	0	0	36	36
Total	319	400	33	57	36	845
Rv3873						
Positive	95	4	1	1	0	101
Negative	187	393	30	0	0	610
Borderline	27	3	0	0	0	30
Indeterminate	10	0	2	56	0	68
Missing	0	0	0	0	36	36
Total	319	400	33	57	36	845

continued

TABLE 24 Cross-tabulation of individual IGRA results against T-SPOT.*TB* results (*continued*)

T-SPOT.*TB*, *n*						
IGRA result	**Positive**	**Negative**	**Borderline**	**Indeterminate**	**Missing**	**Total**
Rv2654						
Positive	118	9	3	0	0	130
Negative	169	384	27	0	0	580
Borderline	21	7	1	0	0	29
Indeterminate	11	0	2	57	0	70
Missing	0	0	0	0	36	36
Total	319	400	33	57	36	845
Rv3879c						
Positive	109	6	3	1	0	119
Negative	170	390	28	0	0	588
Borderline	30	4	0	0	0	34
Indeterminate	10	0	2	56	0	68
Missing	0	0	0	0	36	36
Total	319	400	33	57	36	845
ESAT-6						
Positive	247	0	0	0	0	247
Negative	46	400	19	0	0	465
Borderline	21	0	13	0	0	34
Indeterminate	5	0	1	57	0	63
Missing	0	0	0	0	36	36
Total	319	400	33	57	36	845
CFP-10						
Positive	267	0	0	0	0	267
Negative	41	400	10	0	0	451
Borderline	8	0	22	0	0	30
Indeterminate	3	0	1	57	0	61
Missing	0	0	0	0	36	36
Total	319	400	33	57	36	845

Tables 25 and *26* show comparisons of the sensitivity and specificity of the different test combinations versus those of T-SPOT.*TB*. Indeterminate test results were excluded from these analyses (see *Appendix 10*, *Tables 72* and *73* for the results of sensitivity analyses with indeterminates included). The sensitivity of T-SPOT.*TB* was lower than those of any of the four test combinations (see *Table 25*), but had higher specificity (see *Table 26*). The sensitivity of the two-antigen combination of CFP-10 and Rv3615c was 7% higher than that of T-SPOT.*TB*, with a relative sensitivity of 1.07 (95% CI 1.03 to 1.11). However, the specificity of the combined CFP-10 and Rv3615c was 5% lower than that of T-SPOT.*TB*, with a relative specificity of 0.95 (95% CI 0.92 to 0.98).

TABLE 25 Comparison of sensitivity of IGRA combinations and T-SPOT.*TB*

Test	*n*[a]	Sensitivity, % (95% CI)	Relative sensitivity[b] (95% CI)	*p*-value
T-SPOT.*TB* (ESAT-6 + CFP-10)	328	82.3 (77.7 to 86.1)	–	–
CFP-10 + Rv3615c	327	88.4 (84.5 to 91.5)	1.07 (1.03 to 1.12)	< 0.001
CFP-10 + Rv3615c + Rv3879c	327	89.0 (85.1 to 92.0)	1.08 (1.04 to 1.12)	< 0.001
ESAT-6 + CFP-10 + Rv3615c	328	89.9 (86.2 to 92.8)	1.09 (1.06 to 1.13)	< 0.001
ESAT-6 + CFP-10 + Rv3615c + Rv3879c	328	89.9 (86.2 to 92.8)	1.09 (1.06 to 1.13)	< 0.001

a Number of test results.
b Sensitivity of combination test divided by sensitivity of T-SPOT.*TB*.
Note
Indeterminate test results were excluded in these analyses. See *Appendix 10* for the results of the sensitivity analyses with indeterminates included as test positives.

TABLE 26 Comparison of specificity of IGRA combinations and T-SPOT.*TB*

Test	*n*[a]	Specificity, % (95% CI)	Relative specificity[b] (95% CI)	*p*-value
T-SPOT.*TB* (ESAT-6 + CFP-10)	386	82.5 (78.4 to 86.0)	–	–
CFP-10 + Rv3615c	387	78.0 (73.6 to 81.9)	0.95 (0.91 to 0.98)	0.001
CFP-10 + Rv3615c + Rv3879c	388	76.3 (71.8 to 80.3)	0.93 (0.89 to 0.96)	< 0.001
ESAT-6 + CFP-10 + Rv3615c	387	76.5 (72.0 to 80.5)	0.93 (0.90 to 0.96)	< 0.001
ESAT-6 + CFP-10 + Rv3615c + Rv3879c	388	74.7 (70.2 to 78.8)	0.91 (0.88 to 0.94)	< 0.001

a Number of test results.
b Sensitivity of combination test divided by sensitivity of T-SPOT.*TB*.
Note
Indeterminate test results were excluded in these analyses. See *Appendix 10* for the results of the sensitivity analyses with indeterminates included as test positives.

Evaluation of the tuberculin skin test

The TST was not performed as part of the IDEA study. We did not aim to evaluate the accuracy of the TST and so we did not impose a standard protocol for the conduct of TSTs; each centre followed local trust policy. Five of the nine hospital trusts included in the analyses performed TSTs, but only on a subset of patients, ranging between 20% and 74% of those recruited. We do not know why certain patients were selected for the TST and others were not but the reasons may involve a mix of patient- and clinician-specific factors. Altogether, the TST results were available for only 38% of the study population. The TST was mainly performed at two centres (London North West Healthcare NHS Trust and Imperial College Healthcare NHS Trust), and these centres together accounted for 83% (266/322) of the available TST results. Furthermore, practice varied across the centres in the IDEA study.

Nonetheless, for completeness we estimated the diagnostic accuracy of the TST alone. In addition, because one of our primary objectives was to develop an evidence-based optimal testing algorithm that defines the role of IGRAs in the diagnostic workup of suspected active TB, we considered combinations with T-SPOT.*TB* or QFT-GIT. The results are presented in *Appendix 13*.

The sensitivity of the TST was evaluated at each of the three prespecified thresholds (≥ 5 mm, ≥ 10 mm and the stratified threshold). The findings presented in *Appendix 13* should be interpreted cautiously given the limited number of data and variation in practice between centres. Moreover, the findings may be

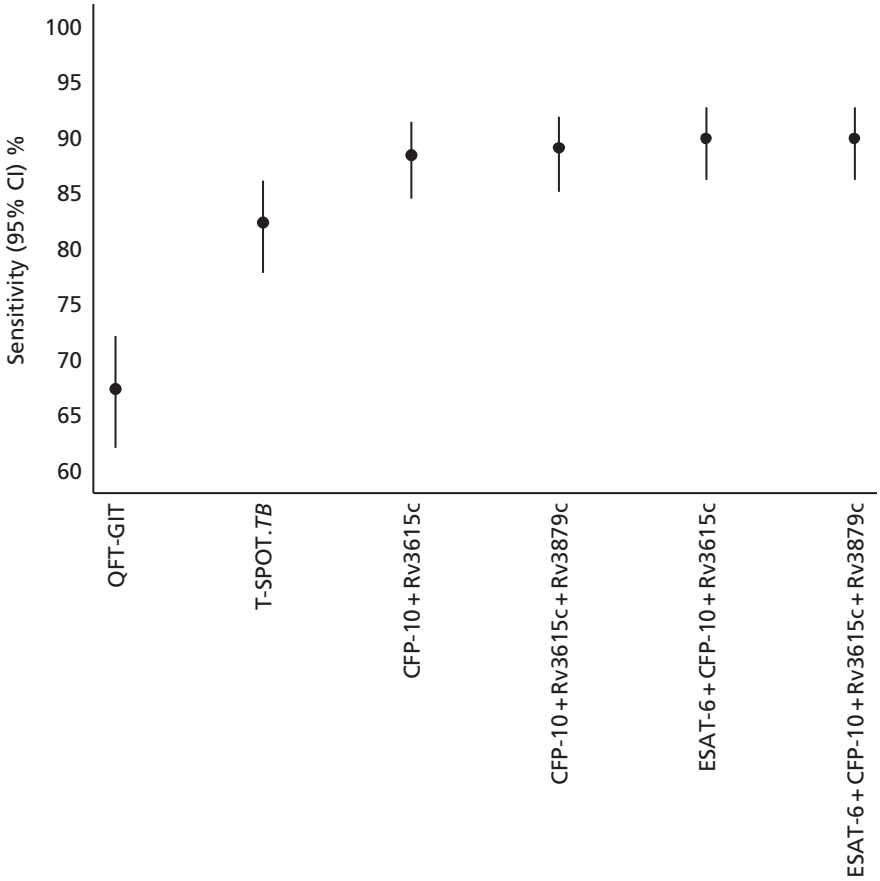

FIGURE 5 Summary of sensitivities of IGRAs. The sensitivities of the six tests are sorted on the plot in increasing order.

subject to bias. The final diagnosis of patients was verified using a composite reference standard applied by a panel of clinicians to anonymised patient clinical data. The panel were blinded to routine and study IGRA results. However, if TSTs were performed, then the results were available thus creating a potential for incorporation bias in the evaluation of the diagnostic accuracy of the TST.

Discussion

This large prospective study of the diagnostic accuracy of T-SPOT.*TB*, QFT-GIT, TSTs and second-generation IGRAs was conducted in secondary care in a population representative of clinical practice. The differences observed between inpatients and outpatients reflect the case mix of acute admissions, in which, for instance, pneumonia-like illnesses and advanced malignancy presentations are more common. However, the final TB diagnosis in these two groups was broadly similar and allows for generalisability of our findings. *Figure 5* summarises the sensitivities of T-SPOT.*TB*, QFT-GIT, TSTs, and combinations of TSTs and IGRAs. *Figure 5* also shows the sensitivities of combinations of novel antigens. T-SPOT.*TB* showed higher sensitivity than QFT-GIT, a relative increase of 22% (95% CI 14% to 31%), thus indicating greater utility as a rule-out test.

Literature searches in PubMed were regularly conducted during the IDEA study to update the SSC on developments in the field. *Appendix 11*, *Table 74*, gives a summary of 31 IGRA studies and one systematic review published between 1 January 2013 and 16 March 2016. To the authors' knowledge, the IDEA study is the largest prospective and most recent comparative evaluation of the accuracy of T-SPOT.*TB* and QFT-GIT for diagnosis of active TB in a high-income setting. None of the studies in *Appendix 11* compared T-SPOT.*TB* and QFT-GIT either prospectively or retrospectively. QFT-GIT was compared with the TST in 14 (45.2%) studies and T-SPOT.*TB* was compared with the TST in four (12.9%) studies; the remaining

13 (41.9%) studies evaluated only T-SPOT.*TB* or QFT-GIT. According to the systematic review of pleural TB published in 2015,[28] there was only one good-quality study out of the 21 studies and two small studies compared QFT-GIT and T-SPOT.*TB*. The authors of the review pooled results across all QuantiFERON and T-SPOT.*TB* assays and so results are not comparable with the findings of the IDEA study.

The rate of indeterminate IGRA results was higher for QFT-GIT (9.6%) than for T-SPOT.*TB* (7.0%). For both IGRAs, most of the indeterminate results were among patients without active TB, with 59% (47/79) and 65% (37/57) occurring for QFT-GIT and T-SPOT.*TB*. *Appendix 12* (see *Table 75*) summarises some key characteristics of patients with indeterminate IGRA results. Indeterminates were excluded from the main analyses and the impact on findings was investigated in sensitivity analyses by including indeterminate IGRA results as test positives, that is, as true positives for those with culture-confirmed or highly probable active TB and as false positives for those without active TB. Based on this rule, an increase in sensitivities, a decrease in specificities and no change in NPVs was expected. Generally, small increases were seen in sensitivity accompanied by decreases in specificity.

Human immunodeficiency virus-positive patients were a key subgroup in which analyses of T-SPOT.*TB* and QFT-GIT were planned. However, the number of patients ($n = 135$) recruited in the main cohort was small. The IDEA study was extended to facilitate recruitment of additional patients. The results for the entire HIV cohort (i.e. including patients recruited during the extension period) are presented in *Chapter 5*. For the main study cohort, the sensitivities of T-SPOT.*TB* and QFT-GIT were lower in those with HIV co-infection than the HIV-negative subgroup. Similarly, the two IGRAs had lower sensitivities in those patients with diabetes mellitus than those without. Specificity was higher in HIV-positive patients than HIV-negative patients but was lower in those with diabetes mellitus than those without diabetes mellitus. Although there appeared to be differences in test performance between subgroups, there was no statistical evidence of an effect of HIV infection status or diabetes mellitus on relative test performance. The findings from these analyses should be taken with caution, as the number of test results in some subgroups was small. Subgroup analyses were not possible for the other two key subgroups: those with end-stage renal failure and those on immune suppressants.

The most promising novel antigen was RV3615c, with a sensitivity that was higher than that of the other five antigens, including CFP-10 and ESAT-6 (the two antigens that constitute T-SPOT.*TB*). Test combinations including Rv3615c showed higher sensitivity than T-SPOT.*TB*, with limited gains in sensitivity when more antigens were added to a combination. Thus, the two-antigen combination of CFP-10 and Rv3615c or the three-antigen combination of ESAT-6, CFP-10 and Rv3615c seem promising. The latter combination had a sensitivity and specificity of 90.5% (95% CI 86.9% to 93.2%) and 70.0% (95% CI 65.4% to 74.2%), compared with 83.2% (95% CI 78.9% to 86.8%) and 75.4% (95% CI 71.1% to 79.3%) for T-SPOT.*TB*. The added value of Rv3615c to T-SPOT.*TB* was a 9% (95% CI 5% to 12%) relative increase in sensitivity at the expense of specificity with a relative decrease of 7% (95% CI 4% to 10%). The incremental gain in sensitivity to 90% is likely to be clinically useful and warrants further investigation.

Chapter 5 Substudy of human immunodeficiency virus-positive participants

Recruitment of human immunodeficiency virus-positive patients

Following the closure of the main study to recruitment, two additional NHS trusts (King's College Healthcare NHS Foundation Trust and Barts Health NHS Trust) were invited to participate in the study to facilitate the recruitment of HIV-positive patients. A total of 263 patients were recruited from 12 NHS trusts between 25 November 2011 and 19 December 2014. The number of patients recruited at each centre is shown in *Table 27*. Over half of the patients were recruited from two trusts: Royal Free London NHS Foundation Trust (29.3%) and Imperial College Healthcare NHS Trust (25.9%).

The flow of patients through the study is shown in *Figure 6*. Of the 263 HIV-positive patients recruited, 201 were included in the analyses. Reasons for exclusion are shown in *Figure 6*. Two patients from Frimley Health NHS Foundation Trust were excluded, as mentioned earlier in *Chapter 3* (see *Recruitment of participants into main study cohort*).

The final diagnosis of four (2.0%) patients was clinically indeterminate (*Figure 6* and *Table 28*). Among the remaining 197 patients, there were 32 (16.2%) cases of active TB and 165 (83.8%) cases of non-active TB. Of the 165 non-active TB cases, 68 (41.2%) were category 4D.

Baseline characteristics of the human immunodeficiency virus-positive cohort

The demographic characteristics of the 201 patients are given in *Table 29*. The median age of patients was 43 years (range 18–79 years) and the majority (67.7%) were male. A substantial number of HIV-positive patients were of black (45.3%) or white (37.8%) ethnicity. A total of 53 countries of birth were represented; countries with at least three participants are shown in *Table 29*. Many patients (97/201, 48.3%) were in paid employment.

TABLE 27 Recruitment of HIV-positive patients by centre

Hospital trust	Patients recruited, *n* (%)
Imperial College Healthcare NHS Trust	68 (25.9)
Heart of England NHS Foundation Trust	7 (2.7)
Chelsea and Westminster Hospital NHS Foundation Trust	49 (18.6)
Royal Free London NHS Foundation Trust	77 (29.3)
St George's University Hospitals NHS Foundation Trust	6 (2.3)
Frimley Health NHS Foundation Trust	2 (0.8)
University Hospitals of Leicester NHS Trust	28 (10.7)
London North West Healthcare NHS Trust	8 (3)
Oxford University Hospitals NHS Trust	6 (2.3)
Sandwell and West Birmingham Hospitals NHS Trust	1 (0.4)
King's College Healthcare NHS Foundation Trust	3 (1.1)
Barts Health NHS Trust	8 (3)
Total	263 (100)

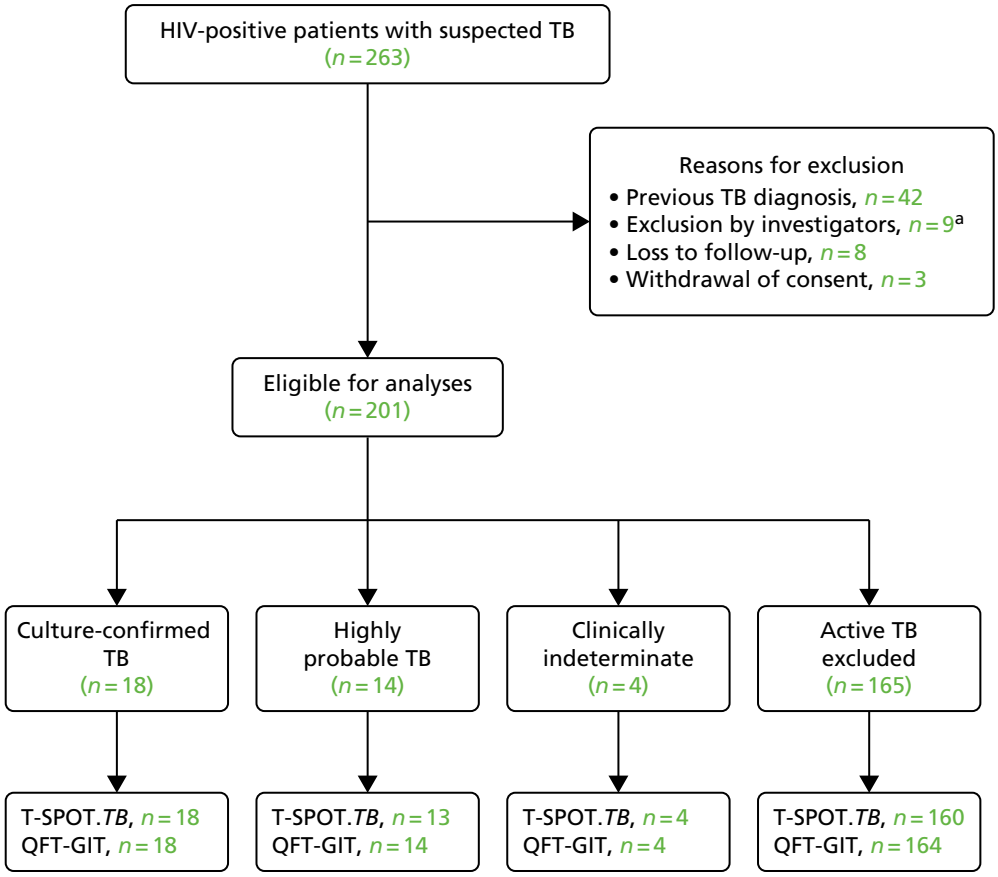

FIGURE 6 Study flow diagram of HIV-positive patients with suspected active TB. The final four boxes show the number of patients with available IGRA results. a, On advice from the Study Steering Committee, and following consultation between the study management group and data management groups, nine patients were excluded from analyses. Two patients recruited from Frimley Health NHS Foundation Trust were excluded because all patients recruited from that site had diagnoses of confirmed or highly probable TB (categories 1 and 2) due an error of implementation of recruitment criteria, i.e. the natural spectrum of patients with *suspected* TB was not being recruited. Two patients were, on review, considered by the expert diagnostic panel to be ineligible (before unblinding IGRA results) on the basis that they were being investigated for TB (due to an incidental abnormal chest X-ray, known contact with active TB, or screening for anti-TNF treatment), but did not present with symptoms or signs suggestive of TB. An additional five patients were excluded due to invalid consent forms.

TABLE 28 Reference standard categories for HIV-positive patients

Diagnostic category	Patients, *n* (%)
1: Culture-confirmed TB	18 (9.0)
2: Highly probable TB	14 (7.0)
3: Clinically indeterminate	4 (2.0)
4: Active TB excluded	
4A	2 (1.0)
4B	2 (1.0)
4C	93 (46.3)
4D	68 (33.8)
Total	201 (100)

TABLE 29 Demographic characteristics of HIV-positive cohort

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Age (years), median (range)	43 (26–62)	40 (24–66)	55 (30–62)	44 (18–79)	43 (18–79)
Male, *n* (%)	13 (72.2)	10 (71.4)	2 (50.0)	111 (67.3)	136 (67.7)
Ethnic origin, *n* (%)					
Asian	1 (5.6)	2 (14.3)	0 (0.0)	3 (1.8)	6 (3.0)
Black	12 (66.7)	7 (50.0)	2 (50.0)	70 (42.4)	91 (45.3)
Hispanic	0 (0.0)	1 (7.1)	0 (0.0)	7 (4.2)	8 (4.0)
Indian subcontinent	3 (16.7)	0 (0.0)	1 (25.0)	8 (4.8)	12 (6.0)
Middle Eastern	1 (5.6)	0 (0.0)	0 (0.0)	2 (1.2)	3 (1.5)
Mixed	0 (0.0)	1 (7.1)	0 (0.0)	4 (2.4)	5 (2.5)
White	1 (5.6)	3 (21.4)	1 (25.0)	71 (43.0)	76 (37.8)
Country of birth, *n* (%)					
UK	2 (11.1)	1 (7.1)	1 (25.0)	57 (34.5)	61 (30.3)
Zimbabwe	1 (5.6)	2 (14.3)	0 (0.0)	16 (9.7)	19 (9.5)
Nigeria	2 (11.1)	0 (0.0)	1 (25.0)	6 (3.6)	9 (4.5)
India	2 (11.1)	0 (0.0)	1 (25.0)	4 (2.4)	7 (3.5)
Uganda	0 (0.0)	0 (0.0)	0 (0.0)	6 (3.6)	6 (3.0)
Ethiopia	2 (11.1)	2 (14.3)	0 (0.0)	1 (0.6)	5 (2.5)
Ireland	0 (0.0)	0 (0.0)	0 (0.0)	5 (3.0)	5 (2.5)
Kenya	0 (0.0)	1 (7.1)	0 (0.0)	4 (2.4)	5 (2.5)
Portugal	0 (0.0)	0 (0.0)	0 (0.0)	5 (3.0)	5 (2.5)
South Africa	0 (0.0)	1 (7.1)	0 (0.0)	4 (2.4)	5 (2.5)
Brazil	0 (0.0)	1 (7.1)	0 (0.0)	3 (1.8)	4 (2.0)
Jamaica	1 (5.6)	1 (7.1)	0 (0.0)	2 (1.2)	4 (2.0)
Poland	0 (0.0)	0 (0.0)	0 (0.0)	4 (2.4)	4 (2.0)
Angola	1 (5.6)	0 (0.0)	0 (0.0)	2 (1.2)	3 (1.5)
The Democratic Republic of the Congo	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.8)	3 (1.5)
Malawi	1 (5.6)	1 (7.1)	0 (0.0)	1 (0.6)	3 (1.5)
Sierra Leone	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.8)	3 (1.5)
Sri Lanka	1 (5.6)	0 (0.0)	0 (0.0)	2 (1.2)	3 (1.5)
Thailand	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.8)	3 (1.5)
Other[a]	5 (27.8)	4 (28.6)	1 (25.0)	32 (19.4)	42 (20.9)
Years in the UK, median (range)	9.6 (0.05–33.5)	10.1 (1.0–37.1)	8.8 (6.2–10.6)	11.9 (0.2–60.3)	11.6 (0.05–60.3)

continued

TABLE 29 Demographic characteristics of HIV-positive cohort (*continued*)

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Profession, *n* (%)[b]					
Paid employment	9 (50.0)	6 (42.9)	1 (25.0)	81 (49.1)	97 (48.3)
Unpaid employment	6 (33.3)	3 (21.4)	3 (75.0)	61 (37.0)	73 (36.3)
Student	1 (5.6)	0	0	8 (4.8)	9 (4.5)
Health-care/laboratory worker	1 (5.6)	3 (21.4)	0	9 (5.5)	13 (6.5)
Social/prison worker	1 (5.6)	1 (7.1)	0	3 (1.8)	5 (2.5)
Sex worker	0	1 (7.1)	0	2 (1.2)	3 (1.5)
Unknown	0	0	0	1 (0.6)	1 (0.5)

a Summarises results from 34 countries with fewer than three patients.
b Some patients had more than one profession.

The clinical characteristics of the patients are presented in *Table 30*. Of the 201 patients, 71 (35.3%) had no other comorbidities. There were 10 (5.0%) patients with pre-existing diabetes mellitus, four (2.0%) with chronic/end-stage renal failure and none was on immunosuppressive therapy. Vitamin D measurement was unknown for most (85.1%) patients (see *Appendix 5*, *Table 55*, for thresholds used to categorise vitamin D status). *Table 30* shows CD4 (cluster of differentiation 4) count grouped into four categories. CD4 count was missing for eight patients. Most of the 193 patients had a CD4 count of $\geq$ 200 cells/μl (46.8%). The median CD4 count was 285 cells/μl (range 0–1228 cells/μl). The distribution of CD4 count in the cohort is shown in *Figure 7*.

TABLE 30 Clinical characteristics of HIV-positive cohort

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Height (m), median (range)	1.7 (1.4–1.8)	1.8 (1.6–1.9)	1.7 (1.6–1.8)	1.7 (1.3–1.9)	1.7 (1.3–1.9)
Weight (kg), median (range)	69 (52–91)	65 (42–116)	62 (57–110)	67 (43–112)	67 (42–116)
BMI (kg/m^2), median (range)	24 (18–48)	22 (16–40)	21 (21–34)	23 (12–39)	23 (12–48)
BCG vaccinated, *n* (%)	15 (83.3)	12 (85.7)	4 (100.0)	136 (82.4)	167 (83.1)
BCG scar visible, *n* (%)					
Yes	13 (72.2)	12 (85.7)	2 (50.0)	114 (69.1)	141 (70.1)
No	0 (0.0)	0 (0.0)	0 (0.0)	17 (10.3)	17 (8.5)
Unsure	2 (11.1)	0 (0.0)	2 (50.0)	17 (10.3)	21 (10.4)
Missing	3 (16.7)	2 (14.3)	0 (0.0)	17 (10.3)	22 (10.9)
Known TB contact, *n* (%)	3 (16.7)	2 (14.3)	1 (25.0)	19 (11.5)	25 (12.4)

TABLE 30 Clinical characteristics of HIV-positive cohort (*continued*)

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Other pre-existing conditions/comorbidities, *n* (%)[a]					
None	9 (50.0)	8 (57.1)	2 (50.0)	52 (31.5)	71 (35.3)
Diabetes	2 (11.1)	0 (0.0)	1 (25.0)	7 (4.2)	10 (5.0)
Hepatitis B	4 (22.2)	1 (7.1)	0 (0.0)	8 (4.8)	13 (6.5)
Hepatitis C	0 (0.0)	0 (0.0)	0 (0.0)	10 (6.1)	10 (5.0)
Chronic/end-stage renal failure	0 (0.0)	0 (0.0)	1 (25.0)	3 (1.8)	4 (2.0)
Cancer	0 (0.0)	0 (0.0)	0 (0.0)	7 (4.2)	7 (3.5)
Organ transplantation	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.5)
Asthma	0 (0.0)	2 (14.3)	0 (0.0)	12 (7.3)	14 (7.0)
Other	8 (44.4)	4 (28.6)	2 (50.0)	101 (61.2)	115 (57.2)
Medication, *n* (%)[a]					
None/missing	4 (22.2)	7 (50.0)	0 (0.0)	46 (27.9)	57 (28.4)
Chemotherapy	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.5)
Corticosteroids ≥ 15 mg/day	1 (5.6)	1 (7.1)	0 (0.0)	14 (8.5)	16 (8.0)
Corticosteroids < 15 mg/day	0 (0.0)	1 (7.1)	0 (0.0)	5 (3.0)	6 (3.0)
Ciclosporin, tacrolimus or everolimus	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.5)
Methotrexate	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.5)
Other	18 (100.0)	14 (100.0)	4 (100.0)	164 (99.4)	200 (99.5)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.5)
Vitamin D deficiency, *n* (%)					
Deficient	3 (16.7)	1 (7.1)	1 (25.0)	9 (5.5)	14 (7.0)
Insufficient	1 (5.6)	2 (14.3)	0 (0.0)	2 (1.2)	5 (2.5)
Normal	0 (0.0)	1 (7.1)	2 (50.0)	7 (4.2)	10 (5.0)
Not known	14 (77.8)	10 (71.4)	1 (25.0)	147 (89.1)	172 (85.6)
CD4 count, *n* (%)					
< 50 cells/μl	1 (5.6)	4 (28.6)	0 (0.0)	48 (29.1)	53 (26.4)
≥ 50 cells/μl and < 100 cells/μl	2 (5.6)	0 (0.0)	0 (0.0)	15 (9.1)	16 (8.0)
≥ 100 cells/μl and < 200 cells/μl	5 (27.8)	3 (21.4)	2 (50.0)	20 (12.1)	30 (14.9)
≥ 200 cells/μl	8 (44.4)	6 (42.9)	2 (50.0)	78 (47.3)	94 (46.8)
Missing	3 (16.7)	1 (7.14)	0 (0.0)	4 (2.4)	8 (4.0)
CD4 count, cells/μl (range)	293 (14–670)	267 (0–669)	370 (183–800)	283 (0–1228)	285 (0–1228)

BMI, body mass index.

a Some patients had multiple comorbidities and so the numbers do not add up to 201. The percentages are column percentages for each row. The same applies to medication.

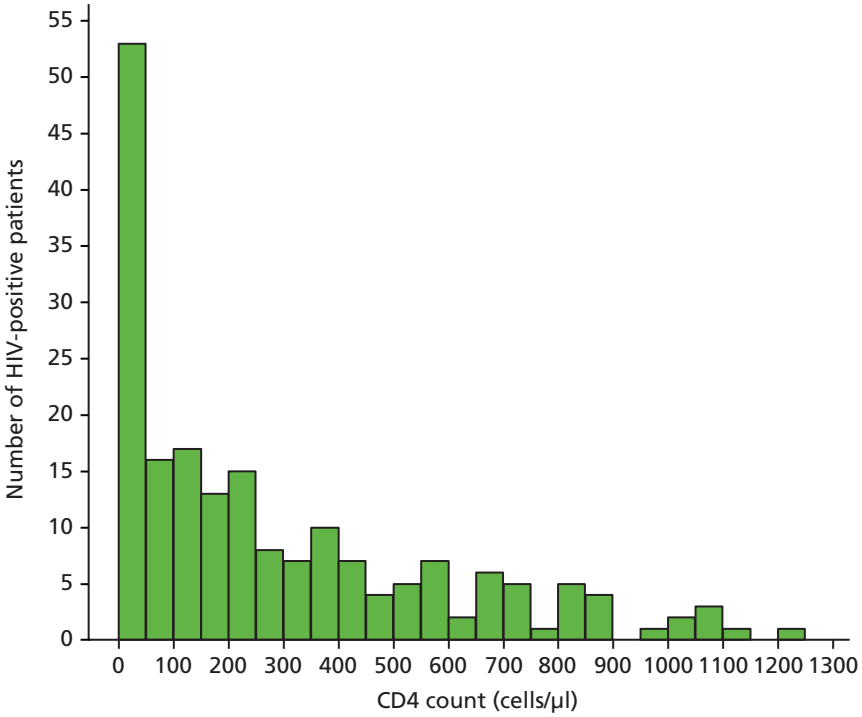

FIGURE 7 Distribution of CD4 counts in HIV-positive substudy cohort.

Table 31 gives the social history of patients. Smoking history was missing for one patient; almost half (48.8%) of the patients had never smoked, whereas the remaining were current (31.3%) or ex-smokers (19.4%). Fourteen (7.0%) patients had a history of alcohol misuse and the history of recreational drug use was unknown for many (64.7%) patients. Few patients (6.0%) had a history of homelessness and 7.5% had a history of imprisonment.

TABLE 31 Social history of the HIV-positive cohort

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Smoking history, *n* (%)					
Never smoked	13 (72.2)	7 (50.0)	3 (75.0)	75 (45.5)	98 (48.8)
Ex-smoker	3 (16.7)	2 (14.3)	0 (0.0)	34 (20.6)	39 (19.4)
Current smoker	2 (11.1)	5 (35.7)	1 (25.0)	55 (33.3)	63 (31.3)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.5)
Pack years if current smoker, *n* (%)					
≤ 10 years	0 (0.0)	1 (20.0)	0 (0.0)	9 (16.4)	10 (15.9)
11–20 years	0 (0.0)	0 (0.0)	0 (0.0)	2 (3.6)	2 (3.2)
21–50 years	0 (0.0)	0 (0.0)	0 (0.0)	3 (5.5)	3 (4.8)
Unknown	2 (100.0)	4 (80.0)	1 (100.0)	41 (74.5)	48 (76.2)

TABLE 31 Social history of the HIV-positive cohort (*continued*)

Characteristic	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
History of alcohol use, *n* (%)					
Non-drinker	10 (55.6)	6 (42.9)	3 (75.0)	67 (40.6)	86 (42.8)
Ex-drinker	1 (5.6)	1 (7.1)	0 (0.0)	17 (10.3)	19 (9.5)
Current drinker	7 (38.9)	7 (50.0)	1 (25.0)	79 (47.9)	94 (46.8)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.2)	2 (1.0)
Units/week if current drinker, median (range)	5 (0–75)	8 (0–19)	10 (10–10)	5 (0–126)	5 (0–126)
History of alcohol misuse, *n* (%)	1 (5.6)	0 (0.0)	0 (0.0)	13 (7.9)	14 (7.0)
History of recreational drug use, *n* (%)					
Non-user	6 (33.3)	2 (14.3)	2 (50.0)	31 (18.8)	41 (20.4)
Ex-user	1 (5.6)	0 (0.0)	0 (0.0)	7 (4.2)	8 (4.0)
Current user	0 (0.0)	3 (21.4)	0 (0.0)	19 (11.5)	22 (11.0)
Unknown	11 (61.1)	9 (64.3)	2 (50.0)	108 (65.5)	130 (64.7)
History of homelessness, *n* (%)	1 (5.6)	1 (7.1)	0 (0.0)	10 (6.1)	12 (6.0)
Years homeless if currently or previously homeless, median (range)	0 (0.0)	12	–	5 (0–12)	5 (0–12)
History of imprisonment, *n* (%)	0 (0.0)	1 (7.1)	0 (0.0)	14 (8.5)	15 (7.5)

The frequency of symptoms is shown in *Table 32* based on data from 197 patients. The four other patients were recruited on the basis of abnormal clinical signs rather than symptoms. The main symptoms were cough, fever, night sweats, weight loss and lethargy. Patients generally presented with multiple symptoms; the median number of symptoms was three (range 1–10). Cough, as a symptom, was often present (68.5%) in patients.

TABLE 32 Symptoms at presentation for the HIV-positive cohort

Symptom[a]	Dosanjh category				Total
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Cough, *n* (%)	13 (72.2)	4 (30.8)	3 (75.0)	115 (71.0)	135 (68.5)
Fever, *n* (%)	13 (72.2)	8 (61.5)	3 (75.0)	85 (52.5)	109 (55.3)
Night sweats, *n* (%)	11 (61.1)	9 (69.2)	1 (25.0)	86 (53.1)	107 (54.3)
Weight loss, *n* (%)	12 (66.7)	11 (84.6)	2 (50.0)	89 (54.9)	114 (57.9)
Haemoptysis, *n* (%)	2 (11.1)	1 (7.7)	0	18 (11.1)	21 (10.7)
Lethargy, *n* (%)	7 (38.9)	8 (61.5)	1 (25.0)	81 (50.0)	97 (49.2)
Other, *n* (%)	11 (61.1)	5 (38.5)	2 (50.0)	79 (48.8)	97 (49.2)
Number of symptoms, median (range)	4 (1–8)	4 (3–6)	3.5 (1–5)	3 (1–9)	3 (1–10)

a Based on 197 patients.

Final diagnosis in human immunodeficiency virus-positive patients

The diagnostic tests performed during the diagnostic workup of patients are shown in *Table 33*. Chest radiography and culture were the most common tests performed (with each performed in 90.6% of patients). The number of T-SPOT.*TB*, QFT-GIT and TST tests performed as part of routine care, at each centre, is shown in *Appendix 14* (see *Table 78*). These routine IGRA results were not used in the final diagnosis of patients in the IDEA study. TSTs were performed in only 12 patients at 3 of the 11 centres; TST results were available for all 12 patients.

The final diagnoses of active TB patients are summarised in *Table 34*. Of the 32 patients with active TB, 19 (59.4%) had smear-negative TB. A total of 13 patients (40.6%) had pulmonary TB, 15 (46.9%) patients had extrapulmonary TB, and the remaining four (12.5%) patients had both forms of TB. The most common sites of infection were the lungs (53.1%) and lymph nodes (31.1%). Of the 31 culture tests performed, 20 (64.5%) had no drug resistance. *Table 35* shows the final diagnosis of non-active TB patients. A patient may have multiple conditions, but pneumonia was the most frequent diagnosis (40.0%).

Test results for human immunodeficiency virus-positive patients

Of the 201 patients, 194 (96.5%) had results for both T-SPOT.*TB* and QFT-GIT; reasons for missing T-SPOT.*TB* and QFT-GIT results are shown in *Table 36*. The cross-classified results of the two tests are given in *Appendix 14*, *Table 79*. *Table 37* shows the results for T-SPOT.*TB*, QFT-GIT and the TST according to the four diagnostic categories. *Table 38* shows the cross-classification of the two tests for those with (categories 1 and 2) and without active TB (category 4).

TABLE 33 Diagnostic tests performed in the diagnostic workup of active TB in HIV-positive patients

Test	Dosanjh category, *n* (%)				Total, *n* (%)
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
BAL investigation	5 (27.8)	4 (28.6)	3 (75.0)	68 (41.2)	80 (39.8)
CXR	14 (77.8)	14 (100.0)	3 (75.0)	151 (91.5)	182 (90.5)
CSF investigation	3 (16.7)	6 (42.9)	1 (25.0)	17 (10.3)	27 (13.4)
CT	14 (77.8)	12 (85.7)	3 (75.0)	122 (73.9)	151 (75.1)
Culture	18 (100.00)	13 (92.9)	3 (75.0)	148 (89.7)	182 (90.5)
Histology or biopsy	3 (16.7)	5 (35.7)	0 (0.0)	49 (29.7)	57 (28.4)
MRI	4 (22.2)	9 (64.3)	1 (25.0)	28 (17.0)	42 (20.9)
PCR	10 (55.6)	5 (35.7)	1 (25.0)	42 (25.5)	58 (28.9)
Smear test	14 (77.8)	14 (100.0)	2 (50.0)	139 (84.2)	169 (84.1)

CSF, cerebrospinal fluid; CXR, chest radiography.

TABLE 34 Final diagnosis of TB in HIV-positive patients

Characteristic	Category of TB, *n* (%)		Total, *n* (%)
	1	2	
All TB	18 (56.3)	14 (43.7)	32 (100)
Smear-positive TB	8 (44.4)	1 (7.1)	9 (28.1)
Smear-negative TB	6 (33.3)	13 (92.9)	19 (59.4)
Pulmonary TB	10 (55.6)	3 (21.4)	13 (40.6)
Extrapulmonary TB	5 (27.8)	10 (71.4)	15 (46.9)
Pulmonary/extrapulmonary TB	3 (16.7)	1 (7.1)	4 (12.5)
Site of infection[a]			
Brain	0 (0.0)	2 (14.3)	2 (6.3)
Central nervous system	0 (0.0)	1 (7.1)	1 (3.1)
Lung	13 (72.2)	4 (28.6)	17 (53.1)
Lymph node	4 (22.2)	6 (42.9)	10 (31.3)
Miliary TB (disseminated)	3 (16.7)	0 (0.0)	3 (9.4)
Pericardium	0 (0.0)	1 (7.1)	1 (3.1)
Pleura	1 (5.6)	0 (0.0)	1 (3.1)
Spine	1 (5.6)	1 (7.1)	2 (6.3)
Multidrug resistance[b]			
None	18 (100)	2 (15.4)	20 (64.5)
Not tested	0 (0.0)	11 (84.6)	11 (35.5)

a There were multiple TB infection sites for some patient and so the percentages do not add up to 100%.
b Available where culture tests were performed during diagnostic workup ($n = 31$).

TABLE 35 Final diagnosis of non-TB in HIV-positive patients

Diagnosis	Number of patients (%)
Cancer	17 (10.3)
Chest infection	2 (1.2)
LRTI	13 (7.9)
LTBI – treatment indicated	1 (0.6)
Pneumonia	66 (40.0)
Sarcoidosis	3 (1.8)
Other	75 (45.5)

LRTI, lower respiratory tract infection; LTBI, latent TB infection.

TABLE 36 Reasons for missing IGRA results for HIV-positive patients

Reason	Test, *n*	
	QFT-GIT	T-SPOT.*TB*
No sample could be taken	1	–
Sample destroyed for laboratory reasons	–	2
Sample unsuitable for testing	–	4
Total	1	6

TABLE 37 Results for T-SPOT.*TB* and QFT-GIT, by final diagnosis, in the cohort of HIV-positive patients

Index test	Dosanjh category								Total
	1	2	3	4A	4B	4C	4D	4A–D	
***T-SPOT.*TB**									
Positive	10	6	1	0	1	8	3	12	29
Negative	5	3	2	1	1	57	42	101	111
Borderline	0	1	1	0	0	6	2	8	10
Indeterminate	3	3	0	1	0	19	19	39	45
Missing	0	1	0	0	0	3	2	5	6
Total	18	14	4	2	2	93	68	165	201
Median SFCs ESAT-6 (range)	17 (0–462)	8 (0–395)	1 (0–3)	0	0	0 (0–66)	0 (0–74)	0 (0–74)	0 (0–462)
Median SFCs CFP-10 (range)	6 (0–136)	0 (0–315)	3 (0–20)	0	8 (0–16)	0 (0–67)	0 (0–56)	0 (0–67)	0 (0–315)
QFT-GIT									
Positive	11	6	1	0	1	8	2	11	29
Negative	7	6	2	1	1	61	54	117	132
Indeterminate	0	2	1	1	0	24	11	36	39
Missing	0	0	0	0	0	0	1	1	1
Total	18	14	4	2	2	93	68	165	201
Median IFN-γ levels (range)	0.76 (0–3.78)	0.03 (0–10)	0.01 (0–0.54)	0.04 (0–0.07)	1.33 (0–2.65)	0 (0–7.3)	0 (0–1.56)	0 (0–7.3)	0.01 (0–10)

Diagnostic accuracy of T-SPOT.*TB* and QuantiFERON GOLD In-Tube in human immunodeficiency virus-positive cohort

The proportion of indeterminate test results was 23.1% (45/195) for T-SPOT.*TB* and 19.5% (39/200) for QFT-GIT. The difference between the two proportions was 3.6% (95% CI –4.5% to 11.6%; $p = 0.4$). Sensitivity was 68.0% (95% CI 48.4% to 82.8%) for T-SPOT.*TB* and 56.7% (95% CI 39.2% to 72.6%) for QFT-GIT (*Table 39*). The specificities were 83.5% (95% CI 75.8% to 89.0%) and 91.4% (95% CI 85.3% to 95.1%). The PPVs for T-SPOT.*TB* and QFT-GIT were 46.0% (95% CI 31.0% to 61.6%) and 60.7% (95% CI 42.4% to 76.4%), respectively, and the NPVs were 92.7% (95% CI 86.2% to 96.2%) and 90.0% (95% CI 83.6% to 94.1%), respectively.

TABLE 38 Cross-tabulation of T-SPOT.*TB* and QFT-GIT against final diagnosis in the cohort of HIV-positive patients

T-SPOT.*TB*		TB positive (categories 1 and 2)						TB negative (category 4)					
		Positive	Negative	Borderline	Indeterminate	Missing	Total	Positive	Negative	Borderline	Indeterminate	Missing	Total
QFT-GIT	Positive	14	1	0	2	0	17	4	3	1	3	0	11
	Negative	1	6	1	4	1	13	5	75	7	26	4	117
	Indeterminate	1	1	0	0	0	2	3	22	0	10	1	36
	Missing	0	0	0	0	0	0	0	1	0	0	0	1
	Total	16	8	1	6	1	32	12	101	8	39	5	165

TABLE 39 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT for diagnosis of active TB in the HIV-positive cohort

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	17/25	68.0 (48.4 to 82.8)	17/30	56.7 (39.2 to 72.6)
Culture-positive TB	10/15	66.7 (41.7 to 84.8)	11/18	61.1 (38.6 to 79.7)
Culture-negative TB	7/10	70.0 (39.7 to 89.2)	6/11	54.5 (28.0 to 78.7)
Smear-positive TB	4/8	50.0 (21.5 to 78.5)	6/9	66.7 (35.4 to 87.9)
Smear-negative TB	9/13	69.2 (42.4 to 87.3)	8/17	47.1 (26.2 to 69.0)
Pulmonary TB	7/10	70.0 (39.7 to 89.2)	9/13	69.2 (42.4 to 87.3)
Extrapulmonary TB	8/11	72.7 (43.4 to 90.3)	6/14	42.9 (21.4 to 67.4)
Specificity for a diagnosis of active TB				
Active TB excluded	101/121	83.5 (75.8 to 89.0)	117/128	91.4 (85.3 to 95.1)
Active TB excluded, TST negative, no risk factors for LTBI	42/47	89.4 (77.4 to 95.4)	54/56	96.4 (87.9 to 99.0)
Predictive value				
PPV	17/37	46.0 (31.0 to 61.6)	17/28	60.7 (42.4 to 76.4)
NPV	101/109	92.7 (86.2 to 96.2)	117/130	90.0 (83.6 to 94.1)
Likelihood ratios				
Positive likelihood ratio	–	4.11 (2.54 to 6.66)	–	6.59 (3.46 to 12.6)
Negative likelihood ratio	–	0.38 (0.22 to 0.68)	–	0.47 (0.31 to 0.72)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were excluded from these analyses. Sensitivity analyses with indeterminates as test positives are presented in *Appendix 14*, *Table 80*.

Using only culture-confirmed active TB cases, the sensitivity of T-SPOT.*TB* decreased slightly to 66.7% (95% CI 41.7% to 84.8%), whereas that of QFT-GIT increased to 61.1% (95% CI 38.6% to 79.7%). When the analyses were restricted to category 4D patients without active TB, specificities were higher than using all patients without active TB (see *Table 39*). In sensitivity analyses with indeterminate test results included as test positives, the sensitivity of T-SPOT.*TB* was 74.2% (95% CI 56.8% to 86.3%) and 59.4% (95% CI 42.3% to 74.5%) for QFT-GIT. The specificity of T-SPOT.*TB* was 63.1% (95% CI 55.4% to 70.2%) and 71.3% (95% CI 64.0% to 77.7%) for QFT-GIT. See *Appendix 14*, *Table 80*, for full results.

The diagnostic performance of the two IGRAs is shown stratified by CD4 count in *Table 40*. The estimates within each stratum give results comparable to the entire cohort in terms of the higher sensitivity of T-SPOT.*TB* and higher specificity of QFT-GIT. However, the estimates are based on very small numbers of active TB and non-active TB cases and are presented solely for illustrative purposes.

TABLE 40 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT stratified by CD4 count in the HIV-positive substudy cohort

	Test			
	T-SPOT.*TB*		QFT-GIT	
CD4 count	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
< 50 cells/μl	1/3	33.3 (6.15 to 79.2)	1/4	25.0 (4.56 to 70.0)
≥ 50 cells/μl and < 100 cells/μl	0/1	0.00 (0.00 to 79.4)	0/1	0.00 (0.00 to 79.4)
≥ 100 cells/μl and < 200 cells/μl	5/6	83.3 (43.7 to 97.0)	4/7	57.1 (25.1 to 84.2)
≥ 200 cells/μl	10/12	83.3 (55.2 to 95.3)	11/14	78.6 (52.4 to 92.4)
Specificity for a diagnosis of active TB				
< 50 cells/μl	26/29	89.7 (73.6 to 96.4)	26/29	89.7 (73.6 to 96.4)
≥ 50 cells/μl and < 100 cells/μl	11/13	84.6 (57.8 to 95.7)	11/12	91.7 (64.6 to 98.5)
≥ 100 cells/μl and < 200 cells/μl	10/13	76.9 (49.7 to 91.8)	14/15	93.3 (70.2 to 98.8)
≥ 200 cells/μl	53/64	83.1 (72.2 to 90.3)	63/69	91.3 (82.3 to 96.0)

Note
Sensitivity and specificity are presented as percentages.

Comparison of diagnostic accuracy of T-SPOT.*TB* and QuantiFERON GOLD In-Tube in human immunodeficiency virus-positive cohort

Excluding indeterminate IGRA results, there were 146 T-SPOT.*TB* results and 158 QFT-GIT results. The sensitivity of T-SPOT.*TB* was higher than that of QFT-GIT, with a relative sensitivity of 1.12 (95% CI 0.87 to 1.44). There was no statistical evidence of a difference ($p = 0.4$). In contrast, the specificity of T-SPOT.*TB* was significantly lower than that of QFT-GIT, with a relative specificity of 0.91 (95% CI 0.84 to 0.99) (*Table 41*). When indeterminate IGRA results were included as test positives in a sensitivity analysis, the analysis included 191 T-SPOT.*TB* results and 196 QFT-GIT results. Unlike the primary analysis, there was no statistical evidence of a difference in specificity (see *Appendix 14*, *Table 81*).

TABLE 41 Comparison of diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in HIV-positive patients

Test	Number of test results[a]	Sensitivity (95% CI)	Number of test results[b]	Specificity (95% CI)
T-SPOT.*TB*	25	62.8 (44.1 to 78.3)	121	83.4 (75.7 to 88.9)
QFT-GIT	30	56.1 (38.3 to 72.4)	128	91.7 (85.4 to 95.4)
Ratio (95% CI);[c] *p*-value	–	1.12 (0.87 to 1.44); 0.4	–	0.91 (0.84 to 0.99); 0.02

a Number of test results among those patients with active TB.
b Number of test results among those without active TB.
c The ratio of the sensitivity (or specificity) of T-SPOT.*TB* to that of QFT-GIT. The natural outputs from GEE models are odds ratios. Ratios of sensitivities (relative sensitivity) and ratios of specificities (relative specificity) were computed post estimation of the models. Confidence intervals were obtained using the delta method.

Note
Sensitivities and specificities are presented as percentages.

Discussion

Of the 263 patients recruited, 201 patients were included in the analyses. This is well below the target sample size of 390. Although T-SPOT.*TB* showed higher sensitivity than QFT-GIT, a relative increase of 12%, there was no statistical evidence of a difference in sensitivity. The sensitivity of QFT-GIT in the IDEA study was lower than that in other recent studies[29–31] (see *Appendix 11*, *Table 74*).

The indeterminate rate was lower for QFT-GIT (19.5%) than for T-SPOT.*TB* (23.1%). The indeterminate results were mainly in those without active TB, with 92.3% (36/39) for QFT-GIT and 86.7% (39/45) for T-SPOT.*TB*. The impact of indeterminate results was explored in sensitivity analyses by including indeterminate results as test positives. There was a small increase in the sensitivity of QFT-GIT but a large increase in the sensitivity of T-SPOT.*TB*. This is because of the higher indeterminate rate for T-SPOT.*TB* (18.2%) among category 1 and 2 active TB cases than for QFT-GIT (5.9%).

Chapter 6 Economic evaluation methods

In this chapter we present methods for assessment of the cost-effectiveness of IGRAs as rule-out tests for active TB. That is, we consider using an IGRA as an initial test, with a negative result indicating that a patient does not have TB, thus accelerating diagnosis of the actual cause of disease in such patients. The use of QFT-GIT and T-SPOT.*TB* was compared against current practice, as determined by analysis of patient records. We considered which diagnostic tests were performed, their costs and the time taken between decision points involving each test. The time taken to diagnose or rule out TB is a key consideration. Our report adheres to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.[32]

Decision tree model

We developed a decision tree model to calculate the incremental costs and incremental health utilities (quality-adjusted life-year; QALYs) of changing from current practice to using an IGRA as an initial rule-out test. Current practice was determined by the analysis of patient records. The model structure representing current practice is shown in *Figure 7*. Adding a rule-out test to the diagnostic pathway introduces additional delay in the diagnosis of active TB in those patients who have the disease, as it introduces an additional step in the pathway. Patients who were not initially diagnosed with active TB have a follow-up consultation after approximately 2 months; those who had a false-negative rule-out test result, that is, they had TB incorrectly ruled out, can have TB identified at this point. The final diagnostic outcomes were the four categories described in Dosanjh *et al.*,[3] herein referred to as 'Dosanjh categories' (see *Appendix 2*, *Table 52*). The health economic analysis was undertaken from a NHS perspective. No discounting was required, as the diagnostic process occurs over a relatively short time period.

The model contained two levels of uncertainty.

1. Individual-level uncertainty: patient records revealed variation in the number and type of tests used for TB diagnosis and time to diagnosis.
2. Parameter uncertainty: uncertainty in the costs of tests and procedures, and the sensitivity and specificity of IGRAs.

A (balanced) bootstrap sample of TB and non-TB patients was created for each simulation of the decision tree, which retained the subsample sizes. Individual patient costs and time to diagnosis were jointly sampled to preserve the dependency between them. A Bayesian forward sampling approach generated values from all of these distributions to simulate a particular model outcome. This was then repeated in a Monte Carlo framework to obtain a sample of several thousand model runs that encapsulated the first-order (patient variability) and second-order (parameter) uncertainty in the model outputs. *Table 42* summarises the model parameters. The model was implemented in the statistical programming language R (version R-3.3.0; R Foundation for Statistical Computing, Vienna, Austria).

Distributional formulation of individual-level/sample uncertainty

The number of individuals who enter the diagnostic pathway was defined as n. Of these, we defined n_+ as active TB cases and the remainder as non-active TB cases. We defined the probability of a given patient being an active TB case as n_+/n and so the probability of being a non-active TB case is $1-(n_+/n)$. This can therefore be considered a draw from a Bernoulli distribution and so for the total sample the binomial distribution gives the number of TB cases in a sample population as:

$$\begin{aligned} X_{TB} &\sim Bin(^{n_+}/_n, n) \\ X_{\sim TB} &= n - X_{TB}, \end{aligned} \tag{1}$$

TABLE 42 Decision tree model parameters and values from sensitivity analyses

Parameter	Symbol	Main model values	Sensitivity range
Rule-out test unit cost (£)	$C_{ruleout}$	T-SPOT.*TB* or QFT-GIT	1–200
Follow-up time for those not diagnosed with TB (days)	T_{FN}	Direct estimate from the IDEA study clinical data set	54–127[a]
Cohort active TB prevalence	n_+/n	Direct estimate from the IDEA study clinical data set	0.1–0.5
Rule-out test time (days)	$T_{ruleout}$	Uniform(2,14)	
Quality-of-life detriment for active TB symptoms	D_{QoL}	Triangle(0.11,0.21,0.31)[33,34]	
Current time to diagnosis by TB status	$\bar{t}_i$	Direct estimate from the IDEA study clinical data set	From the IDEA study data set
Current combined cost of diagnosis by TB status	$\bar{c}_i$	Direct estimate from the IDEA study clinical data set	From the IDEA study data set

a 5% and 95% quantiles of the time to follow-up visit observed in the IDEA study clinical data.

that is,

$$p(X_{TB}|{}^{n_+}\!/_n) = \binom{n}{X_{TB}} ({}^{n_+}\!/_n)^{X_{TB}} (1 - {}^{n_+}\!/_n)^{(n - X_{TB})}. \tag{2}$$

By the same principle, patients are then randomly split between those who are ruled out and those who are not ruled out from the active TB and non-active TB subgroups:

$$\begin{aligned} X_+^{TB} &\sim Bin(p_+, X_{TB}) \\ X_-^{TB} &= X_{TB} - X_+^{TB} \\ X_+^{\sim TB} &\sim Bin(1 - p_-, X_{\sim TB}) \\ X_-^{\sim TB} &= X_{\sim TB} - X_+^{\sim TB}. \end{aligned} \tag{3}$$

This process results in a final random subdivision of the sample population into one of the end states.

Health economics outcomes

Each of the terminal nodes (outcomes) of the decision tree has an associated cost and health utility (measured in QALYs). Incremental QALY differences are due to differences in the time taken to start appropriate treatment, leading to differences in morbidity. For simplicity of notation, we shall denote $1 - D_{QoL}$ by q. For the patient cohort these were defined as:

$$\begin{aligned} C_{current} &= \bar{c} \\ C_{novel} &= ({}^{n_+}\!/_n)(\bar{c} + C_{ruleout}) + ({}^{n_-}\!/_n)((1 - p_-)\bar{c} + C_{ruleout}) \\ E_{current} &= q\bar{t} \\ E_{novel} &= ({}^{n_+}\!/_n)q(\bar{t} + T_{ruleout} + (1 - p_+)T_{FN}) + ({}^{n_-}\!/_n)q((1 - p_-)\bar{t} + T_{ruleout}). \end{aligned} \tag{4}$$

Estimation of costs used in the model

The costs and distributions used in the sensitivity analyses are summarised in *Tables 43–45*, using 2014/15 prices. When necessary, costs were inflated from previous years using the Hospital and Community Health Service pay and price index.[37]

TABLE 43 Health-care professional consultation visit monetary costs incurred

Consultation type	Cost (£)	Distribution	Sources
First visit: respiratory medicine, multiprofessional	241	Gamma(53.3,4.52)	*National Tariff Payment System 2014/15*. Annex 5A[35]
	167 (SE 33)		Hughes *et al.*, 2012[36]
Follow-up visit: respiratory medicine, multiprofessional	143	Gamma(18.78,7.62)	*National Tariff Payment System 2014/15*. Annex 5A[35]
	167 (SE 33)		Hughes *et al.*, 2012[36]

SE, standard error.

TABLE 44 Test and sampling procedure costs for cost-effectiveness calculations

Test	Unit cost (£) [min., max.]	Distribution	Sources
Culture	22.29 (SE 2.23)	Gamma(100,0.22)	Drobniewski *et al.*, 2015[38]
Sputum smear microscopy	7	Gamma(106,0.07)	NICE's *Tuberculosis: Prevention, Diagnosis, Management and Service Organisation. NICE Guideline 33*, 2016[39]
	1.56 (SE 0.68)		Hughes *et al.*, 2012[36]
TST	17.48		Auguste *et al.*, 2016[40,41]
	16 [8, 32]	Uniform(8,36)	NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]
T-SPOT.*TB*	59.57		Sutcliffe, 2016[41]
	55 [45, 99]	Uniform(50,106)	NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]
QFT-GIT	58 [29, 87]	Uniform(29,87)	Pareek *et al.*, 2013[43]
CXR	35		NICE's *Tuberculosis: Prevention, Diagnosis, Management and Service Organisation. NICE Guideline 33*, 2016[39]
	28 [19, 34]	Uniform(23,43)	NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]
Bronchoalveolar lavage	23.24	–	St Mary's R&D office
	[11.62, 46.48]	Uniform (11.62, 46.48)	Proportions from NICE's *Clinical Diagnosis and Management of Tuberculosis and Measures for its Prevention and Control. NICE Guideline 33*, 2006,[44] Pareek *et al.*, 2013[43] and NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]

continued

TABLE 44 Test and sampling procedure costs for cost-effectiveness calculations (*continued*)

Test	Unit cost (£) [min., max.]	Distribution	Sources
EBUS	2634	–	St Mary's R&D office
Bronchoscopy procedure	612	–	St Mary's R&D office
	[306, 1224]	Uniform(306,1224)	Proportions from NICE's *Clinical Diagnosis and Management of Tuberculosis and Measures for its Prevention and Control. NICE Guideline 33*, 2006,[44] Pareek *et al.*, 2013[43] and NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]
Histology from biopsy	25	–	St Mary's R&D office
	[12.5, 50]	Uniform(12.5,50)	Proportions from NICE's *Clinical Diagnosis and Management of Tuberculosis and Measures for its Prevention and Control. NICE Guideline 33*, 2006,[44] Pareek *et al.*, 2013[43] and NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]
Needle aspirate	90.21	–	St Mary's R&D office
	[45.1, 180.42]	Uniform(45.1,180.42)	Proportions from NICE's *Clinical Diagnosis and Management of Tuberculosis and Measures for its Prevention and Control. NICE Guideline 33*, 2006,[44] Pareek *et al.*, 2013[43] and NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]
PCR	202.45	–	St Mary's R&D office
	[101.2, 404.9]	Uniform(101.2,404.9)	Proportions from NICE's *Clinical Diagnosis and Management of Tuberculosis and Measures for its Prevention and Control. NICE Guideline 33*, 2006,[44] Pareek *et al.*, 2013[43] and NICE's *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*, 2010[42]

CXR, chest radiography; EBUS, endobronchial ultrasound; max., maximum; min., minimum; SE, standard error.
Note
Uncertainty distributions for each test are estimated either for uniform or gamma distributions depending on the form of the data available.

When uncertainty bounds were not available in the recent sources used, uncertainty ranges were informed by previous studies. Uniform distributions were used when the upper and lower limits were available and gamma distributions were used when the standard error of the average cost was available.

When the lower, upper and mean values were available, the range between the lower and upper bound was defined by the proportional decrease and increase from the mean, respectively. These proportions were then used with an alternative, more appropriate, mean to calculate associated upper and lower

TABLE 45 Active TB treatment costs for the first 60 days

Drug	Dosage (mg/day)[a]	Dosage by patient weight (mg/kg/day)[a]	Batch cost[b] (£)	Quantity per batch (capsule or tablet)[b]	Dosage per capsule or tablet[b]	60-day total cost (£)
Rifampicin	600	–	48.00	100	300	57.60
Isoniazid	300	–	19.24	56	50	123.69
Pyrazinamide	2000	–	38.34	30	500	306.72
Ethambutol hydrochloride	–	15	42.74	56	400	116.74

a Data taken from the *British National Formulary*.[45]
b Data taken from NICE's *Tuberculosis: Prevention, Diagnosis, Management and Service Organisation. NICE Guideline 33* (2016).[39]

limits. For example, a lower bound of half and an upper bound of twice the point estimate values were used.[42,43,46] Skewed distributions were represented using a gamma distribution.

As the end point of the analysis is diagnosis of TB or ruling out TB, treatment costs after final diagnosis are out of scope. However, when a patient was started on TB treatment and then a lack of response to that treatment informed a decision that the patient did not in fact have TB, the cost of this treatment was included, as it is part of the cost of ruling out TB for those patients.

Tests used in the diagnostic pathways are either specific to the diagnosis of active TB or, in the case of imaging tests [CT, MRI and positron emission tomography (PET)], can aid the diagnosis of multiple diseases. For patients who do not have TB, these imaging tests will be used to inform the ultimate diagnosis.

The treatment costs for those patients who have TB ruled out after starting treatment are given in *Table 45*. Following the NICE guidelines for active TB management,[39] it was assumed that such patients are on treatment until their 2-month follow-up appointment, when they are reassessed for response to treatment. The regimen in this period is a daily treatment with rifampicin, isoniazid, pyrazinamide and ethambutol. The *British National Formulary*[45] provides fixed dosages for adults, except for ethambutol hydrochloride, which is determined by patient weight. The mean weight at time of first presentation of 67.98 kg was used in the model. *Table 46* summarises the sensitivity and specificity values previously given in *Chapter 4, Comparison of diagnostic accuracy of T-SPOT.*TB *and QuantiFERON GOLD In-Tube* and *Tables 15* and *61* along with values for the beta PERT (project evaluation and review technique) distributions used for the probabilistic sensitivity analyses.

TABLE 46 Diagnostic test performance and distributions for different patient strata

Patient strata	Test	Indeterminate IGRA results	Sensitivity	95% CI	Beta distribution, Beta (a, b)	Specificity	95% CI	Beta distribution, Beta (a, b)
All	QFT-GIT	Excluded	0.673	0.620 to 0.721	(147,72)	0.804	0.761 to 0.841	(315,77)
All	T-SPOT.*TB*	Excluded	0.823	0.778 to 0.861	(299,64)	0.826	0.786 to 0.861	(293,60)
All	QFT-GIT	Included	0.697	0.647 to 0.742	(147,64)	0.715	0.671 to 0.756	(145,58)
All	T-SPOT.*TB*	Included	0.832	0.789 to 0.868	(290,59)	0.754	0.711 to 0.793	(348,112)
HIV positive	QFT-GIT	Excluded	0.565	0.368 to 0.744	(15,11)	0.920	0.843 to 0.961	(67,6)
HIV positive	T-SPOT.*TB*	Excluded	0.632	0.410 to 0.809	(14,8)	0.899	0.813 to 0.948	(74,7)
HIV positive	QFT-GIT	Included	0.600	0.407 to 0.766	(17,11)	0.748	0.658 to 0.820	(70,24)
HIV positive	T-SPOT.*TB*	Included	0.708	0.508 to 0.851	(18,7)	0.664	0.570 to 0.746	(73,36)
HIV negative	QFT-GIT	Excluded	0.681	0.627 to 0.731	(147,69)	0.770	0.718 to 0.814	(226,68)
HIV negative	T-SPOT.*TB*	Excluded	0.835	0.790 to 0.872	(287,57)	0.808	0.760 to 0.848	(250,59)
HIV negative	QFT-GIT	Included	0.704	0.653 to 0.751	(146,61)	0.704	0.652 to 0.752	(208,88)
HIV negative	T-SPOT.*TB*	Included	0.841	0.797 to 0.877	(280,53)	0.785	0.736 to 0.827	(264,72)

Chapter 7 Economic evaluation results

Introduction

This chapter presents the results of the economic evaluation. First, a data analysis with a health economics focus is given. This demonstrates the patterns and variability in times and costs between patients and also indicates why a simple direct estimation of the relevant summary statistics is not appropriate. Second, the modelling results are given and, finally, the main base-case scenario results are presented.

Results

The idealised diagnostic pathway representing current practice is shown in *Figure 8*. Patient records were analysed to determine what proportion of patients followed the pathways in this diagram. For each Dosanjh category, the particular tests expected to be performed and their corresponding results were determined and then compared with patient records. *Tables 47* and *48* present the frequencies of different tests performed, stratified by final diagnostic outcome. Importantly, the process of TB diagnosis rarely followed the idealised diagnostic pathways. Although culture, IGRAs (QFT-GIT and T-SPOT.*TB*), TSTs, sputum smear microscopy and chest radiography were frequently used, there was substantial variation between patients. These results are presented graphically in *Figure 9*, in which the number of patients who traverse each branch are indicated by the width of the branch. From the individual-level patient data, we used the empirical distributions of time to diagnosis and total test costs in the probabilistic sensitivity analysis in the following health economic evaluation.

Table 49 gives distributional statistics of the cost of testing for each final diagnosis category. There are long right-hand tails, making single centrality summary statistics misleading.

The decision tree model structure is shown in *Figure 10* and includes true-negative, false-negative, true-positive and false-positive rule-out test results. Patients receiving a positive rule-out test result (i.e. TB is *not* ruled out) then follow the standard diagnostic pathway, whereas patients receiving a negative rule-out test result are regarded as not having TB. If the rule-out test result is a false negative, then at the 2-month follow-up consultation the patient's persistent symptoms leads to them entering the standard TB diagnostic pathway.

For patients who are TB negative and who receive a true negative in the rule-out test, the rule-out test results in a faster diagnosis of the true cause of their illness. However, for other patients, the rule-out test increases the delay in their ultimate diagnosis, both for those who have TB and those who do not.

The 'standard pathway' branch represents the range of variation observed in the patient cohort. Probabilities are shown below branches following a circular chance node and costs are below branches following a square decision node. The rule-out test is either T-SPOT.*TB* or QFT-GIT.

Results of the cost-effectiveness analysis are summarised in *Table 50*. *Figure 11* shows cost-effectiveness planes for the use of T-SPOT.*TB* or QFT-GIT as rule-out tests, including three clouds of results for all patients, HIV-negative patients and HIV-positive patients. *Figures 12* and *13* show the cost-effectiveness planes with uncertainty represented by ellipses showing the 95% and 50% ranges of uncertainty, respectively.

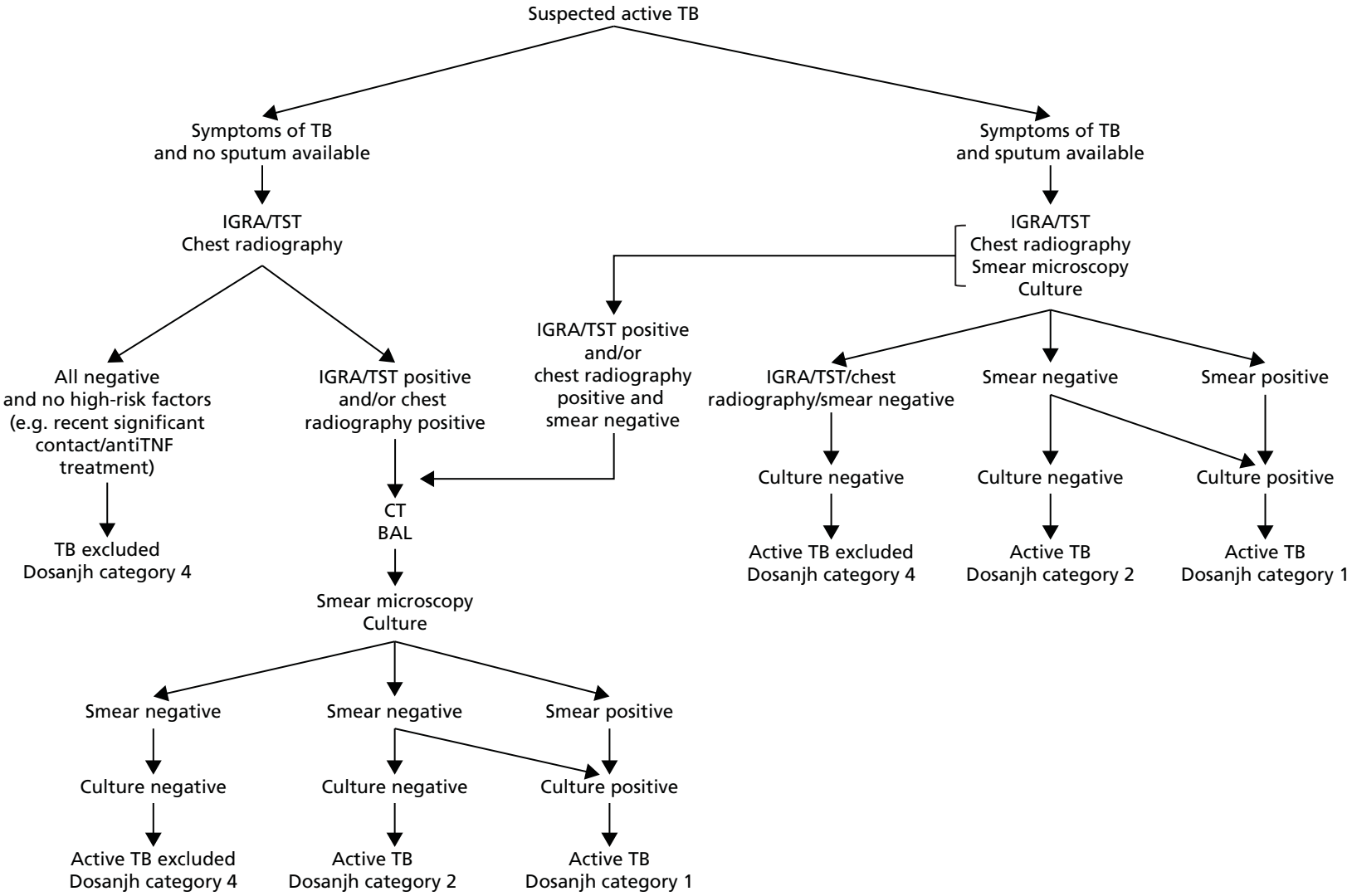

FIGURE 8 Idealised TB diagnostic pathway.

TABLE 47 Test result combinations of culture, sputum smear microscopy and chest radiography by final diagnosis

			Dosanjh category, *n*				
Culture	Sputum smear	Chest radiography	1	2	3	4	Total, *n*
Not performed	Not performed	Not performed	0	0	3	13	16
Not performed	Not performed	Indeterminate	0	0	2	13	15
Not performed	Not performed	Negative	0	0	7	61	68
Not performed	Not performed	Positive	0	3	0	1	4
Not performed	Negative	Not performed	0	0	0	3	3
Not performed	Negative	Indeterminate	0	0	1	2	3
Not performed	Negative	Negative	0	0	1	6	7
Not performed	Negative	Positive	0	0	0	1	1
Indeterminate	Not performed	Negative	0	0	1	0	1
Indeterminate	Negative	Negative	0	0	0	1	1
Indeterminate	Positive	Positive	0	0	0	1	1
Negative	Not performed	Not performed	0	0	2	8	10
Negative	Not performed	Indeterminate	0	1	1	10	12
Negative	Not performed	Negative	0	1	3	18	22
Negative	Not performed	Positive	0	2	0	1	3
Negative	Negative	Not performed	0	2	2	34	38
Negative	Negative	Indeterminate	0	4	4	65	73
Negative	Negative	Negative	0	7	16	235	258
Negative	Negative	Positive	0	11	4	48	63
Negative	Positive	Not performed	0	0	0	2	2
Negative	Positive	Indeterminate	0	0	0	1	1
Negative	Positive	Negative	0	0	0	4	4
Negative	Positive	Positive	0	1	0	1	2
Positive	Not performed	Not performed	1	0	0	0	1
Positive	Not performed	Indeterminate	3	0	0	0	3
Positive	Not performed	Positive	4	0	0	0	4
Positive	Negative	Not performed	8	0	0	0	8
Positive	Negative	Indeterminate	17	0	0	0	17
Positive	Negative	Negative	14	0	0	0	14
Positive	Negative	Positive	57	0	0	0	57
Positive	Positive	Not performed	3	0	0	0	3
Positive	Positive	Indeterminate	5	0	0	0	5
Positive	Positive	Negative	2	0	0	0	2
Positive	Positive	Positive	47	0	0	0	47
Total			161	32	47	529	769

TABLE 48 Tests performed at least once in the in-practice routine diagnostic pathways for suspected active TB

	Dosanjh category, *n*			
Test/sampling method	**1**	**2**	**3**	**4**
Culture	161 (1.00)	29 (0.91)	33 (0.70)	429 (0.81)
QFT-GIT	15 (0.09)	8 (0.25)	9 (0.19)	45 (0.09)
T-SPOT.*TB*	29 (0.18)	6 (0.19)	19 (0.40)	150 (0.28)
TST	55 (0.34)	10 (0.31)	21 (0.45)	164 (0.31)
Sputum smear microscopy	153 (0.95)	25 (0.78)	28 (0.60)	404 (0.76)
Bronchoalveolar lavage	50 (0.31)	14 (0.44)	9 (0.19)	135 (0.26)
Histology from biopsy	23 (0.14)	11 (0.34)	14 (0.30)	113 (0.21)
Needle aspirate	28 (0.17)	6 (0.19)	10 (0.21)	60 (0.11)
PCR	70 (0.43)	5 (0.16)	7 (0.15)	84 (0.16)
CXR	148 (0.92)	29 (0.91)	40 (0.85)	465 (0.88)
CT	85 (0.53)	22 (0.69)	27 (0.57)	323 (0.61)
MRI	16 (0.10)	3 (0.09)	7 (0.15)	57 (0.11)
PET	1 (0.01)	0 (0)	5 (0.11)	13 (0.02)

CXR, chest radiography.
Note
For each test, the proportion for each Dosanjh category is given in brackets.

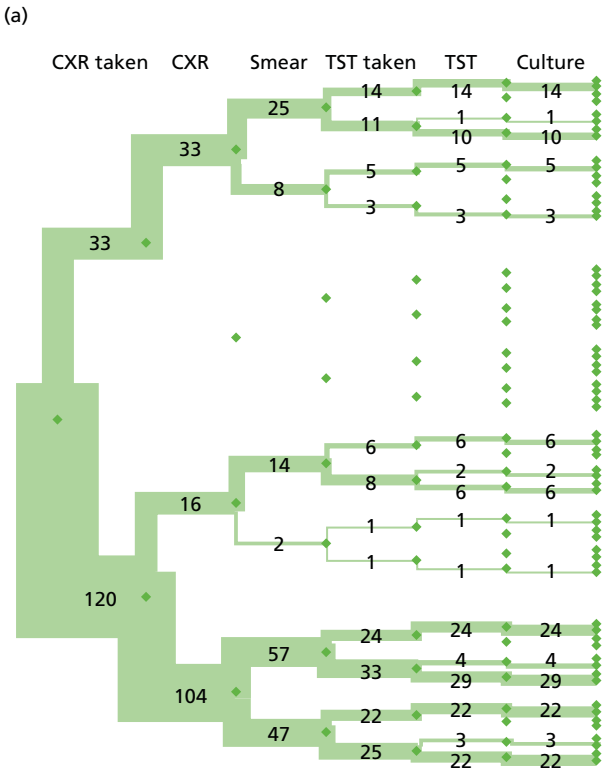

FIGURE 9 Representation of numbers of patients undergoing various TB tests, stratified by final diagnosis. (a) Culture-confirmed TB, Dosanjh category 1; (b) TB non-culture-confirmed, Dosanjh category 2; (c) Indeterminate, Dosanjh category 3; and (d) TB excluded, Dosnajh category 4. CXR, chest radiography. Patient numbers are shown on the branches. The frequencies of patients in each branch are proportional to the width of the branch. Upwards branches are for negative test results or if the decision was taken not to take a test (for chest radiography and culture) and the downwards branches are for positive results or a decision to test. (*continued*)

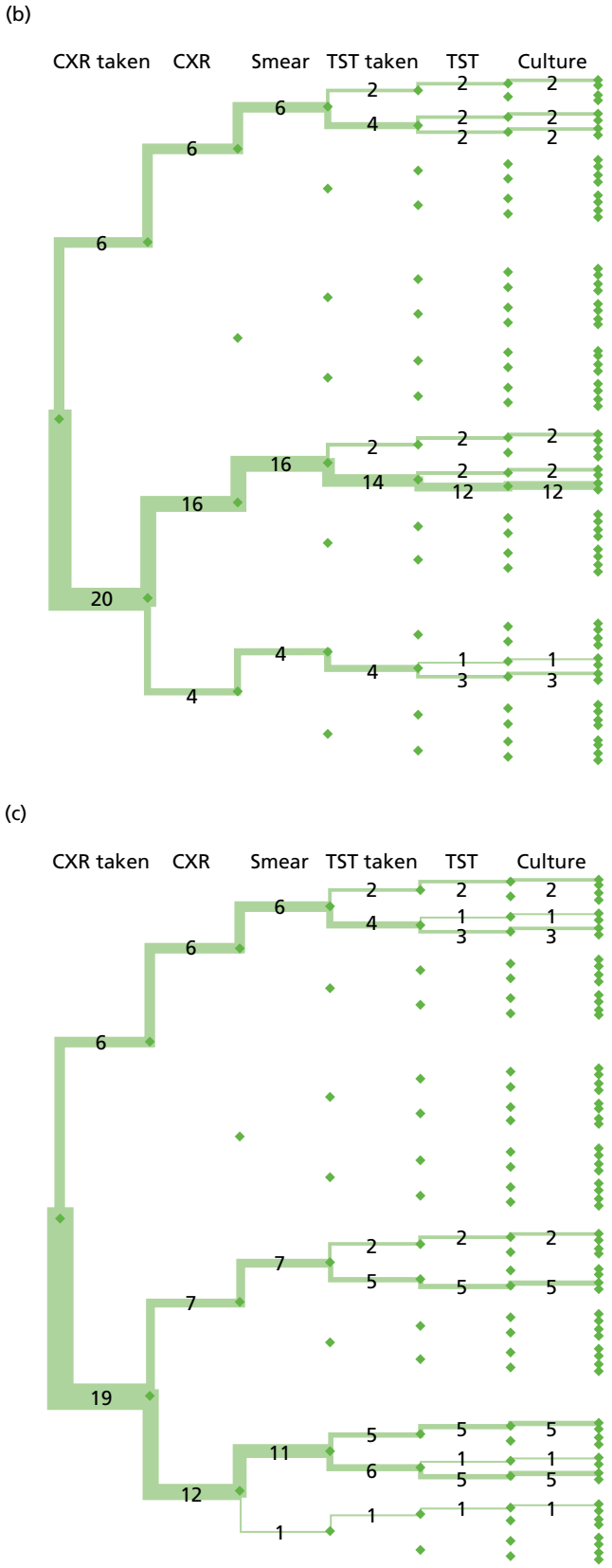

FIGURE 9 Representation of numbers of patients undergoing various TB tests, stratified by final diagnosis. (a) Culture-confirmed TB, Dosanjh category 1; (b) TB non-culture-confirmed, Dosanjh category 2; (c) Indeterminate, Dosanjh category 3; and (d) TB excluded, Dosnajh category 4. CXR, chest radiography. Patient numbers are shown on the branches. The frequencies of patients in each branch are proportional to the width of the branch. Upwards branches are for negative test results or if the decision was taken not to take a test (for chest radiography and culture) and the downwards branches are for positive results or a decision to test. (*continued*)

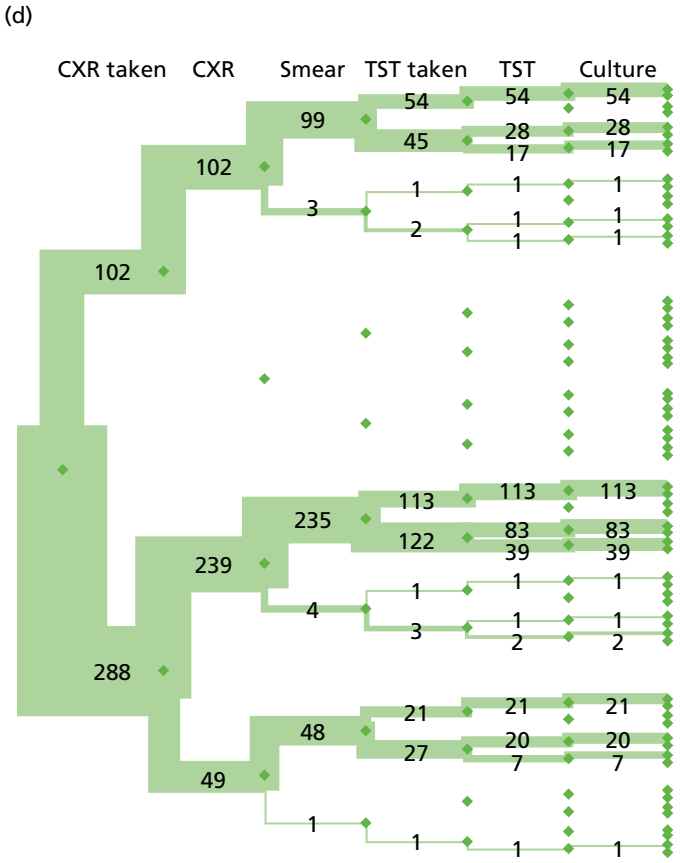

FIGURE 9 Representation of numbers of patients undergoing various TB tests, stratified by final diagnosis. (a) Culture-confirmed TB, Dosanjh category 1; (b) TB non-culture-confirmed, Dosanjh category 2; (c) Indeterminate, Dosanjh category 3; and (d) TB excluded, Dosnajh category 4. CXR, chest radiography. Patient numbers are shown on the branches. The frequencies of patients in each branch are proportional to the width of the branch. Upwards branches are for negative test results or if the decision was taken not to take a test (for chest radiography and culture) and the downwards branches are for positive results or a decision to test.

TABLE 49 Summary statistics for cost of diagnosis for the patient cohort by final diagnosis category

	Diagnosis cost (£)			
Dosanjh category	Lower quartile	Median	Mean	Upper quartile
1	292.31 (29.93)	442.92 (17.64)	469.77 (20.59)	640.03 (30.78)
2	383.73 (66.96)	476.12 (29.6)	474.73 (37.87)	572.03 (53.04)
3	254.08 (72.77)	502.04 (49.58)	535.48 (57.55)	644.12 (83.71)
4	202.00 (17.21)	433.39 (7.69)	445.61 (12.25)	588.47 (17.75)

Note
A total of 1000 bootstrap samples were used to give means and standard errors (in brackets) for each statistic.

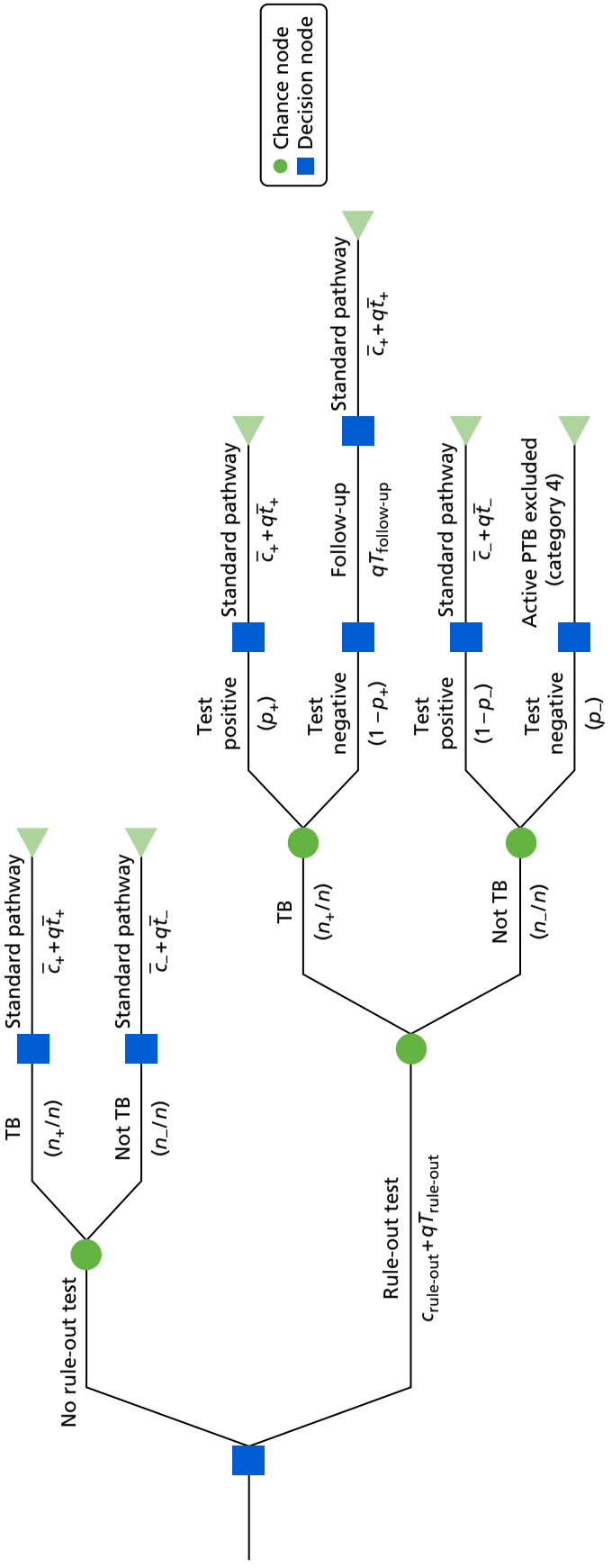

FIGURE 10 Decision tree comparing current practice ('no rule-out test') with a diagnostic pathway incorporating an initial rule-out test ('rule-out test'). PTB, pulmonary tuberculosis.

TABLE 50 Main results from cost-effectiveness analyses for the entire patient cohort

Patient strata	Test	Indeterminate test results	Incremental QALYs	Incremental cost (£000)	ICER (× 10^3)	Probability cost-effective at £20,000 per QALY	Probability cost-effective at £30,000 per QALY
All	QFT-GIT	Excluded	–6.22	–86.85	13.97	0.08	0.05
All	T-SPOT.*TB*	Excluded	–3.58	–78.81	22.01	0.26	0.22
All	QFT-GIT	Included	–6.50	–70.14	10.79	0.06	0.04
All	T-SPOT.*TB*	Included	–4.14	–65.12	15.73	0.21	0.17
HIV positive	QFT-GIT	Excluded	–7.14	–106.09	14.85	0.09	0.07
HIV positive	T-SPOT.*TB*	Excluded	–5.91	–86.87	14.70	0.12	0.08
HIV positive	QFT-GIT	Included	–7.86	–72.64	9.24	0.04	0.02
HIV positive	T-SPOT.*TB*	Included	–6.34	–42.11	6.64	0.04	0.03
HIV negative	QFT-GIT	Excluded	–6.66	–80.27	12.06	0.06	0.04
HIV negative	T-SPOT.*TB*	Excluded	–3.52	–75.31	21.40	0.26	0.21
HIV negative	QFT-GIT	Included	–6.47	–69.35	10.72	0.05	0.03
HIV negative	T-SPOT.*TB*	Included	–3.57	–71.45	20.04	0.24	0.19

ICER, incremental cost-effectiveness ratio.

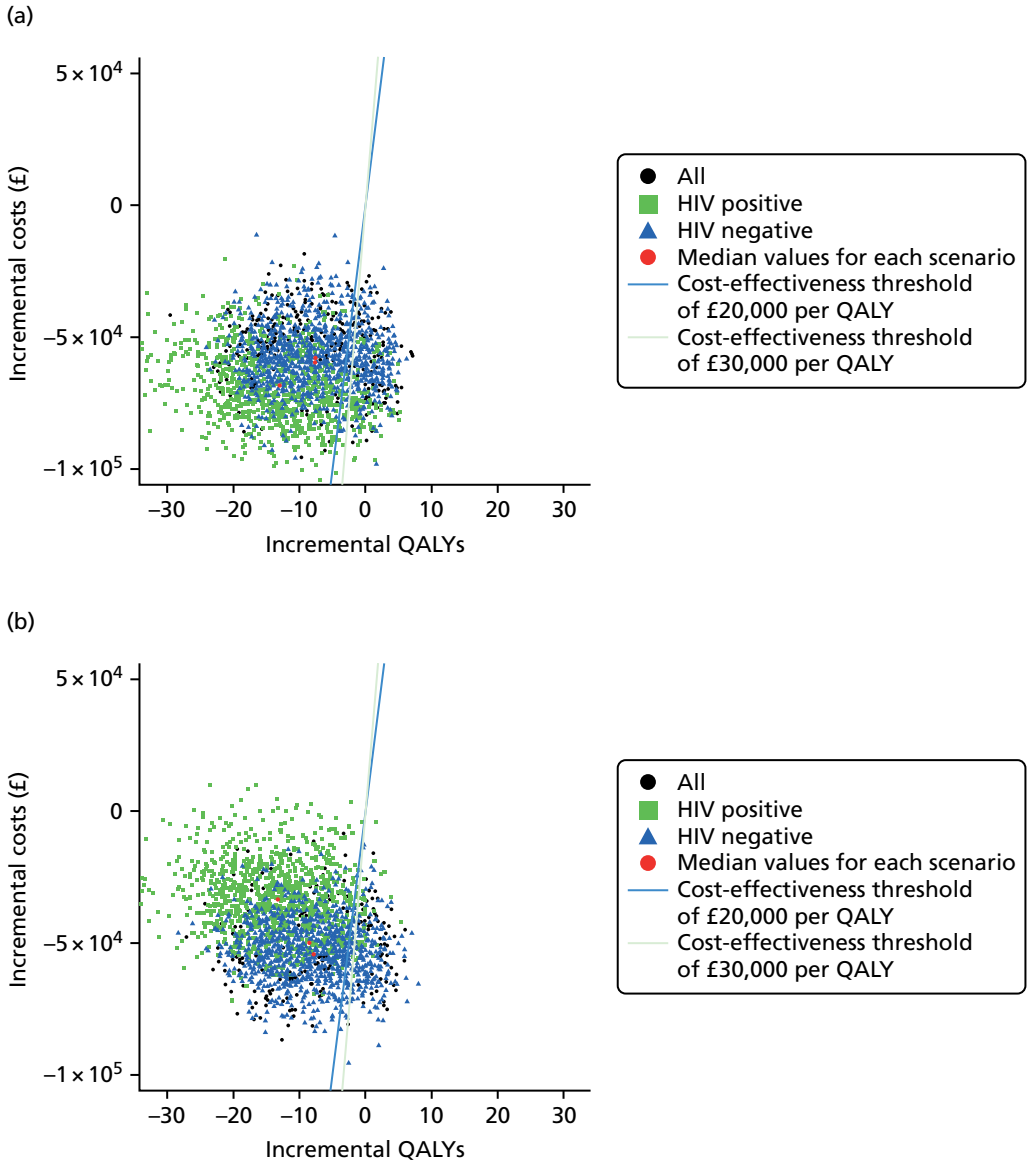

FIGURE 11 Cost-effectiveness planes comparing the use of T-SPOT.*TB* or QFT-GIT as rule-out tests with current practice. Results for (a) T-SPOT.*TB* with indeterminate diagnostic outcomes excluded; (b) T-SPOT.*TB* with indeterminate diagnostic outcomes included; (c) QFT-GIT with indeterminate diagnostic outcomes excluded; and (d) QFT-GIT with indeterminate diagnostic outcomes included. Statistics were calculated for the entire patient cohort. The upper panels show results for T-SPOT.*TB*, whereas the lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate diagnostic outcomes excluded and included, respectively. Note that an indeterminate IGRA result should not be confused with a diagnostic categorisation of a patient as being 'clinically indeterminate' with regard to having active TB (i.e. Dosanjh category 3). The analyses considered all patients (black circle), HIV-positive patients (green square) and HIV-negative patients (blue triangle), with each point showing the results of a single simulation result. The red points indicate the median values for each scenario. Diagonal lines indicate the cost-effectiveness thresholds of £20,000 per QALY (blue vertical line) and £30,000 per QALY (green vertical line). A total of 1000 simulations were run for each scenario. (*continued*)

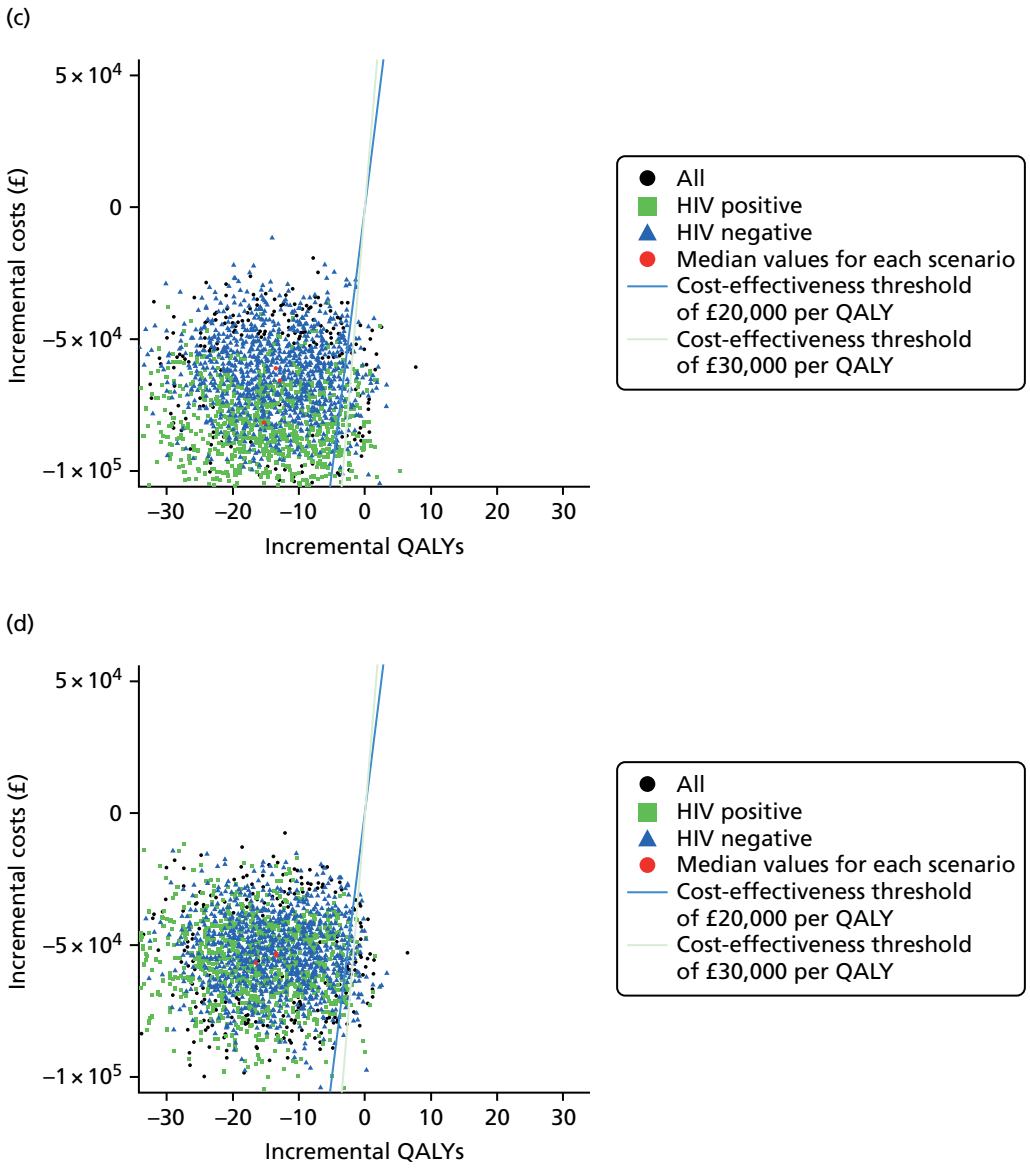

FIGURE 11 Cost-effectiveness planes comparing the use of T-SPOT.*TB* or QFT-GIT as rule-out tests with current practice. Results for (a) T-SPOT.*TB* with indeterminate diagnostic outcomes excluded; (b) T-SPOT.*TB* with indeterminate diagnostic outcomes included; (c) QFT-GIT with indeterminate diagnostic outcomes excluded; and (d) QFT-GIT with indeterminate diagnostic outcomes included. Statistics were calculated for the entire patient cohort. The upper panels show results for T-SPOT.*TB*, whereas the lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate diagnostic outcomes excluded and included, respectively. Note that an indeterminate IGRA result should not be confused with a diagnostic categorisation of a patient as being 'clinically indeterminate' with regard to having active TB (i.e. Dosanjh category 3). The analyses considered all patients (black circle), HIV-positive patients (green square) and HIV-negative patients (blue triangle), with each point showing the results of a single simulation result. The red points indicate the median values for each scenario. Diagonal lines indicate the cost-effectiveness thresholds of £20,000 per QALY (blue vertical line) and £30,000 per QALY (green vertical line). A total of 1000 simulations were run for each scenario.

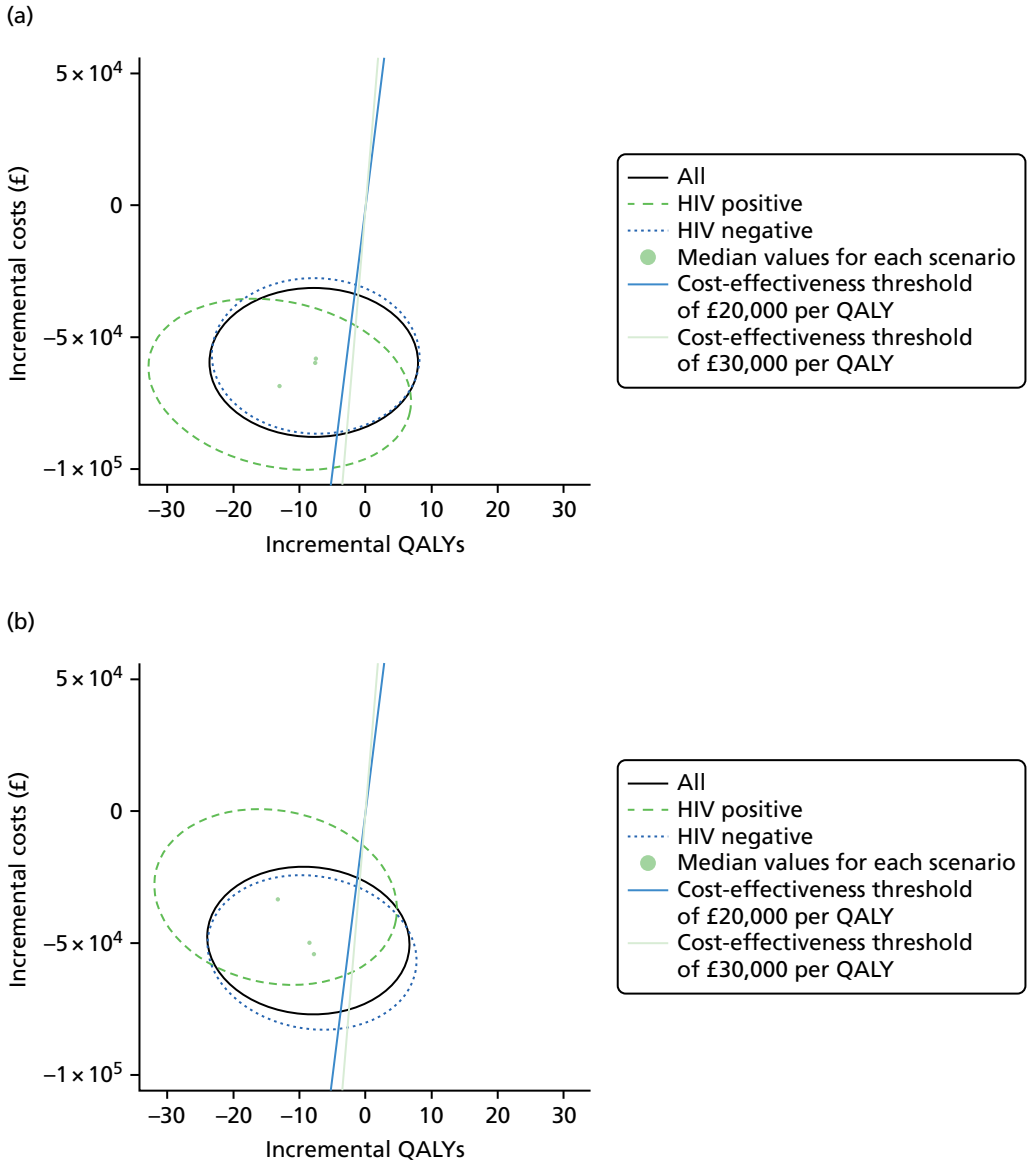

FIGURE 12 Cost-effectiveness planes comparing the use of T-SPOT.*TB* or QFT-GIT as rule-out tests with current practice (95% density of simulation results). Results for (a) T-SPOT.*TB* with indeterminate diagnostic outcomes excluded; (b) T-SPOT.*TB* with indeterminate diagnostic outcomes included; (c) QFT-GIT with indeterminate diagnostic outcomes excluded; and (d) QFT-GIT with indeterminate diagnostic outcomes included. Statistics were calculated for the entire patient cohort. The upper panels show results for T-SPOT.*TB*, whereas the lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate diagnostic outcomes excluded and included, respectively. Ellipses show 95% density of simulation results. The analyses considered all patients (black lines), HIV-positive patients (green dashed lines) and HIV-negative patients (blue dotted lines). The green dots indicate the median values for each scenario. Diagonal lines indicate the cost-effectiveness thresholds of £20,000 per QALY (blue vertical line) and £30,000 per QALY (green vertical line). A total of 1000 simulations were run for each scenario. (*continued*)

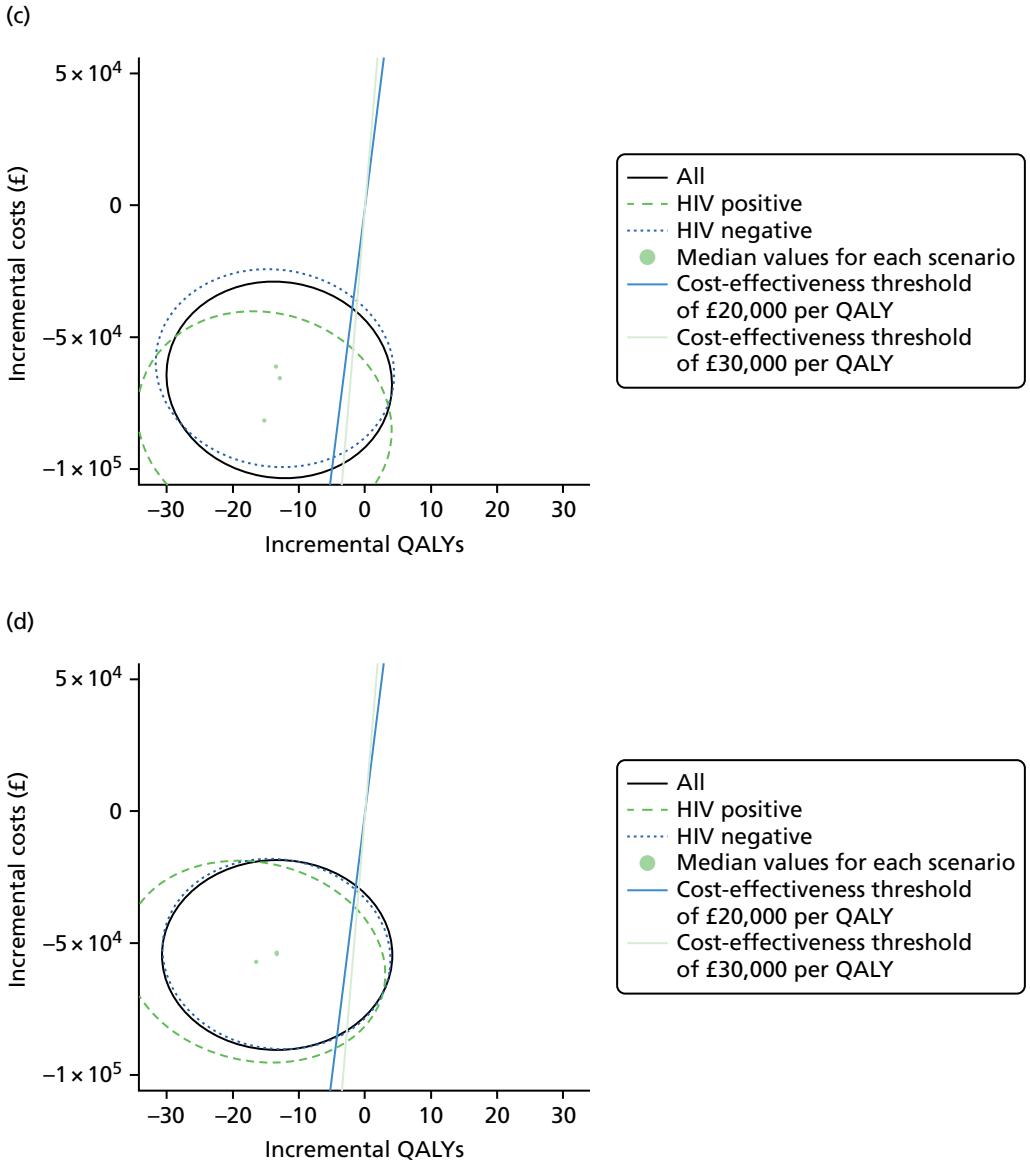

FIGURE 12 Cost-effectiveness planes comparing the use of T-SPOT.*TB* or QFT-GIT as rule-out tests with current practice (95% density of simulation results). Results for (a) T-SPOT.*TB* with indeterminate diagnostic outcomes excluded; (b) T-SPOT.*TB* with indeterminate diagnostic outcomes included; (c) QFT-GIT with indeterminate diagnostic outcomes excluded; and (d) QFT-GIT with indeterminate diagnostic outcomes included. Statistics were calculated for the entire patient cohort. The upper panels show results for T-SPOT.*TB*, whereas the lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate diagnostic outcomes excluded and included, respectively. Ellipses show 95% density of simulation results. The analyses considered all patients (black lines), HIV-positive patients (green dashed lines) and HIV-negative patients (blue dotted lines). The green dots indicate the median values for each scenario. Diagonal lines indicate the cost-effectiveness thresholds of £20,000 per QALY (blue vertical line) and £30,000 per QALY (green vertical line). A total of 1000 simulations were run for each scenario.

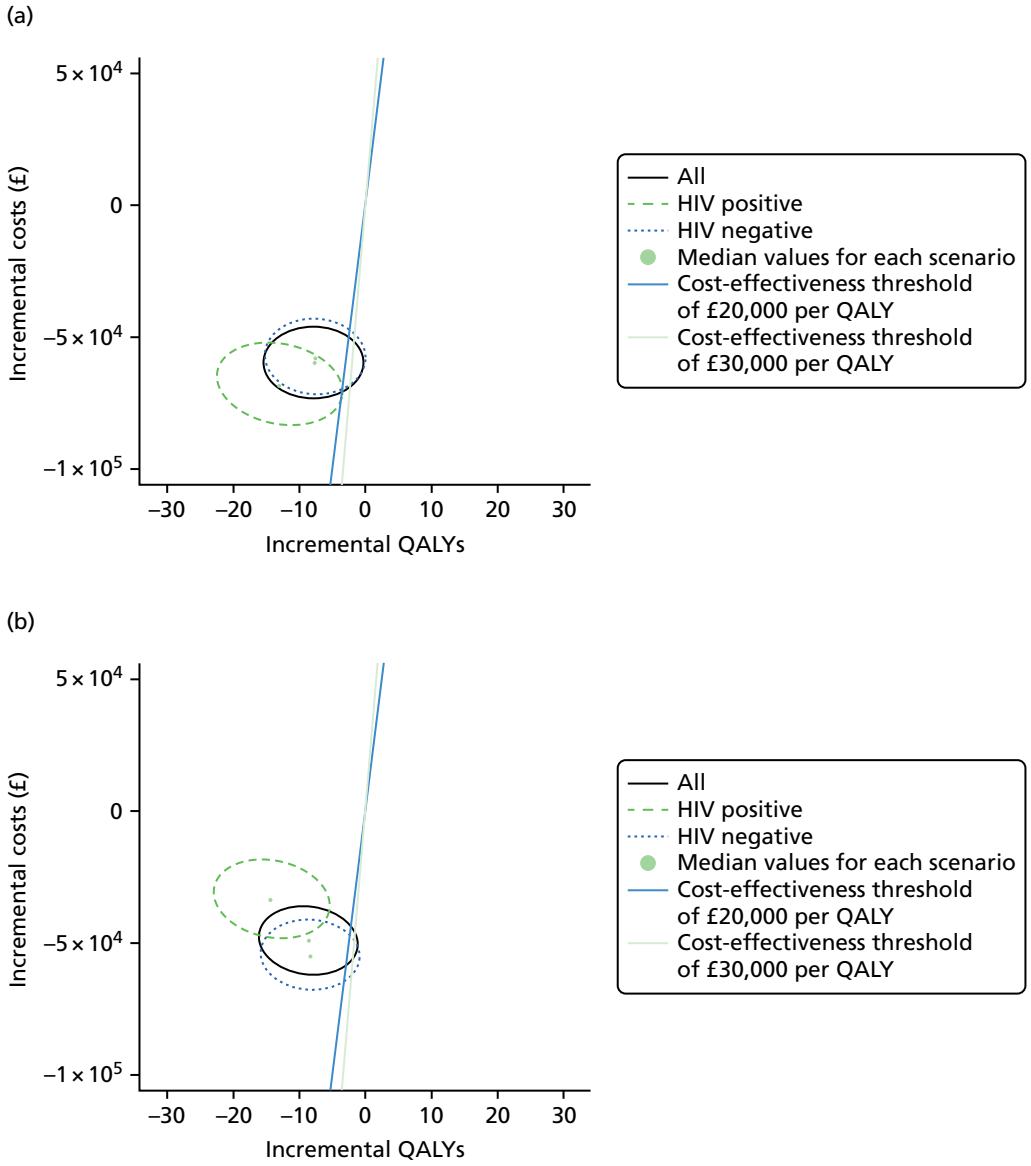

FIGURE 13 Cost-effectiveness planes comparing the use of T-SPOT.*TB* or QFT-GIT as rule-out tests with current practice (50% density of simulation results). Results for (a) T-SPOT.*TB* with indeterminate diagnostic outcomes excluded; (b) T-SPOT.*TB* with indeterminate diagnostic outcomes included; (c) QFT-GIT with indeterminate diagnostic outcomes excluded; and (d) QFT-GIT with indeterminate diagnostic outcomes included. Upper panels show results for T-SPOT.*TB*, whereas lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate diagnostic outcomes excluded and included, respectively. Ellipses show 50% density of simulation results. The analyses considered all patients (black lines), HIV-positive patients (green dashed lines) and HIV-negative patients (blue dotted lines). The green dots indicate the median values for each scenario. Diagonal lines indicate the cost-effectiveness thresholds of £20,000 per QALY (blue vertical line) and £30,000 per QALY (green vertical line). A total of 1000 simulations were run for each scenario. (*continued*)

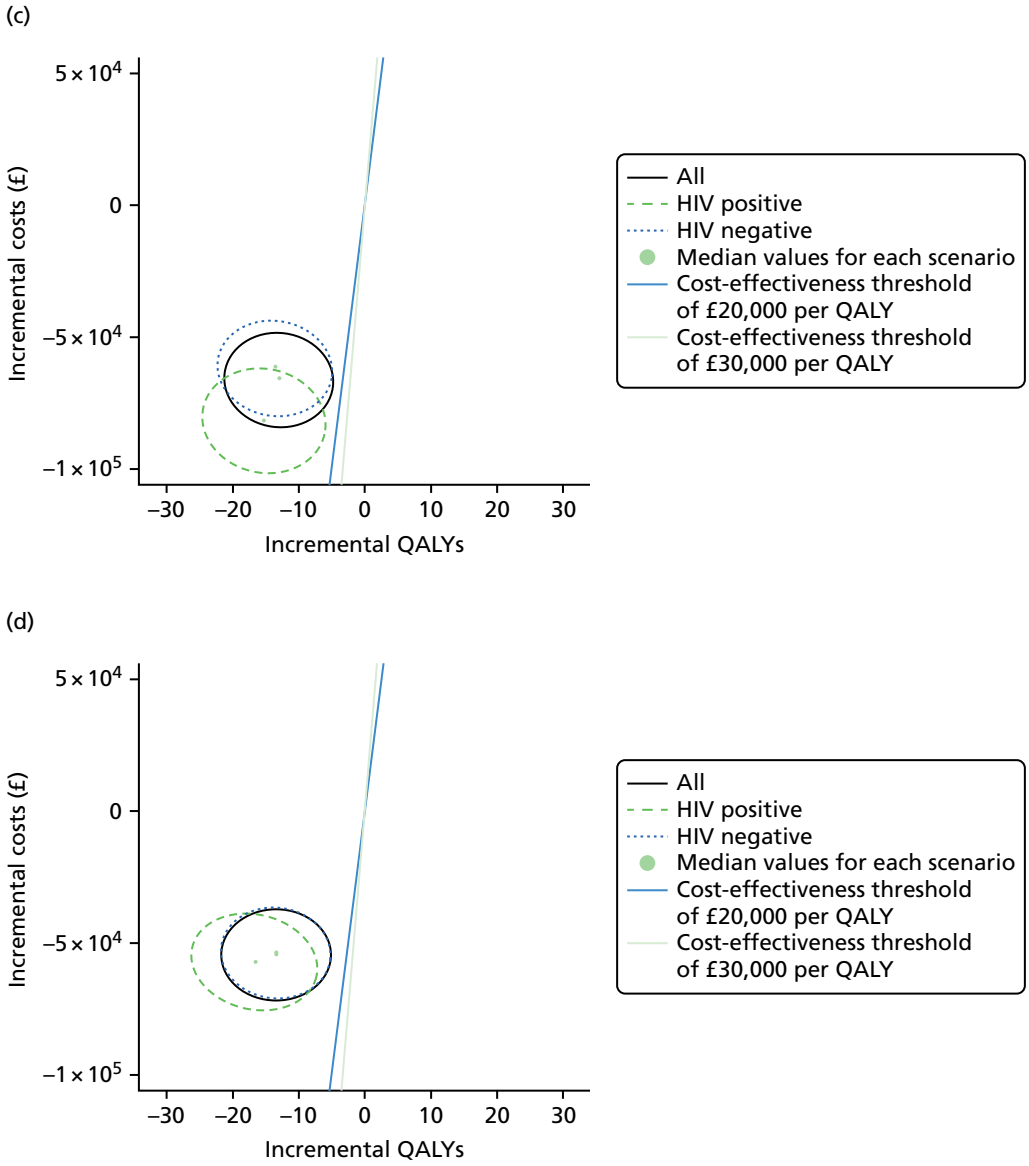

FIGURE 13 Cost-effectiveness planes comparing the use of T-SPOT.*TB* or QFT-GIT as rule-out tests with current practice (50% density of simulation results). Results for (a) T-SPOT.*TB* with indeterminate diagnostic outcomes excluded; (b) T-SPOT.*TB* with indeterminate diagnostic outcomes included; (c) QFT-GIT with indeterminate diagnostic outcomes excluded; and (d) QFT-GIT with indeterminate diagnostic outcomes included. Upper panels show results for T-SPOT.*TB*, whereas lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate diagnostic outcomes excluded and included, respectively. Ellipses show 50% density of simulation results. The analyses considered all patients (black lines), HIV-positive patients (green dashed lines) and HIV-negative patients (blue dotted lines). The green dots indicate the median values for each scenario. Diagonal lines indicate the cost-effectiveness thresholds of £20,000 per QALY (blue vertical line) and £30,000 per QALY (green vertical line). A total of 1000 simulations were run for each scenario.

Using IGRAs as a rule-out test is likely to be cost saving (except perhaps for T-SPOT.*TB* when used with HIV-positive patients), but also harmful to health. The magnitude of the health detriment is such that it would not be cost-effective if a QALY is valued at either £20,000 or £30,000 (indicated by diagonal lines).

As the use of IGRAs as a rule-out test is detrimental to health (because of the increased average time to diagnosis) but cost saving, the cost-effectiveness acceptability curves (*Figure 14*) show a high probability of rule-out testing being cost-effective when the value of a QALY is low, with the probability declining steeply as the value of a QALY increases. This is because if a QALY has a low value then for a given cost saving a relatively large loss of QALYs would be considered acceptable, whereas if a QALY has a high value then

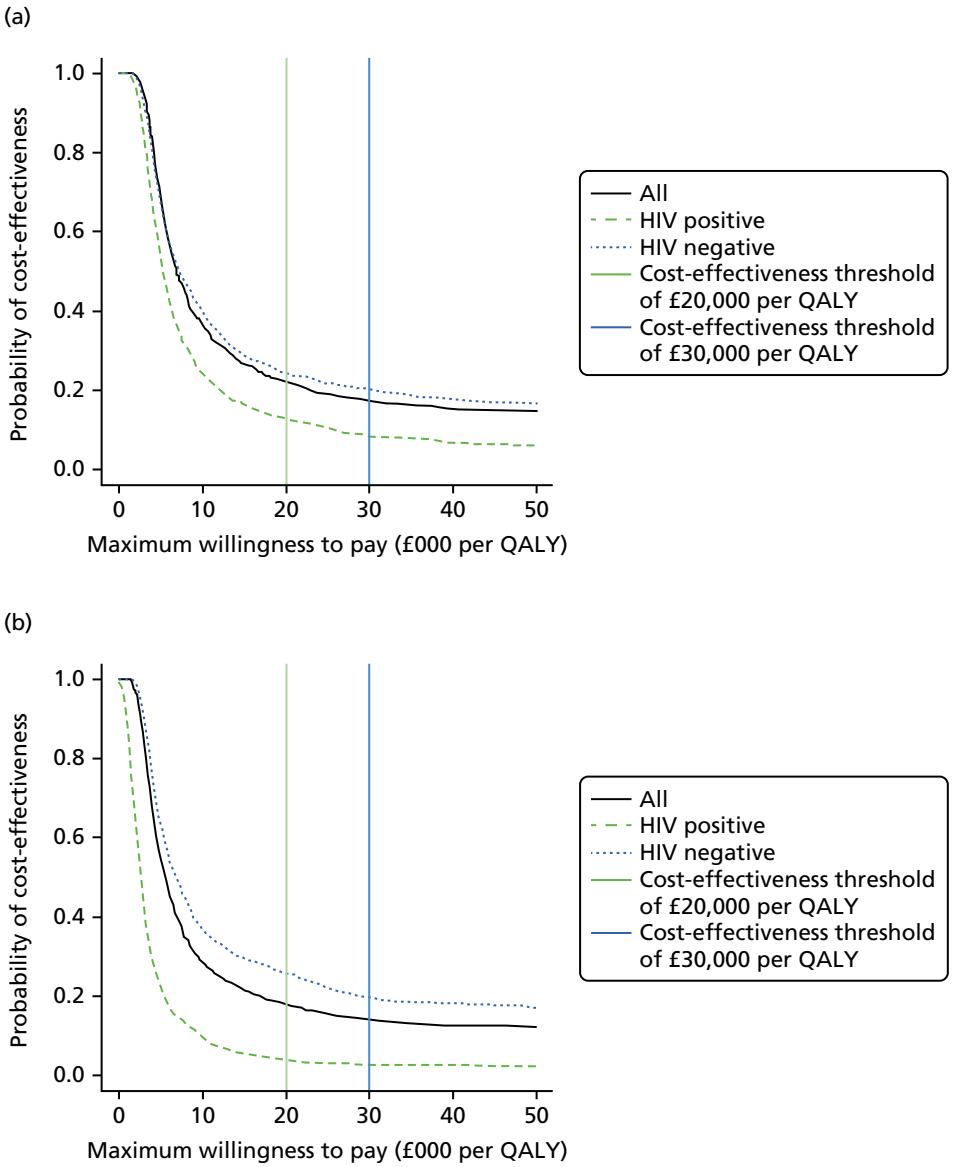

FIGURE 14 Cost-effectiveness acceptability curves for T-SPOT.*TB* or QFT-GIT as rule-out tests compared with current practice. Results for (a) T-SPOT.*TB* with indeterminate IGRA result patients excluded; (b) T-SPOT.*TB* with indeterminate IGRA result patients included; (c) QFT-GIT with indeterminate IGRA result patients excluded; and (d) QFT-GIT with indeterminate IGRA result patients included. Upper panels show results for T-SPOT.*TB*, whereas lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate IGRA result patients excluded and included, respectively. The analyses considered all patients (black lines), HIV-positive patients (green dashed lines) and HIV-negative patients (blue dotted lines). Vertical lines indicate thresholds of £20,000 per QALY (green vertical line) and £30,000 per QALY (blue vertical line). (*continued*)

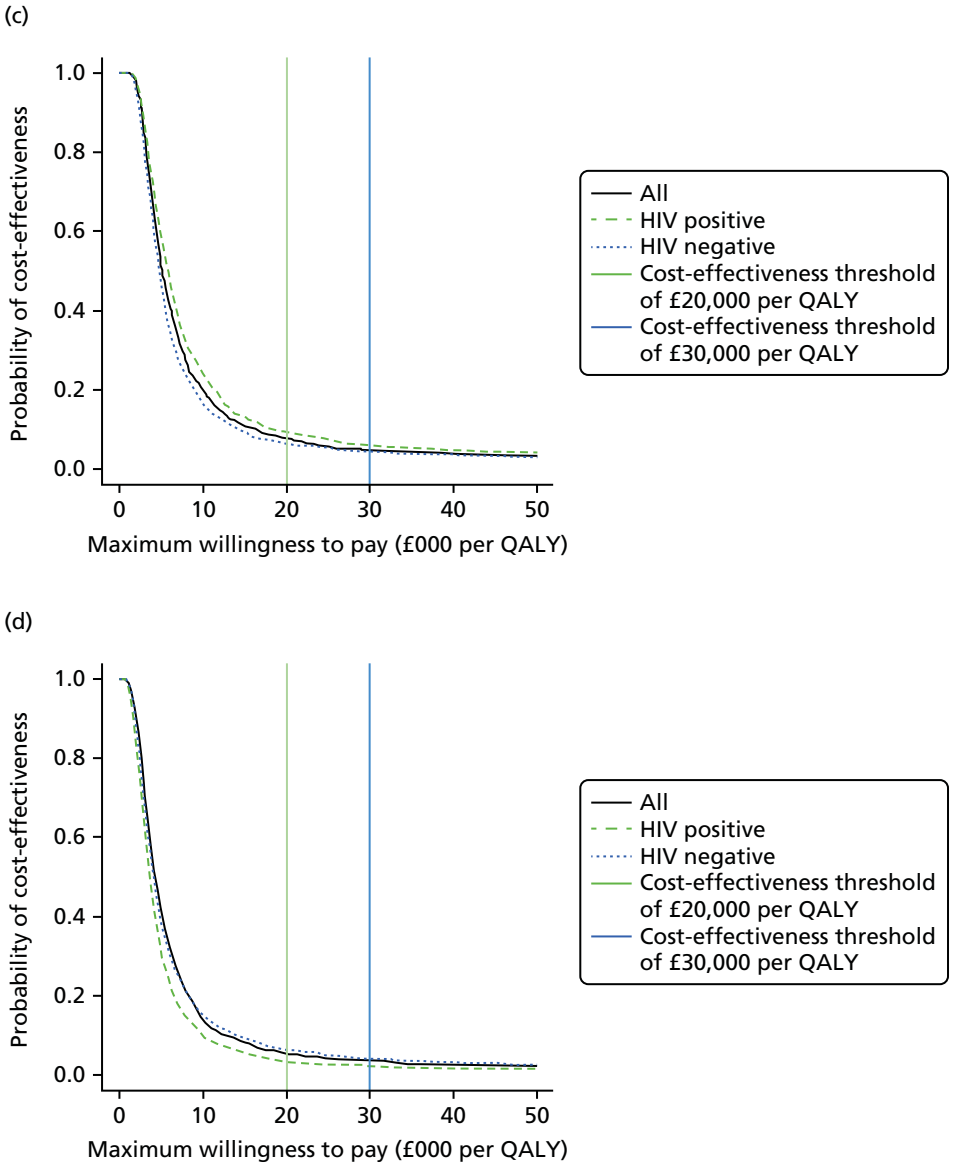

FIGURE 14 Cost-effectiveness acceptability curves for T-SPOT.*TB* or QFT-GIT as rule-out tests compared with current practice. Results for (a) T-SPOT.*TB* with indeterminate IGRA result patients excluded; (b) T-SPOT.*TB* with indeterminate IGRA result patients included; (c) QFT-GIT with indeterminate IGRA result patients excluded; and (d) QFT-GIT with indeterminate IGRA result patients included. Upper panels show results for T-SPOT.*TB*, whereas lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate IGRA result patients excluded and included, respectively. The analyses considered all patients (black lines), HIV-positive patients (green dashed lines) and HIV-negative patients (blue dotted lines). Vertical lines indicate thresholds of £20,000 per QALY (green vertical line) and £30,000 per QALY (blue vertical line).

such a QALY loss would not be acceptable. If a QALY is valued at £20,000, then all scenarios have a probability of < 30% of being cost-effective and many of them have a probability of < 10%. For £30,000 all scenarios have a probability of < 25% of being cost-effective, but many of them are in single figures. *Figure 15* shows the corresponding upper and lower 95% binomial CIs for the cost-effectiveness acceptability curves in *Figure 14*, calculated using the normal approximation to the binomial distribution.

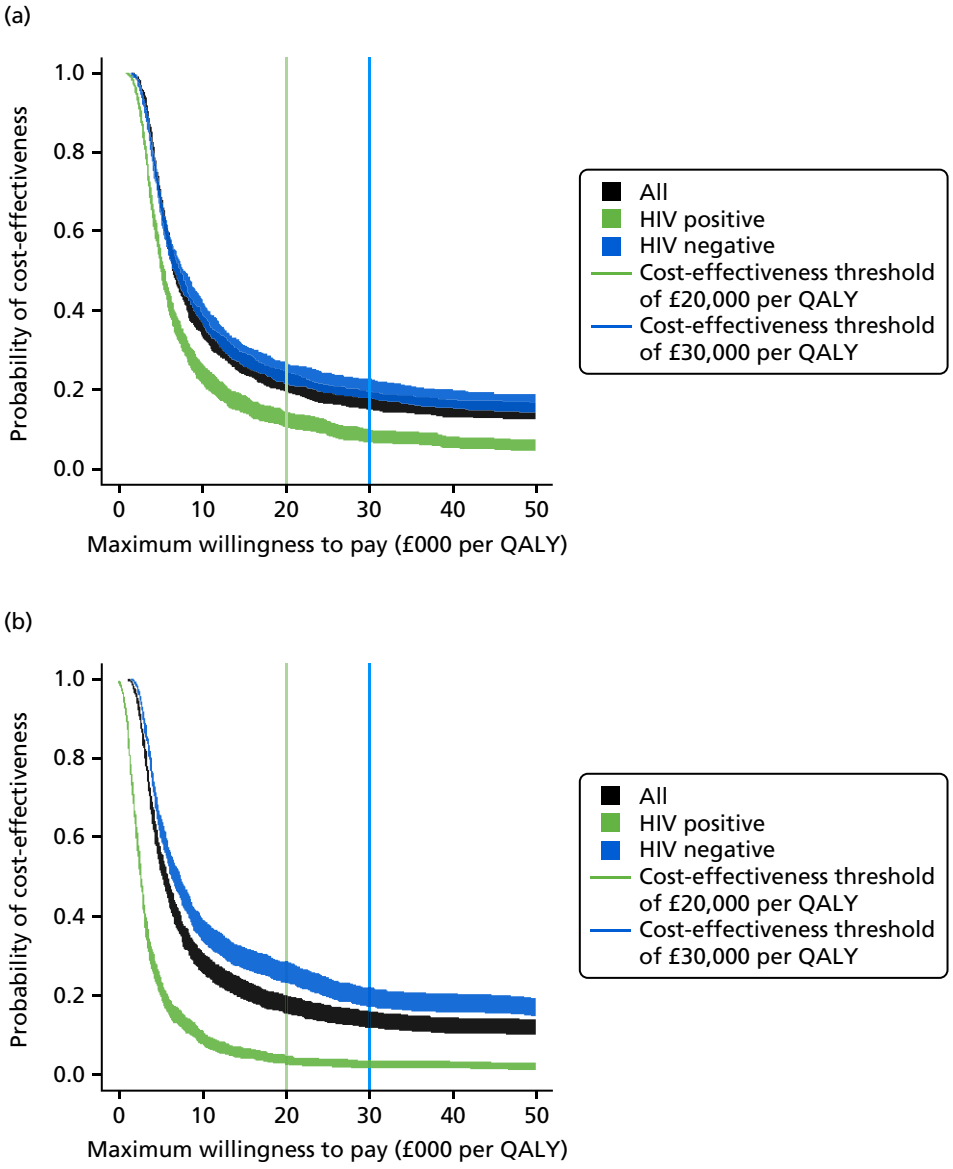

FIGURE 15 Cost-effectiveness acceptability curves (with 95% CI values) for T-SPOT.*TB* or QFT-GIT as rule-out tests compared with current practice. Results for (a) T-SPOT.*TB* with indeterminate IGRA result patients excluded; (b) T-SPOT.*TB* with indeterminate IGRA result patients included; (c) QFT-GIT with indeterminate IGRA result patients excluded; and (d) QFT-GIT with indeterminate IGRA result patients included. Upper and lower 95% CI values shown assuming normal approximation to the binomial distribution. Upper panels show results for T-SPOT.*TB*, whereas lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate IGRA result patients excluded and included, respectively. The analyses considered all patients (black solid lines), HIV-positive patients (green lines) and HIV-negative patients (blue lines). Vertical lines indicate thresholds of £20,000 per QALY (green vertical line) and £30,000 per QALY (blue vertical line). (*continued*)

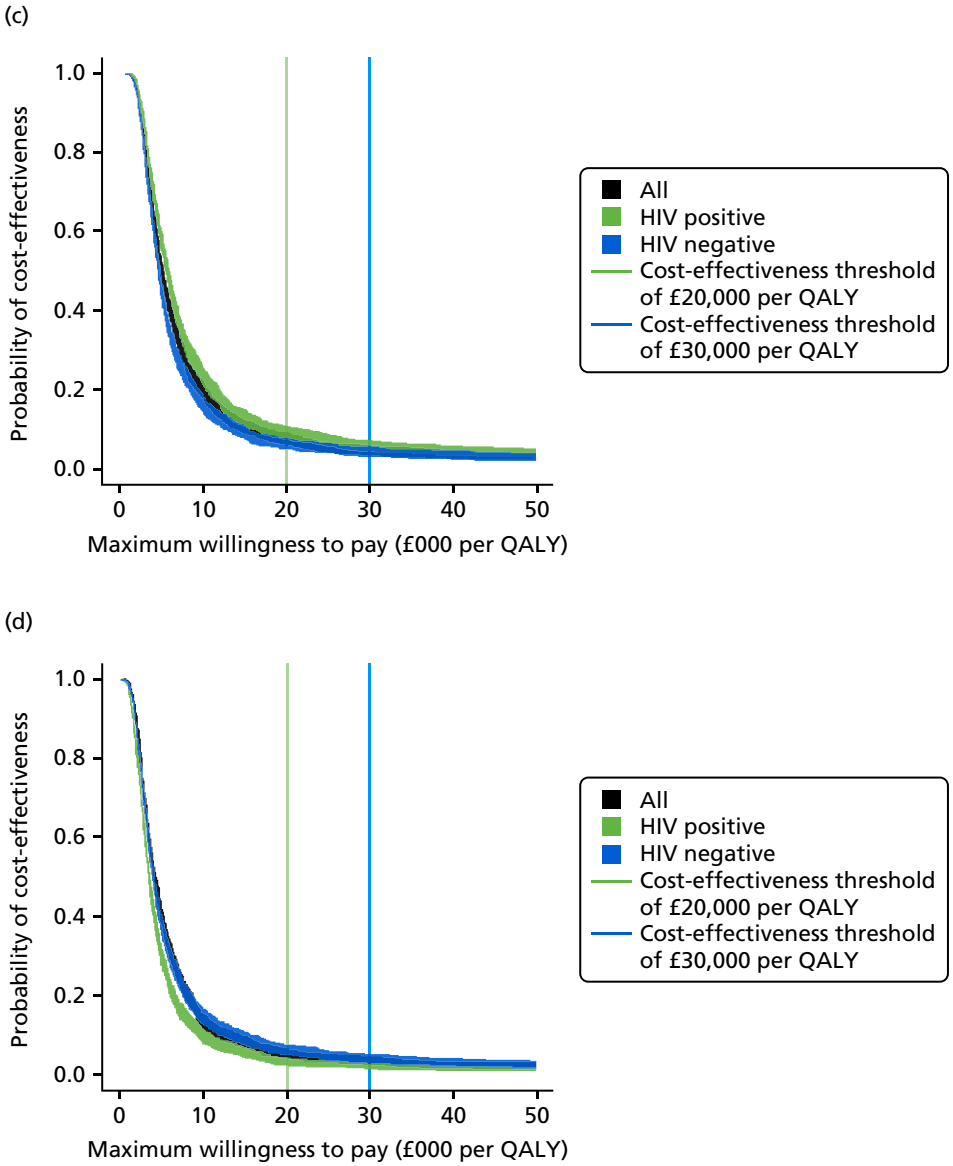

FIGURE 15 Cost-effectiveness acceptability curves (with 95% CI values) for T-SPOT.*TB* or QFT-GIT as rule-out tests compared with current practice. Results for (a) T-SPOT.*TB* with indeterminate IGRA result patients excluded; (b) T-SPOT.*TB* with indeterminate IGRA result patients included; (c) QFT-GIT with indeterminate IGRA result patients excluded; and (d) QFT-GIT with indeterminate IGRA result patients included. Upper and lower 95% CI values shown assuming normal approximation to the binomial distribution. Upper panels show results for T-SPOT.*TB*, whereas lower panels show results for QFT-GIT. Left- and right-hand panels present results with indeterminate IGRA result patients excluded and included, respectively. The analyses considered all patients (black solid lines), HIV-positive patients (green lines) and HIV-negative patients (blue lines). Vertical lines indicate thresholds of £20,000 per QALY (green vertical line) and £30,000 per QALY (blue vertical line).

Discussion

An important finding of this study is that TB diagnosis rarely follows the idealised diagnostic pathways, meaning that costs and time delays until diagnosis may be very different from what economic analyses typically assume. In particular, costs of diagnosis may be typically underestimated, particularly when scanning is involved. Furthermore, there is considerable individual-level variation in the costs and time taken for TB diagnosis, which needs to be represented in the analysis. Therefore, we included the empirical distributions of time to diagnosis and total test costs in the probabilistic sensitivity analysis in the health economic evaluation.

The use of current IGRA tests for ruling out active TB would be unlikely to be considered cost-effective if a QALY were to be valued at £20,000 or £30,000; although it is cost saving, the health detriment is large. Health detriment occurs because of a delay in diagnosing active TB, prolonging illness, for two reasons. First, adding a rule-out test to the diagnostic pathway adds a step and increases the costs and time taken to ultimate diagnosis for all patients except those who do not have TB and receive a true-negative rule-out test result so that they do not undergo the remainder of TB-specific tests. Second, some patients with TB receive a false-negative result from the rule-out test, which delays their diagnosis of TB until after their follow-up appointment. However, for patients who do not have TB, and who have TB correctly ruled out by the initial test, the time to diagnosis of the cause of their illness is reduced, producing a health gain, and there is a cost saving owing to their not having further tests for TB. Whether there is a net health detriment or gain for the patient cohort as a whole depends on the prevalence of active TB in the patients, the performance characteristics of the rule-out test and the length of delay introduced by adding the initial rule-out test.

Limitations and generalisability

The multicentre design means that the study population is representative of the general population with clinically relevant risk of TB in the UK and is therefore representative of the greatest TB burden in the country. In areas where TB rates are low, the patient populations might be substantially different, and this has not been tested in the study. The model is flexible and is suitable for analysing rule-out tests with a range of performance characteristics, and could also be applied to patient populations in lower-burden settings if suitable data were available. A limitation is that, although we have quantified the health detriment of delayed treatment due to prolonged morbidity, we were unable to account for any additional detriment due to a potentially poorer prognosis, because of a lack of suitable data. In addition to a lack of data, the assumed time taken to apply the hypothetical rule-out test is an assumption based on expert opinion, and we have assumed that the time to follow-up of those patients who receive a negative hypothetical rule-out test result can be inferred from our existing data set. The timings of in-practice events were estimated from the study population. The variability in times between patients were explicitly included in the model.

Conclusions and recommendations

The use of current IGRA tests for ruling out active TB would be unlikely to be considered cost-effective if a QALY were to be valued at £20,000 or £30,000. A health detriment in patients with active TB whose diagnosis and treatment are delayed needs to be balanced against health gains in patients who do not have active TB whose diagnosis of the true cause of illness and commencement of appropriate treatment are accelerated, with consideration given to cost savings of faster ruling out of active TB. Although the performance of current IGRA tests means they are not cost-effective, improved test technology or alternative algorithms using current technology could potentially have a performance that is cost-effective.

Future research recommendations are that improved testing technology, including combinations of tests, be investigated with an equivalent analysis. The knowledge base in this field of research is improved data and understanding at the national and local level.

Chapter 8 Discussion

Diagnostic accuracy findings for T-SPOT.*TB* and QFT-GIT were described in detail in *Chapters 4* and *5*, and the results of the economic evaluation were discussed in *Chapter 7*. Therefore, this chapter focuses on the principal findings, strengths and limitations of the study, and implications for health care and research.

Principal findings

This large multicentre study of a consecutive series of patients being investigated for possible TB is representative of routine clinical practice in the UK. The study clearly showed that T-SPOT.*TB* is more sensitive than QFT-GIT (relative sensitivity 1.22, 95% CI 1.14 to 1.31; $p < 0.001$), but that specificities are similar (relative specificity 1.02, 95% CI 0.97 to 1.08; $p = 0.3$). For T-SPOT.*TB* and QFT-GIT, the sensitivities were 82.3% (95% CI 77.7% to 85.9%) and 67.3% (95% CI 62.1% to 72.2%), respectively, whereas the specificities were 82.6% (95% CI 78.6% to 86.1%) and 80.4% (95% CI 76.1% to 84.1%), respectively. In the substudy of HIV-positive patients, the highest-risk subgroup for TB, T-SPOT.*TB* also showed higher sensitivity than QFT-GIT (a relative increase of 12%), but there was no statistical evidence of a difference in sensitivity. The sensitivity of T-SPOT.*TB* in all patients (82.3%) or among HIV co-infected patients (68.0%) is not of sufficient clinical utility as a rule-out test for the diagnosis of active TB in routine clinical practice and cannot be used in isolation as a rule-out test. This is further supported by the economic evaluation. The use of current IGRA tests for ruling out active TB would be unlikely to be considered cost-effective if a QALY were to be valued at £20,000 or £30,000. There are cost savings, but the health detriment is large because of the delay in diagnosing active TB.

The specificities of the IGRAs were also not adequate for IGRAs to be recommended as rule-in tests. However, in category 4D patients (active TB excluded, TST negative and no risk factors for LTBI), both T-SPOT.*TB* and QFT-GIT showed high specificity: 92.3% (95% CI 85.4% to 96.4%) and 93.4% (95% CI 86.4% to 96.9%), respectively. This suggests that IGRAs may have potential value for ruling in TB in settings with a low probability of active TB. However, it should be noted that the number of category 4D patients was small.

For key subgroups in our main study cohort – patients with pre-existing diabetes mellitus, end-stage renal failure or iatrogenic immunosuppression – data were limited. Of the 845 patients included in our analyses, 88 (10.4%) had diabetes mellitus, 12 (1.4%) had chronic/end-stage renal failure and 105 (12.4%) patients were on immunosuppressive therapy. Analysis of patients with and without diabetes mellitus showed that the sensitivities and specificities of both tests were higher in those without diabetes mellitus than in those with diabetes mellitus. However, there was no statistical evidence of an effect on the relative test performance of T-SPOT.*TB* and QFT-GIT. Although this finding should be interpreted with caution because of the small number of active TB cases in the analyses, association between diabetes mellitus and IGRA performance has been reported. Faurholt-Jepsen *et al.*[47] reported an association between diabetes mellitus and lower levels of Mtb antigen-specific IFN-γ and the impact on QFT-GIT results. Our findings are similar, even though diabetes mellitus was self-reported in the IDEA study. We found no study that has evaluated the effect of diabetes mellitus on the diagnostic performance of T-SPOT.*TB*; the IDEA study appears to be the first study to suggest differences in the performance of T-SPOT.*TB* between diabetic and non-diabetic patients.

A new generation of QFT-GIT, QuantiFERON-TB Gold Plus (QFT®-Plus; Qiagen GmbH, Hilden, Germany), was recently launched. QFT-Plus includes a set of peptides designed to stimulate Mtb-specific cluster of differentiation 8-positive T cells.[48] There is very little published evidence about the performance of the test. According to Barcellini *et al.*,[48] who reported the first independent assessment of QFT-Plus, the test had sensitivity of 87.9% (95% CI 80.8% to 92.7%) in 116 TB patients and specificity of 97.2% (95% CI 92.0% to 99.0%) in 106 low-risk controls. In the IDEA study, second-generation IGRAs utilising novel

antigens showed potential as rule-out tests. In particular, the use of a combination of existing antigens – ESAT-6 and CFP-10 – and the newer antigens – Rv3615c and Rv3879c – achieved a sensitivity of 89.9% (95% CI 86.2% to 92.7%) based on all active TB cases and 94.4% (95% CI 90.7% to 96.7%) among culture-confirmed cases. Similar results were obtained for the two-antigen combination of CFP-10 and Rv3615c and the three-antigen combination of ESAT-6, CFP-10 and Rv3615c. The added value of Rv3615c to T-SPOT.*TB* was a 9% (95% CI 5% to 12%) relative increase in sensitivity at the expense of specificity with a relative decrease of 7% (95% CI 4% to 10%). The incremental gain in sensitivity to 90% is likely to be clinically useful in ruling out active TB.

Strengths and limitations

To our knowledge, the IDEA study is the largest, prospective comparative accuracy study of the role of IGRAs for the diagnosis of active TB. Furthermore, we recruited a consecutive series of patients who were representative of UK clinical practice in a high-income setting. We ensured completeness and quality of the data such that missing data were minimal. Therefore, this well- designed and well-conducted study enabled robust and precise estimation of the relative performance of T-SPOT.*TB* and QFT-GIT in the main study cohort.

Our study has wide applicability as we did not exclude key subgroups, such as HIV-positive patients, but aimed to also compare the clinical performance of T-SPOT.*TB* and QFT-GIT in this population. In spite of not achieving the target sample size for the HIV-positive subgroup, compared with the five comparative studies identified in a systematic review published in 2012 and our literature review,[15] the IDEA study remains the largest prospective head-to-head comparison of the two IGRAs in a HIV co-infected population.

The final diagnosis of TB is challenging in non-culture-confirmed cases and relies on a combination of epidemiological, radiological and diagnostic parameters. We used a composite reference standard applied by a panel of experienced clinicians. The panel followed a strict protocol. In the group of patients with non-microbiologically defined highly probable TB, there was a rigorous process to ensure that the clinical panel reviewed all cases using case histories, imaging, other test results and follow-up data to ensure that categorisation was as accurate as possible. The panel was blinded to the study and routine IGRA results to avoid potential for bias.

The IDEA study has limitations. First, owing to the low number of bronchoalveolar samples analysed, we were unable to fully characterise the value of IGRAs using BAL fluid. Therefore, we did not fulfil the secondary objective of determining the diagnostic accuracy of the two IGRAs applied to BAL samples in patients with suspected pulmonary TB who were sputum smear negative, as stated in the protocol.

Second, as already alluded to, we did not achieve our recruitment target for the HIV co-infected population. The IDEA study was extended to facilitate recruitment and more centres were included. We did achieve our original target of 200 patients, which was based on a prevalence of active TB of 50%. However, the final diagnosis of active TB in the 201 HIV-positive patients included in the analyses of the substudy was even lower (32/201, 15.9%) than the 20% used in our revised sample size calculation.

Finally, for the economic evaluation, we found that TB diagnosis rarely followed idealised diagnostic pathways, implying that costs and time delays until diagnosis may be very different from what economic analyses typically assume. In particular, costs of diagnosis may be typically underestimated, particularly when other modalities of imaging apart from plain chest radiology such as CT, ultrasound or magnetic resonance scanning are involved.

Implications for health care

Despite the significantly higher diagnostic sensitivity of T-SPOT.*TB* over QFT-GIT, neither of the two IGRAs can be used routinely as a reliable rule-out test for suspected active TB in this patient population in secondary care. Both IGRAs were also not cost-effective in this setting. However, for patients in which there is a suspicion of TB, but the pre-test probability is low, the NPV of a negative T-SPOT.*TB* result would be correspondingly higher.

The specificity of both IGRAs for a diagnosis of active TB was similar and too low to use as a rule-in test. However, in patients with suspected active TB, a positive IGRA result could help in certain circumstances to keep TB in the differential diagnosis and guide further diagnostic testing towards confirming or excluding a diagnosis of TB. A positive result in the setting of a HIV-infected patient with suspected TB provides clinically useful information, as specificity in this setting was higher than in HIV-negative patients, especially for QFT-GIT.

The incorporation of novel antigens into T-SPOT.*TB*, in particular Rv3615c, yielded a high sensitivity coupled with a modest reduction in specificity. The sensitivity and NPV of 90% in this high-prevalence population in secondary care are compatible with the use of this assay to exclude a diagnosis of TB in patients with lower pre-test probabilities. The 95% sensitivity in culture-confirmed TB is similar to that of the D-dimer assay, which is routinely used as a rapid rule-out test for suspected venous thromboembolism in patients with a low to moderate pre-test probability.

Notably, replacing ESAT-6 with Rv3615c also conferred higher sensitivity than T-SPOT.*TB*. Indeed, the diagnostic sensitivity and NPV of 89% were similar to that achieved by the incorporation of Rv3615c alongside both ESAT-6 and CFP-10. This observation is relevant for TB control internationally because one of the leading TB vaccine candidates currently in clinical trials, H56/IC31, incorporates ESAT-6. The vaccine is protective in the non-human primate model and if it proves to be protective in humans, it is likely to be licensed. If rolled out, vaccinated individuals will likely develop T-cell responses to ESAT-6, which would give false-positive IGRA results, akin to the current scenario with BCG vaccination inducing false-positive TST results. Replacing ESAT-6 with Rv3615c may be a potential solution because a CFP-10- and Rv3615c-based IGRA would have significantly higher sensitivity than existing IGRAs and specificity would not be compromised in H56/IC31-vaccinated individuals.

Recommendations for research

The second-generation IGRAs evaluated in this study do not need to be re-evaluated in a UK routine practice setting because this study enabled an equally rigorous evaluation of these novel assays as it did for conventional IGRAs. Precise estimates of diagnostic accuracy were obtained. However, it would be of interest to evaluate these new assays and their combinations in distinct clinical settings with much lower or much higher prevalence of active TB. It will also be important to assess how these novel IGRAs perform in immunosuppressed subgroups, including HIV-infected patients, diabetic patients with chronic renal impairment and those on immunosuppressive therapy. A comparative accuracy study of the novel assays and QFT-Plus may also be needed to determine how their sensitivity compares in routine practice.

Acknowledgements

We are grateful to clinical and research staff involved in facilitating recruitment and data collection at all participating centres. We would like to thank:

Independent members of the SSC for their advice and support throughout the study.

Lee Potiphar, Senior Research Nurse, for his contribution to the original study protocol and management of the IDEA study nursing team between September 2011 and June 2013.

Anil Sharma, Senior Research Nurse, for managing the IDEA study nursing team between June 2013 and September 2013.

The IDEA study Research Nurses: Samuel Bremang (December 2011 to February 2015), Anil Sharma (November 2011 to February 2014), Gary Hahn (November 2011 to November 2013), Lisa Grass (November 2011 to February 2014), Linda Haimbodi (September 2012 to January 2014) and Hiromi Uzu (May 2014 to February 2015).

Laboratory staff (Dennis Afram, Luigi Marongiu and Bianca Donges) and other staff (Ann-Kathrin Reuschl and Nazneen Siddiqui) involved in blood processing.

Dhansuklal Solanki and Denise Gardner for administrative assistance.

We thank the National Institute for Health Research (NIHR) Health Technology Assessment programme for funding, and the NIHR Health Protection Research Unit, Imperial College London for support. We especially thank all IDEA participants for their involvement in the study.

Staff who facilitated recruitment and data collection

Centre	PI	Staff that recruited patients and performed data collection
Imperial College Healthcare NHS Trust	Professor Ajit Lalvani and Professor Onn Min Kon	Dr Matthew Berry, Dr Frances Sanderson, Dr Graham Cooke, Dr Tommy Pasvol, Mr Lee Potiphar, Mr Anil Sharma, Mrs Lisa Grass, Mr Samuel Bremang, Mrs Amarjit Badhan, Ms Hiromi Uzu, Ms Marie Francis, Miss Helen Piotrowski, Ms Marie O'Donoghue and Ms Irene Zondo
University Hospitals of Leicester NHS Trust	Dr Gerrit Woltmann and Dr Martin Wiselka	Mrs Kate Ellis, Mrs Hilary Pateman, Mrs Linda Mashonganyika and Mr Adam Lewszuk
Sandwell and West Birmingham Hospitals NHS Trust	Dr Naz Nathani	Mrs Harriet Goddard and Mrs Frances Lloyd
Heart of England NHS Foundation Trust	Dr Martin Dedicoat and Dr Heinke Kunst (co-investigator)	Ms Bridget Pointon and Ms Mary O'Sullivan
Oxford University Hospitals NHS Foundation Trust	Professor Christopher Conlon	Professor Christopher Conlon and Mr Lee Potiphar
Frimley Health NHS Foundation Trust	Dr Sarah Menzies	Dr John Wiggins, Mrs Amarjit Badhan, Mr Samuel Bremang and Mr Edmund Cox
King's College Hospital NHS Foundation Trust	Dr Frank Post	Dr Frank Post, Ms Emily Wandolo and Ms Lucy Campbell
Barts Health NHS Trust	Dr Rebecca O'Connell	Mrs Nyasha Makoka, Mr Andrew Crawford-Jones and Ms Kaya Widuch

Centre	PI	Staff that recruited patients and performed data collection
St George's University Hospitals NHS Foundation Trust	Professor Derek Macallan and Dr Felix Chua (co-investigator)	Dr Carmen Cabeza, Miss Louise Wootton, Mr Gary Hahn, Mr Lee Potiphar and Mrs Amarjit Badhan
Chelsea and Westminster Hospital NHS Foundation Trust	Dr Anton Pozniak and Dr Mike Loebingher (co-investigator)	Mr Gary Hahn, Mr Lee Potiphar, Mr Anil Sharma, Mrs Amarjit Badhan, Ms Sanjeewa Basnayake, Mrs Lesley Ruta and Miss Kirsty Money
Royal Free London NHS Foundation Trust	Dr Marc Lipman	Dr Simon Bax, Mr Gary Hahn, Miss Janey Sewell, Mr Anil Sharma, Mr Lee Potiphar, Mrs Amarjit Badhan and Mrs Angelita Solamalai
London North West Healthcare NHS Trust	Dr Rob Davison and Dr David Abdoyeku (co-investigator)	Dr Laurence John, Dr Jim Buckley, Miss Linda Haimbodi, Mr Lee Potiphar, Mr Gary Hahn, Mr Anil Sharma, Mrs Amarjit Badhan, Mr Samuel Bremang and Ms Hiromi Uzu
Ealing Hospital for follow-up	Dr Howard Branley and Dr William Lynn	Mr Samuel Bremang

Independent members of the Study Steering Committee

- Professor Khalid Khan (Chairperson of the IDEA Steering Committee and Professor of Women's Health and Clinical Epidemiology), Barts and the London School of Medicine and Dentistry, London, UK.
- Professor Stephen Gordon (Professor of Respiratory Medicine and Director), Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Queen Elizabeth Central Hospital, Blantyre, Malawi and the Clinical Research Group, Liverpool School of Tropical Medicine, Liverpool, UK.
- Dr James Gray (Consultant Microbiologist), Department of Microbiology, Birmingham Children's Hospital NHS Foundation Trust, Birmingham, UK.
- Dr Johannes B Reitsma (Clinical Epidemiologist), Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands.
- Ms Nisha Karnani (PPI member).

Members of the Data Management Group

Professor Ajit Lalvani, Professor Onn Min Kon, Professor Jon Deeks, Dr Yemisi Takwoingi, Dr Melanie Rees-Roberts, Mr Lee Potiphar, Mr Anil Sharma, Dr Hilary Whitworth, Mrs Amarjit Badhan, Miss Aime Boakye and Dr Christopher Partlett.

Members of the Study Management Group

Professor Ajit Lalvani, Professor Onn Min Kon, Dr Melanie Rees-Roberts, Mr Lee Potiphar, Dr Hilary Whitworth, Mr Anil Sharma and Mrs Amarjit Badhan.

The IDEA study collaborators

Dr Yemisi Takwoingi, Dr Hilary Whitworth, Dr Melanie Rees-Roberts, Mrs Amarjit Badhan, Dr Christopher Partlett, Dr Nathan Green, Miss Aime Boakye, Miss Heather Lambie, Mr Luigi Marongiu, Dr Mark Jit, Dr Peter White, Professor Jonathan Deeks, Professor Onn Min Kon, Dr Ajit Lalvani, Dr David Abdoyeku, Dr Howard Branley, Professor Felix Chua, Professor Christopher Conlon, Dr Graham Cooke, Professor Robert Davison, Dr Martin Dedicoat, Dr Heinke Kunst, Dr Marc Lipman, Dr Mike Loebingher, Dr William Lynn, Professor Derek Macallan, Dr Sarah Menzies, Dr Nazim Nathani, Dr Rebecca O'Connell, Dr Frank Post, Dr Anton Pozniak, Dr Martin Wiselka and Dr Gerrit Woltmann.

Contributions of authors

Dr Yemisi Takwoingi (Senior Study Statistician) led the statistical analysis, contributed to the interpretation of the diagnostic accuracy results, led the writing of *Chapters 2–5* and produced the report.

Dr Hilary Whitworth (former Study Co-ordinator and Post-Doctoral Research Associate) was responsible for day-to-day management of the IDEA study and, management of the lead research nurse, writing of *Chapter 1* and contributed to the writing of the report.

Dr Melanie Rees-Roberts (former IDEA Study Co-ordinator) contributed to the management of the IDEA study and built the study databases.

Mrs Amarjit Badhan (Senior Clinical Research Nurse) was responsible for the day-to-day management of the IDEA study with the study co-ordinator and for the management of the nursing team, led the writing of *Chapters 1* and *8* and commented on the draft of the report.

Dr Christopher Partlett (Study Statistician) conducted the statistical analysis and contributed to the writing of the report.

Dr Nathan Green (Research Fellow) performed statistical analysis of diagnostic pathways and health economic modelling, and led the writing of *Chapters 6* and *7*.

Miss Aime Boakye (IDEA Administrator plus laboratory role) contributed to database management, data collection, data cleaning, quality assurance and the writing of *Chapter 2*.

Miss Heather Lambie (Laboratory Staff) performed blood and BAL processing, conducted index tests and entered laboratory data.

Mr Luigi Marongiu (Research Assistant) contributed to the management and analysis of blood samples and database management.

Dr Mark Jit (Senior Mathematical and Economic Modeller, Reader) contributed to the design and interpretation of analysis of diagnostic pathways and the health economic analysis.

Dr Peter White (Reader, Head of Public Health England Modelling and Economics Unit) contributed to the study design and led the statistical analysis of diagnostic pathways and the economic analysis.

Professor Jonathan J Deeks (Professor of Biostatistics and Co-applicant) contributed to the study design and analysis, the interpretation of results and the writing of the report.

Professor Onn Min Kon (Professor and Consultant Physician Imperial College Healthcare) was chief investigator, provided clinical expertise, and contributed to study design, execution and the writing of the report.

Professor Ajit Lalvani (Professor and Consultant Physician Imperial College Healthcare) was chief investigator, provided clinical expertise, and contributed to the study design, execution and the writing of the report.

Publications

Abubakar I, Stagg HR, Whitworth H, Lalvani A. How should I interpret an interferon gamma release assay result for tuberculosis infection? *Thorax* 2013;**68**:298–301.

Whitworth HS1, Scott M, Connell DW, Dongés B, Lalvani A. IGRAs – the gateway to T cell based TB diagnosis. *Methods* 2013;**61**:52–62.

Whitworth HS, Badhan A, Boakye AA, Takwoingi Y, Rees-Roberts M, Partlett C, *et al.* Clinical utility of existing and second-generation interferon-γ release-assays for diagnostic evaluation of tuberculosis: an observational cohort study. *Lancet Infect Dis* 2019;**19**:193–202.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to available anonymised data may be granted following review.

References

1. World Health Organization. *Global Tuberculosis Report 2015*. 20th edn. Geneva: World Health Organization; 2015.
2. Public Health England. *Tuberculosis in England 2015. Report Version 1.1*. London: Public Health England; 2015.
3. Dosanjh DP, Hinks TS, Innes JA, Deeks JJ, Pasvol G, Hackforth S, *et al.* Improved diagnostic evaluation of suspected tuberculosis. *Ann Intern Med* 2008;**148**:325–36. https://doi.org/10.7326/0003-4819-148-5-200803040-00003
4. Health Protection Agency Centre for Infections. *Tuberculosis in the UK: Annual Report on Tuberculosis Surveillance in the UK 2008*. London: Health Protection Agency; 2008.
5. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999;**284**:1520–3. https://doi.org/10.1126/science.284.5419.1520
6. Whitworth HS, Scott M, Connell DW, Dongés B, Lalvani A. IGRAs – the gateway to T cell based TB diagnosis. *Methods* 2013;**61**:52–62. https://doi.org/10.1016/j.ymeth.2012.12.012
7. Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Ann Intern Med* 2008;**149**:177–84. https://doi.org/10.7326/0003-4819-149-3-200808050-00241
8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;**351**:h5527. https://doi.org/10.1136/bmj.h5527
9. Oxford Immunotec. T-SPOT.*TB Package Insert*. URL: www.oxfordimmunotec.com/international/wp-content/uploads/sites/3/PI-TB-IVD-UK-V2.pdf (accessed 4 April 2016).
10. Qiagen. *QuantiFERON TB Gold (QFT). ELISA Package Insert*. 2009. URL: www.quantiferon.com/irm/content/PI/QFT/2PK/UK.pdf (accessed 4 April 2016).
11. NICE. *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control*. Clinical Guideline 117. London: NICE; 2011.
12. Wang L, Turner MO, Elwood RK, Schulzer M, FitzGerald JM. A meta-analysis of the effect of Bacille Calmette Guérin vaccination on tuberculin skin test measurements. *Thorax* 2002;**57**:804–9. https://doi.org/10.1136/thorax.57.9.804
13. Goletti D, Carrara S, Stefania C, Butera O, Amicosante M, Ernst M, *et al.* Accuracy of immunodiagnostic tests for active tuberculosis using single and combined results: a multicenter TBNET-Study. *PLOS ONE* 2008;**3**:e3417. https://doi.org/10.1371/journal.pone.0003417
14. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med* 2002;**21**:835–52. https://doi.org/10.1002/sim.1058
15. Santin M, Muñoz L, Rigau D. Interferon-γ release assays for the diagnosis of tuberculosis and tuberculosis infection in HIV-infected adults: a systematic review and meta-analysis. *PLOS ONE* 2012;**7**:e32482. https://doi.org/10.1371/journal.pone.0032482
16. Rice B, Elford J, Yin Z, Kruijshaar M, Abubakar I, Lipman M, *et al.* Decreasing incidence of tuberculosis among heterosexuals living with diagnosed HIV in England and Wales. *AIDS* 2013;**27**:1151–7. https://doi.org/10.1097/QAD.0b013e32835e2cb1

17. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927;**22**:209–12. https://doi.org/10.1080/01621459.1927.10502953

18. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. *Stat Sci* 2001;**16**:101–17. https://doi.org/10.1214/ss/1009213286

19. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;**44**:763–70. https://doi.org/10.1016/0895-4356(91)90128-V

20. Leisenring W, Pepe MS, Longton G. A marginal regression modelling framework for evaluating medical diagnostic tests. *Stat Med* 1997;**16**:1263–81. https://doi.org/10.1002/(SICI)1097-0258(19970615)16:11<1263::AID-SIM550>3.0.CO;2-M

21. Yen YF, Yen MY, Lin YS, Lin YP, Shih HC, Li LH, *et al.* Smoking increases risk of recurrence after successful anti-tuberculosis treatment: a population-based study. *Int J Tuberc Lung Dis* 2014;**18**:492–8. https://doi.org/10.5588/ijtld.13.0694

22. Chan SF, Deeks JJ, Macaskill P, Irwig L. Three methods to construct predictive models using logistic regression and likelihood ratios to facilitate adjustment for pretest probability give similar results. *J Clin Epidemiol* 2008;**61**:52–63. https://doi.org/10.1016/j.jclinepi.2007.02.012

23. Knottnerus JA. Application of logistic regression to the analysis of diagnostic data: exact modeling of a probability tree of multiple binary variables. *Med Decis Making* 1992;**12**:93–108. https://doi.org/10.1177/0272989X9201200202

24. World Medical Association. *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA* 2013;**310**:2191–4. https://doi.org/10.1001/jama.2013.281053

25. Great Britain. *Data Protection Act 1998*. Chapter 29. London: The Stationery Office; 1998.

26. Rose AM, Watson JM, Graham C, Nunn AJ, Drobniewski F, Ormerod LP, *et al.* Tuberculosis at the end of the 20th century in England and Wales: results of a national survey in 1998. *Thorax* 2001;**56**:173–9. https://doi.org/10.1136/thorax.56.3.173

27. Whitworth HS, Badhan A, Boakye AA, Takwoingi Y, Rees-Roberts M, Partlett C, *et al.* Clinical utility of existing and second-generation interferon-γ release-assays for diagnostic evaluation of tuberculosis: an observational cohort study. *Lancet Infect Dis* 2019;**19**:193–202. https://doi.org/10.1016/S1473-3099(18)30613-3

28. Aggarwal AN, Agarwal R, Gupta D, Dhooria S, Behera D. Interferon gamma release assays for diagnosis of pleural tuberculosis: a systematic review and meta-analysis. *J Clin Microbiol* 2015;**53**:2451–9. https://doi.org/10.1128/JCM.00823-15

29. Danel C, Kabran M, Inwoley A, Badje A, Herrmann JL, Moh R, *et al.* Quantiferon-TB Gold: performance for ruling out active tuberculosis in HIV-infected adults with high CD4 count in Côte d'Ivoire, West Africa. *PLOS ONE* 2014;**9**:e107245. https://doi.org/10.1371/journal.pone.0107245

30. Lagrange PH, Thangaraj SK, Dayal R, Deshpande A, Ganguly NK, Girardi E, *et al.* A toolbox for tuberculosis (TB) diagnosis: an Indian multicentric study (2006-2008). Evaluation of QuantiFERON-TB gold in tube for TB diagnosis. *PLOS ONE* 2013;**8**:e73579. https://doi.org/10.1371/journal.pone.0073579

31. Sauzullo I, Mengoni F, Ermocida A, Massetti AP, D'Agostino C, Russo G, *et al.* Interferon-γ release assay in HIV-infected patients with active tuberculosis: impact of antituberculous drugs on host immune response. *New Microbiol* 2014;**37**:153–61.

32. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, *et al.* Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health* 2013;**16**:e1–5. https://doi.org/10.1016/j.jval.2013.02.010

33. Kruijshaar ME, Lipman M, Essink-Bot ML, Lozewicz S, Creer D, Dart S, *et al.* Health status of UK patients with active tuberculosis. *Int J Tuberc Lung Dis* 2010;**14**:296–302.

34. Kind P, Hardman G, Macran S. *UK Population Norms for EQ-5D. Discussion Paper 172*. York: Centre for Health Economics, University of York; 1999.

35. Monitor and NHS England. *National Tariff Payment System 2014/15*. London: Monitor; 2013.

36. Hughes R, Wonderling D, Li B, Higgins B. The cost effectiveness of nucleic acid amplification techniques for the diagnosis of tuberculosis. *Respir Med* 2012;**106**:300–7. https://doi.org/10.1016/j.rmed.2011.10.005

37. Curtis L BA. *Unit Costs of Health & Social Care*. Canterbury: Personal Social Services Research Unit; 2015.

38. Drobniewski F, Cooke M, Jordan J, Casali N, Mugwagwa T, Broda A, *et al.* Systematic review, meta-analysis and economic modelling of molecular diagnostic tests for antibiotic resistance in tuberculosis. *Health Technol Assess* 2015;**19**(34). https://doi.org/10.3310/hta19340

39. NICE. *Tuberculosis: Prevention, Diagnosis, Management and Service Organisation. NICE Guideline 33*. London: NICE; 2016.

40. Auguste P, Tsertsvadze A, Pink J, Court R, Seedat F, Gurung T, *et al.* Accurate diagnosis of latent tuberculosis in children, people who are immunocompromised or at risk from immunosuppression and recent arrivals from countries with a high incidence of tuberculosis: systematic review and economic evaluation. *Health Technol Assess* 2016;**20**(38). https://doi.org/10.3310/hta20380

41. Auguste P, Tsertsvadze A, Pink J, Court R, Seedat F, Gurung T, *et al. Accurate Diagnosis of Latent Tuberculosis in Children, in People Who are Immunocompromised or at Risk from Immunosuppression, and Recent Arrivals from Countries with a High Incidence of Tuberculosis: Systematic Review and Economic Evaluation. Appendix H: Warwick Evidence Diagnosis of LTBI*. London: NICE; 2016. pp. 1–779. URL: www.nice.org.uk/guidance/ng33/evidence/appendix-h-warwick-evidence-diagnosis-of-ltbi-report-80851860832 (accessed 5 January 2017).

42. NICE. *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control. NICE Guideline. Update of CG117 – Appendix 6. Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for Latent Tuberculosis*. London: NICE; 2010. pp. 1–52.

43. Pareek M, Bond M, Shorey J, Seneviratne S, Guy M, White P, *et al.* Community-based evaluation of immigrant tuberculosis screening using interferon γ release assays and tuberculin skin testing: observational study and economic analysis. *Thorax* 2013;**68**:230–9. https://doi.org/10.1136/thoraxjnl-2011-201542

44. NICE. *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis and Measures for its Prevention and Control. NICE Guideline 33*. London: NICE; 2006.

45. Joint Formulary Committee. *British National Formulary*. 69 edn. London: BMJ Group and Pharmaceutical Press; 2016.

46. Pooran A, Booth H, Miller RF, Scott G, Badri M, Huggett JF, *et al.* Different screening strategies (single or dual) for the diagnosis of suspected latent tuberculosis: a cost effectiveness analysis. *BMC Pulm Med* 2010;**10**:7. https://doi.org/10.1186/1471-2466-10-7

47. Faurholt-Jepsen D, Aabye MG, Jensen AV, Range N, Praygod G, Jeremiah K, *et al.* Diabetes is associated with lower tuberculosis antigen-specific interferon gamma release in Tanzanian tuberculosis patients and non-tuberculosis controls. *Scand J Infect Dis* 2014;**46**:384–91. https://doi.org/10.3109/00365548.2014.885657

48. Barcellini L, Borroni E, Brown J, Brunetti E, Codecasa L, Cugnata F, *et al.* First independent evaluation of QuantiFERON-TB Plus performance. *Eur Respir J* 2016;**47**:1587–90. https://doi.org/10.1183/13993003.02033-2015

49. Great Britain. *Human Tissue Act 2004*. London: The Stationery Office; 2004.

50. Adewole OO, Erhabor GE, Sogaolu MO, Onipede AO, Owiafe PK, Awopeju FO, Ota MO. Diagnostic utility of QuantiFERON-TB gold in-tube in active pulmonary tuberculosis in Nigeria. *West Afr J Med* 2013;**32**:180–5.

51. Jeon YL, Nam YS, You E, Yang JJ, Kim MJ, Cho SY, *et al.* Factors influencing discordant results of the QuantiFERON-TB Gold In-tube test in patients with active TB. *J Infect* 2013;**67**:288–93. https://doi.org/10.1016/j.jinf.2013.06.005

52. Jia H, Pan L, Qin S, Liu F, Du F, Lan T, *et al.* Evaluation of interferon-γ release assay in the diagnosis of osteoarticular tuberculosis. *Diagn Microbiol Infect Dis* 2013;**76**:309–13. https://doi.org/10.1016/j.diagmicrobio.2013.03.030

53. Khalil KF, Ambreen A, Butt T. Comparison of sensitivity of QuantiFERON-TB gold test and tuberculin skin test in active pulmonary tuberculosis. *J Coll Physicians Surg Pak* 2013;**23**:633–6. https://doi.org/09.2013/JCPSP.633636

54. Kim JK, Bang WJ, Oh CY, Yoo C, Cho JS. Feasibility of the interferon-γ release assay for the diagnosis of genitourinary tuberculosis in an endemic area. *Korean J Urol* 2013;**54**:123–6. https://doi.org/10.4111/kju.2013.54.2.123

55. Lavender TW, Barrett A, Magee J, Ong EL. Interferon-γ release assays in the diagnosis of active tuberculosis disease in a low-incident setting: a 5-year review of data. *Clin Microbiol Infect* 2013;**19**:1078–81. https://doi.org/10.1111/1469-0691.12129

56. Lei Y, Yi FM, Zhao J, Luckheeram RV, Huang S, Chen M, *et al.* Utility of in vitro interferon-γ release assay in differential diagnosis between intestinal tuberculosis and Crohn's disease. *J Dig Dis* 2013;**14**:68–75. https://doi.org/10.1111/1751-2980.12017

57. Liu F, Gao M, Zhang X, Du F, Jia H, Yang X, *et al.* Interferon-gamma release assay performance of pleural fluid and peripheral blood in pleural tuberculosis. *PLOS ONE* 2013;**8**:e83857. https://doi.org/10.1371/journal.pone.0083857

58. Lodha R, Mukherjee A, Saini D, Saini S, Singh V, Singh S, *et al.* Role of the QuantiFERON®-TB Gold In-Tube test in the diagnosis of intrathoracic childhood tuberculosis. *Int J Tuberc Lung Dis* 2013;**17**:1383–8. https://doi.org/10.5588/ijtld.13.0348

59. Mahomed H, Ehrlich R, Hawkridge T, Hatherill M, Geiter L, Kafaar F, *et al.* Screening for TB in high school adolescents in a high burden setting in South Africa. *Tuberculosis* 2013;**93**:357–62. https://doi.org/10.1016/j.tube.2013.02.007

60. Fei B, Wu Z, Min K, Zhang J, Ding C, Wu H. Interferon-γ release assay in the diagnosis of laryngeal tuberculosis. *Acta Otolaryngol* 2014;**134**:314–7. https://doi.org/10.3109/00016489.2013.850174

61. Garazzino S, Galli L, Chiappini E, Pinon M, Bergamini BM, Cazzato S, *et al.* Performance of interferon-gamma release assay for the diagnosis of active or latent tuberculosis in children in the first 2 years of age: a multicenter study of the Italian Society of Pediatric Infectious Diseases. *Pediatr Infect Dis J* 2014;**33**:e226–31. https://doi.org/10.1097/INF.0000000000000353

62. Kim CH, Kim JY, Hwang YI, Lee CY, Choi JH, Park YB, *et al.* Interferon-γ enzyme-linked immunospot assay in patients with tuberculosis and healthy adults. *Tuberc Respir Dis* 2014;**76**:23–9. https://doi.org/10.4046/trd.2014.76.1.23

63. Kim CH, Lim JK, Yoo SS, Lee SY, Cha SI, Park JY, Lee J. Diagnostic performance of the QuantiFERON-TB Gold In-Tube assay and factors associated with nonpositive results in patients with miliary tuberculosis. *Clin Infect Dis* 2014;**58**:986–9. https://doi.org/10.1093/cid/ciu045

64. Park H, Shin JA, Kim HJ, Ahn CM, Chang YS. Whole blood interferon-γ release assay is insufficient for the diagnosis of sputum smear negative pulmonary tuberculosis. *Yonsei Med J* 2014;**55**:725–31. https://doi.org/10.3349/ymj.2014.55.3.725

65. Schopfer K, Rieder HL, Bodmer T, Steinlin-Schopfer JF, Chantana Y, Studer P, *et al.* The sensitivity of an interferon-γ release assay in microbiologically confirmed pediatric tuberculosis. *Eur J Pediatr* 2014;**173**:331–6. https://doi.org/10.1007/s00431-013-2161-x

66. Wang X, Wu Y, Wang M, Wang Y. The sensitivity of T-SPOT.*TB* assay in diagnosis of pediatric tuberculosis. *Fetal Pediatr Pathol* 2014;**33**:123–5. https://doi.org/10.3109/15513815.2013.878010

67. Wlodarczyk M, Rudnicka W, Janiszewska-Drobinska B, Kielnierowski G, Kowalewicz-Kulbat M, Fol M, Druszczynska M. Interferon-gamma assay in combination with tuberculin skin test are insufficient for the diagnosis of culture-negative pulmonary tuberculosis. *PLOS ONE* 2014;**9**:e107208. https://doi.org/10.1371/journal.pone.0107208

68. Anwar A, Hamdan AJ, Salim B, Yosra A, Hani M, Abdullah AH. Diagnostic utility of QuantiFERON-TB Gold (QFT-G) in active pulmonary tuberculosis. *J Glob Infect Dis* 2015;**7**:108–12. https://doi.org/10.4103/0974-777X.162231

69. Bao L, Li T, Diao N, Shen Y, Shao L, Zhang Y, *et al.* Fluctuating behavior and influential factors in the performance of the QuantiFERON-TB Gold In-Tube Assay in the diagnosis of tuberculosis. *PLOS ONE* 2015;**10**:e0103763. https://doi.org/10.1371/journal.pone.0103763

70. Sali M, Buonsenso D, Goletti D, D'Alfonso P, Zumbo A, Fadda G, *et al.* Accuracy of QuantiFERON-TB gold test for tuberculosis diagnosis in children. *PLOS ONE* 2015;**10**:e0138952. https://doi.org/10.1371/journal.pone.0138952

71. Shin JA, Chang YS, Kim HJ, Ahn CM, Byun MK. Diagnostic utility of interferon-gamma release assay in extrapulmonary tuberculosis. *Diagn Microbiol Infect Dis* 2015;**82**:44–8. https://doi.org/10.1016/j.diagmicrobio.2015.02.002

72. Sun L, Tian JL, Yin QQ, Xiao J, Li JQ, Guo YJ, *et al.* Performance of the interferon gamma release assays in tuberculosis disease in children five years old or less. *PLOS ONE* 2015;**10**:e0143820. https://doi.org/10.1371/journal.pone.0143820

73. Wong KS, Huang YC, Hu HC, Huang YC, Wen CH, Lin TY. Diagnostic utility of QuantiFERON-TB Gold In-Tube test in pediatric tuberculosis disease in Taiwanese children. *J Microbiol Immunol Infect* 2017;**50**:349–54. https://doi.org/10.1016/j.jmii.2015.07.012

74. Xia H, Wang X, Li F, Longuet C, Vernet G, Goletti D, *et al.* Diagnostic values of the QuantiFERON-TB Gold In-tube assay carried out in China for diagnosing pulmonary tuberculosis. *PLOS ONE* 2015;**10**:e0121021. https://doi.org/10.1371/journal.pone.0121021

75. Uzunhan O, Törün SH, Somer A, Salman N, Köksalan K. Comparison of tuberculin skin test and QuantiFERON®-TB Gold In-Tube for the diagnosis of childhood tuberculosis. *Pediatr Int* 2015;**57**:893–6. https://doi.org/10.1111/ped.12659

76. Azghay M, Bouchaud O, Mechaï F, Nicaise P, Fain O, Stirnemann J. Utility of QuantiFERON-TB Gold In-Tube assay in adult, pulmonary and extrapulmonary, active tuberculosis diagnosis. *Int J Infect Dis* 2016;**44**:25–30. https://doi.org/10.1016/j.ijid.2016.01.004

77. Jia H, Pan L, Du B, Sun Q, Wei R, Xing A, *et al.* Diagnostic performance of interferon-γ release assay for lymph node tuberculosis. *Diagn Microbiol Infect Dis* 2016;**85**:56–60. https://doi.org/10.1016/j.diagmicrobio.2016.02.001

Appendix 1 Reporting checklist for diagnostic accuracy studies

TABLE 51 The Standards for Reporting Diagnostic Accuracy Studies (STARD) checklist

Section and topic	Number	Item	Reported on page number
Title or abstract			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values or AUC)	vii, viii
Abstract			
	2	Structured summary of study design, methods, results and conclusions	vii, viii
Introduction			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	1, 2
	4	Study objectives and hypotheses	2
Methods			
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	3
Participants	6	Eligibility criteria	3
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	3
	8	Where and when potentially eligible participants were identified (setting, location and dates)	3, 15
	9	Whether participants formed a consecutive, random or convenience series	3, *Figure 1*
Test methods	10a	Index test, in sufficient detail to allow replication	5–8
	10b	Reference standard, in sufficient detail to allow replication	8, 9
	11	Rationale for choosing the reference standard (if alternatives exist)	Not applicable
	12a	Definition of and rationale for test positivity cut-off points or result categories of the index test, distinguishing pre-specified from exploratory	6–8
	12b	Definition of and rationale for test positivity cut-off points or result categories of the reference standard, distinguishing pre-specified from exploratory	*Appendix 2*
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	5
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	8, 9

continued

TABLE 51 The Standards for Reporting Diagnostic Accuracy Studies (STARD) checklist (*continued*)

Section and topic	Number	Item	Reported on page number
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	10, 11
	15	How indeterminate index test or reference standard results were handled	11
	16	How missing data on the index test and reference standard were handled	11
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	11
	18	Intended sample size and how it was determined	9, 10
Results			
Participants	19	Flow of participants, using a diagram	15, 16, *Figure 4*
	20	Baseline demographic and clinical characteristics of participants	15–21, *Tables 3–7*, *Appendix 4*
	21a	Distribution of severity of disease in those with the target condition	22, 23, *Table 9*
	21b	Distribution of alternative diagnoses in those without the target condition	22, 24, *Table 10*
	22	Time interval and any clinical interventions between index test and reference standard	8
Test results	23	Cross-tabulation of the index test results (or their distribution) by the results of the reference standard	*Tables 11, 13, 20, 22, 37, 38* and *62*
	24	Estimates of diagnostic accuracy and their precision (such as 95% CIs)	25–41, 54–57, *Appendix 13*
	25	Any adverse events from performing the index test or the reference standard	Not applicable
Discussion			
	26	Study limitations, including sources of potential bias, statistical uncertainty and generalisability	86, 87
	27	Implications for practice, including the intended use and clinical role of the index test	87
Other information			
	28	Registration number and name of registry	Not registered
	29	Where the full study protocol can be accessed	www.nets.nihr.ac.uk/__data/assets/pdf_file/0011/51977/PRO-08-106-02.pdf
	30	Sources of funding and other support; role of funders	viii

AUC, area under the curve.

Appendix 2 Composite reference standard for diagnosis of active tuberculosis

TABLE 52 Diagnostic categories for active TB

Diagnostic category	Criteria
1: Culture-confirmed TB	Microbiological culture of *M. tuberculosis* AND suggestive clinical and radiological findings
2: Highly probable TB	Clinical and radiological features highly suggestive of TB and unlikely to be caused by other disease, AND a decision to treat made by a clinician, AND appropriate response to therapy AND histology supportive, if available
3: Clinically indeterminate	Final diagnosis of TB neither highly probable nor reliably excluded
4: Active TB excluded	
Subclassification	
4A: inactive TB	Stable CXR changes, AND TST positive[a] (if done), AND bacteriologically negative (if done) AND no clinical evidence of active disease
4B: one or more risk factors for TB exposure,[a] TST positive[b]	TST positive,[b] AND bacteriologically negative (if done) AND no clinical evidence of active disease
4C: one or more risk factors for TB exposure,[a] TST negative	History of TB exposure AND TST negative (if done)
4D: no risk factors for TB exposure,[a] TST negative	No history of TB exposure AND TST negative (if done)

CXR, chest radiography.
a Risk factors for TB exposure: recent exposure to active TB patient, born in country of high prevalence or belonging to an ethnic group with a high prevalence of TB (incidence > 100/100,000[26]).
b A TST using the Mantoux test, with a threshold of ≥ 15 mm considered positive.
Diagnostic categories adapted from Dosanjh *et al.*[3]

Appendix 3 Protocol amendments

TABLE 53 Summary of protocol amendments

Amendment number (date)	Details of changes
AM01	
Substantial amendment (April 2011)	1. Changed title of the project to a more memorable acronym after discussion with the project team. The new title was 'IGRA in Diagnostic Evaluation of Active TB' the 'IDEA' project. The protocol, patient information sheet and consent forms were changed accordingly 2. Addition of two new recruiting sites: – Heart of England NHS Foundation Trust with Dr Martin Dedicoat as PI – Sandwell and West Birmingham Hospitals NHS Trust with Dr Naz Nathani as PI A site-specific information form was submitted for each site and R&D approval was obtained before commencing recruitment at the sites 3. Because of staff retirement/staff changes, the PIs at three sites changed after our original submission for ethics approval
AM02	
Substantial amendment (October 2011)	1. Correction of some grammatical and spelling errors in the protocol and patient information sheet 2. Updates to the contact information in the protocol and patient information sheet to include the study co-ordinator's details 3. The following changes were made to the protocol to provide further clarification to sites: – Insertion of a summary of the eligibility criteria to ensure that this stands out. Eligibility criteria were previously contained in the text of the protocol – Addition of table 2 and further text to clarify research sampling time points and time windows for collection of samples. Baseline research samples must be taken after informed consent, but no later than 48 hours after the start of treatment for TB or within 7 days of consent, whichever occurs sooner. Follow-up research samples will be taken at 2 months (± 7 days) and at 6 months (± 7 days) after recruitment – Clarification that surplus BAL samples will only be collected at Imperial College NHS Healthcare Trust, St Mary's Hospital – Clarification of the adverse event and serious adverse event reporting procedures to the study co-ordinator 4. The following changes were made to the patient information sheet: – Insertion of local research nurse contact details to allow patients to gain further information on the study. This is in addition to the PI's details currently present – The following sentence, '*This provision does not apply to claims which arise as a result of HIV/AIDS or any related conditions. This does not affect your legal rights to seek compensation*', was added to correct the sponsor's indemnity and insurance information to reflect that HIV-positive patients will be recruited into the study 5. Update to the consent form to reflect the change in the patient information sheet version and date

continued

TABLE 53 Summary of protocol amendments (*continued*)

Amendment number (date)	Details of changes
AM03	
Substantial amendment (October 2012)	1. Clarification in the protocol that participants with previous history of TB (including previous TB treatment) are eligible for recruitment 2. Inclusion of instructions to investigators that the minimum diagnostic tests carried out for diagnosing TB should follow NICE and local guidelines in order to ensure that all sites assess patients for active TB in the same manner. This was added to section 5.1 under patient recruitment 3. As a patient's diagnosis is not known at the point of recruitment, clarified guidance was added to section 5.4 of the protocol to assist sites in patient follow-up 4. Extension of the follow-up visit time frame to ± 21 days from a specified time point for the 2-month follow-up and ± 8 days from a specified time point for the 6-month follow-up 5. Addition of a data collection section to summarise the data collection forms used in the study 6. Clarification of patient withdrawal as withdrawal of consent only. Patients with a non-TB diagnosis are not considered as withdrawn from follow-up. Information was provided on how to report the inability to follow-up patients with a non-TB diagnosis 7. Addition of information about collection of samples of *Mycobacteria* from culture-positive diagnostic tests. These samples will be bacteria grown in diagnostic tests. As these samples are not classified as 'relevant material' under the Human Tissue Act[49] and are not samples directly from the patient, additional consent to collect this type of sample is not required for participants enrolled in the IDEA study 8. Details of the study's clinical panel that will be responsible for reviewing and confirming all patient diagnosis outcomes (see protocol section 9.3) 9. The details of the DMC were altered to provide details of the DMG, which will oversee data collection and disease prevalence. As there is no reason to stop the study on safety grounds and as there will not be any interim analyses, the DMG was formed as a more appropriate method of oversight of the data collected instead of a DMC
AM04	
Non-substantial (February 2013)	1. Minor typographical errors were corrected 2. Addition of Ealing Hospital NHS Trust as a site in the study in order to obtain clinical information for patients recruited during diagnostic procedures performed under referral at one of our existing sites. The intention is to open this site in order to obtain clinical details of the patient's final diagnosis and treatment in these cases. Recruitment of new participants will also occur at this site 3. Participant recruitment will be extended in order to achieve the 200 HIV-positive, suspected active TB cases initially set out in the protocol. This will result in the overall sample size increasing beyond 1012 (HIV and non-HIV infected participants). Extended recruitment will occur at a reduced number of sites until recruitment targets are achieved. The following sites will close to recruitment on 31 April 2013 and continue the follow-up phase until the end of the study: (a) University Hospitals of Leicester NHS Trust (b) Oxford University Hospitals NHS Trust (c) St George's Healthcare NHS Trust (d) Frimley Health NHS Foundation Trust (e) Sandwell and West Birmingham Hospitals NHS Trust 4. Sites remaining open and continuing to recruit as normal include: (a) London North West Healthcare NHS Trust (b) Imperial College Healthcare NHS Trust (c) Royal Free London NHS Foundation Trust

TABLE 53 Summary of protocol amendments (*continued*)

Amendment number (date)	Details of changes
AM05	
Substantial amendment (October 2013)	1. From preliminary data, the proportion of HIV-positive patients with a final diagnosis of active TB is lower than anticipated. Thus, the revised required population size for this subgroup of HIV-positive patients with suspected active TB is 390, as detailed in a new paragraph in section 9.2 2. The study was extended by 12 months to achieve this increase in sample size. The study duration stated in the protocol was therefore increased from 3 to 4 years 3. A number of existing IDEA study sites will stay open for the extension. In addition, four new sites were added: – University College London Hospitals NHS Foundation Trust (PI: Dr Robert Miller) – King's College Hospital NHS Foundation Trust (PI: Dr Frank Post) – Barts Health NHS Trust (covering St. Bartholomew's hospital) (PI: Dr Guy Baily) – Lewisham and Greenwich NHS Trust (covering Queen Elizabeth Hospital) (PI: Dr Palchaudhuri Paramita) 4. Dr Melanie Scott, the study co-ordinator, went on maternity leave and so contact details were added for Dr Hilary Whitworth (maternity cover) 5. Dr Howard Branley was added as the PI for Ealing Hospital NHS Trust and the reasons for addition of the site were clarified
AM06	
Substantial amendment (April 2014)	1. The lead Research Nurse was changed from Dr Lee Potiphar to Mrs Amarjit Badhan 2. The PI at Barts NHS Trust was changed from Dr Guy Bailey to Dr Rebecca O'Connell 3. The following amendment was made to section 5.3: *In some circumstances, other appropriately qualified staff may carry out recruitment procedures usually performed by a Research Nurse (identifying patients, taking informed consent, taking blood, completing CRFs)* 4. Minor typographical errors were amended
AM07	
Substantial amendment (February 2015)	1. The following amendments were made to participant follow-up in section 5.4: – For some patients, additional follow-up data may be collected for up to 2 years if requested by expert clinical panel) – Imperial College London, TB Research Centre will hold copies of all consent forms for patients recruited to the IDEA study and hold their personal-identifiable data. This will be stored in a locked cabinet in a secured room with limited access – Recurrence of TB cases within 2 years will be verified from the London Tuberculosis Register and Enhanced Tuberculosis Surveillance 2. The following amendment was made to data collection in section 5.5: *The study team will retrospectively collate data on patient hospital admissions, including dates of admission and discharge. This data will be collected by research nurses using patient notes and hospital inpatient records* 3. The study was extended by 10 months to allow for the health economic analysis to be performed. The revised study end date was 31 December 2015 4. The PI at Ealing Hospital NHS Trust was changed from Dr Howard Branley to Dr William Lynn 5. The study co-ordinator was changed to Dr Hilary Whitworth 6. Minor typographical errors were amended

AIDS, acquired immunodeficiency syndrome; DMC, Data Monitoring Committee.

Appendix 4 Country of birth of patients

TABLE 54 Country of birth of all patients included in the analyses

Country of birth	Dosanjh category, *n*				Total, *N*
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Afghanistan	3	0	0	3	6
Algeria	0	0	0	2	2
Angola	1	0	0	1	2
Antigua and Barbuda	0	0	0	1	1
Argentina	0	0	0	2	2
Bangladesh	4	1	1	13	19
Belarus	0	0	0	1	1
Belgium	1	0	0	0	1
The Plurinational State of Bolivia	1	0	0	1	2
Brazil	0	0	0	6	6
Burundi	2	0	0	0	2
Cameroon	0	0	0	1	1
Chile	0	0	0	1	1
China	1	0	0	0	1
Colombia	1	0	0	0	1
Democratic Republic of the Congo	1	0	0	2	3
Cyprus	0	0	0	3	3
Denmark	0	0	0	1	1
Djibouti	0	0	0	1	1
Ecuador	0	0	0	1	1
Egypt	0	0	0	1	1
Eritrea	5	2	3	3	13
Estonia	0	0	0	1	1
Ethiopia	2	2	0	4	8
France	0	0	0	2	2
The Gambia	0	0	0	1	1
Germany	0	0	0	2	2
Ghana	0	1	0	6	7
Grenada	1	0	0	0	1
Guinea-Bissau	0	0	0	1	1

continued

TABLE 54 Country of birth of all patients included in the analyses (*continued*)

Country of birth	Dosanjh category, *n*				Total, *N*
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
Hong Kong	1	0	0	1	2
India	117	42	5	62	226
Indonesia	1	0	0	0	1
Iran	0	0	0	5	5
Iraq	1	0	0	4	5
Ireland	4	1	0	6	11
Italy	0	0	1	2	3
Jamaica	5	1	2	6	14
Kazakhstan	1	0	0	0	1
Kenya	5	2	0	19	26
Kuwait	2	0	1	0	3
Libya	1	0	0	0	1
Lithuania	0	0	0	1	1
Malawi	1	2	1	1	5
Mauritius	0	0	0	1	1
Morocco	1	0	0	4	5
Mozambique	0	0	0	1	1
Nepal	7	5	0	5	17
Niger	0	0	0	1	1
Nigeria	7	1	0	4	12
Pakistan	14	8	4	27	53
Philippines	9	5	1	6	21
Poland	1	1	1	10	13
Portugal	1	0	0	2	3
Romania	3	0	2	1	6
Saudi Arabia	0	0	0	1	1
Sierra Leone	0	0	0	3	3
Somalia	10	9	3	16	38
South Africa	1	2	1	5	9
Spain	0	0	1	0	1
Sri Lanka	7	0	1	13	21
Sudan	1	0	0	4	5
Swaziland	0	1	0	1	2
Sweden	0	0	0	1	1
Switzerland	0	0	0	1	1
Syria	0	0	0	1	1

TABLE 54 Country of birth of all patients included in the analyses (*continued*)

Country of birth	Dosanjh category, *n*				Total, *N*
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	
United Republic of Tanzania	1	0	1	1	3
Thailand	0	1	0	2	3
Uganda	2	0	1	8	11
UK	32	11	12	138	193
USA	0	2	0	1	3
Uruguay	0	0	0	1	1
Yemen	0	0	0	1	1
Zambia	1	0	0	1	2
Zimbabwe	1	2	1	10	14
Total	261	102	43	439	845

Appendix 5 Thresholds used by centres for defining vitamin D status

TABLE 55 Definition of vitamin D status by recruiting centre

	Vitamin D status		
Hospital trust	Deficient	Insufficient	Normal
Imperial College Healthcare NHS Trust	< 40 nmol/l	40–70 nmol/l	70–150 nmol/l
Heart of England NHS Foundation Trust	< 30 nmol/l	Not specified	> 49.9 nmol/l
Chelsea and Westminster Hospital NHS Foundation Trust	< 40 nmol/l	40–70 nmol/l	70–150 nmol/l
Royal Free London NHS Foundation Trust	< 25 nmol/l	25–75 nmol/l	> 75 nmol/l
St George's University Hospitals NHS Foundation Trust	< 50 nmol/l	Not specified	50–200 nmol/l
Frimley Health NHS Foundation Trust	< 50 nmol/l	Not specified	50–150 nmol/l
University Hospitals of Leicester NHS Trust	< 25 nmol/l	25–50 nmol/l	> 50 nmol/l
London North West Healthcare NHS Trust	< 12.5 nmol/l	12.5–50 nmol/l	50–140 nmol/l
Oxford University Hospitals NHS Foundation Trust	< 50 nmol/l	Not specified	> 50 nmol/l
Sandwell and West Birmingham Hospitals NHS Trust	< 30 nmol/l	30–50 nmol/l	> 50 nmol/l
King's College Hospital NHS Foundation Trust	< 20 µg/l	Not specified	20–50 µg/l
Barts Health NHS Trust	< 30 nmol/l	30–50 nmol/l	80–150 nmol/l
Ealing Hospital	< 25 nmol/l	25–50 nmol/l	51–163 nmol/l

Appendix 6 Interferon gamma release assays and tuberculin skin test performed in routine workup of active tuberculosis: main study cohort

TABLE 56 Hospital trusts performing T-SPOT.*TB*, QFT-GIT and/or the TST in the diagnostic workup of active TB in all patients

Hospital trust	Tests performed, *n*			Number of patients[a]
	T-SPOT.*TB*	QFT-GIT	TST	
Imperial College Healthcare NHS Trust	125	14	87	238
Heart of England NHS Foundation Trust	29	28	21	83
Chelsea and Westminster Hospital NHS Foundation Trust	26	0	0	40
Royal Free London NHS Foundation Trust	1	2	0	41
St George's University Hospitals NHS Foundation Trust	29	1	29	43
University Hospitals of Leicester NHS Trust	4	22	0	100
London North West Healthcare NHS Trust	2	2	191	257
Oxford University Hospitals NHS Foundation Trust	0	0	0	2
Sandwell and West Birmingham Hospitals NHS Trust	6	2	8	41
Total	222	71	336[b]	845

a Number of patients included from each centre in the analyses of the IDEA study.
b TST results available for 322 of the 336 patients. Results were missing for four, eight and two patients from Imperial College Healthcare NHS Trust, London North West Healthcare NHS Trust and Sandwell and West Birmingham Hospitals NHS Trust, respectively.

Appendix 7 Additional T-SPOT.*TB* and QFT-GIT results in all patients in the main study cohort

TABLE 57 Cross-tabulation of T-SPOT.*TB* and QFT-GIT results in all patients in main study cohort

		T-SPOT.*TB*, *n* (%)					
		Positive	Negative	Borderline	Indeterminate	Missing	Total
QFT-GIT, *n* (%)	Positive	234 (73.4)	45 (11.3)	9 (27.3)	13 (22.8)	7 (19.4)	308 (36.4)
	Negative	65 (20.4)	307 (76.8)	20 (60.6)	35 (61.4)	6 (16.7)	433 (51.2)
	Indeterminate	19 (6.0)	45 (11.3)	4 (12.1)	9 (15.8)	2 (5.6)	79 (9.3)
	Missing	1 (0.3)	3 (0.8)	0 (0.0)	0 (0.0)	21 (58.3)	25 (3.0)
	Total	319 (100)	400 (100)	33 (100)	57 (100)	36 (100)	845 (100)

TABLE 58 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT: sensitivity analyses with borderline T-SPOT.*TB* results excluded

	T-SPOT.*TB*		QFT-GIT	
Test performance	*n/N*	Estimate (95% CI)	*n/N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	253/311	81.4 (76.6 to 85.3)	220/327	67.3 (62.0 to 72.1)
Culture-positive TB	185/218	84.9 (79.5 to 89.0)	163/231	70.6 (64.4 to 76.1)
Culture-negative TB	58/83	69.9 (59.3 to 78.7)	48/84	57.1 (46.5 to 67.2)
Smear-positive TB	45/55	81.8 (69.7 to 89.8)	42/56	75.0 (62.3 to 84.5)
Smear-negative TB	169/206	82.0 (76.2 to 86.7)	148/222	66.7 (60.2 to 72.5)
Pulmonary TB	79/105	75.2 (66.2 to 82.5)	79/115	68.7 (59.7 to 76.5)
Extrapulmonary TB	141/169	83.4 (77.1 to 88.3)	113/171	66.1 (58.7 to 72.8)
Specificity for a diagnosis of active TB				
Active TB excluded	319/370	86.2 (82.3 to 89.4)	304/378	80.4 (76.1 to 84.1)
Active TB excluded, TST negative, no risk factors for LTBI	87/93	93.5 (86.6 to 97.0)	85/91	93.4 (86.4 to 96.9)
Predictive values				
PPV	253/304	83.2 (78.6 to 87.0)	220/294	74.8 (69.6 to 79.5)
NPV	319/377	84.6 (80.6 to 87.9)	304/411	74.0 (69.5 to 78.0)
Likelihood ratios				
Positive likelihood ratio	–	5.90 (4.55 to 7.66)	–	3.44 (2.76 to 4.27)
Negative likelihood ratio	–	0.22 (0.17 to 0.27)	–	0.41 (0.35 to 0.48)

Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were excluded in these analyses.

For T-SPOT.*TB*, there were 68 test positives out of 93 highly probable TB cases, with a sensitivity (95% CI) of 73.1% (63.3% to 81.1%). For QFT-GIT, there were 57 test positives out of 96 highly probable TB cases, with a sensitivity (95% CI) of 59.4% (49.4% to 68.7%). Indeterminate IGRA results were excluded from all analyses.

TABLE 59 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT: sensitivity analyses with indeterminate IGRA results included

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	287/345	83.2 (78.9 to 86.8)	246/353	69.7 (64.7 to 74.2)
Culture-positive TB	213/246	86.6 (81.8 to 90.3)	184/252	73.0 (67.2 to 78.1)
Culture-negative TB	64/88	72.7 (62.6 to 80.9)	53/89	59.6 (49.2 to 69.1)
Smear-positive TB	53/63	84.1 (73.2 to 91.1)	53/67	79.1 (67.9 to 87.1)
Smear-negative TB	192/229	83.8 (78.5 to 88.0)	159/232	68.2 (62.0 to 73.9)
Pulmonary TB	94/120	78.3 (70.2 to 84.8)	89/125	71.2 (62.7 to 78.4)
Extrapulmonary TB	155/183	84.7 (78.8 to 89.2)	127/185	68.6 (61.6 to 74.9)
Specificity for a diagnosis of active TB				
Active TB excluded	319/423	75.4 (71.1 to 79.3)	304/425	71.5 (67.1 to 75.6)
Active TB excluded, TST negative, no risk factors for LTBI	87/111	78.4 (69.8 to 85.0)	85/110	77.3 (68.6 to 84.1)
Predictive values				
PPV	287/391	73.4 (68.8 to 77.5)	246/367	67.0 (62.1 to 71.6)
NPV	319/377	84.6 (80.6 to 87.9)	304/411	74.0 (69.5 to 78.0)
Likelihood ratios				
Positive likelihood ratio	–	3.38 (2.85 to 4.03)	–	2.45 (2.07 to 2.89)
Negative likelihood ratio	–	0.22 (0.18 to 0.28)	–	0.42 (0.36 to 0.50)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were included in these analyses.

TABLE 60 Comparison of diagnostic accuracy of T-SPOT.*TB* and QFT-GIT: sensitivity analysis with borderline T-SPOT.*TB* results excluded

Test	Number of test results[a]	Sensitivity (95% CI)	Number of test results[b]	Specificity (95% CI)
T-SPOT.*TB*	311	80.7 (76.1 to 84.7)	370	86.5 (82.7 to 89.6)
QFT-GIT	327	67.3 (62.0 to 72.1)	378	81.1 (76.9 to 84.7)
Ratio[c] (95% CI); *p*-value	–	1.20 (1.12 to 1.29); < 0.001	–	1.07 (1.02 to 1.12); 0.004

a Number of test results among those with active TB.
b Number of test results among those without active TB.
c Ratio of the sensitivity (or specificity) of T-SPOT.*TB* to that of QFT-GIT. The natural outputs from GEE models are odds ratios. Ratios of sensitivities (relative sensitivity) and ratios of specificities (relative specificity) were computed post estimation of the models. CIs were obtained using the delta method.

Note
Sensitivities and specificities are presented as percentages.

TABLE 61 Comparison of T-SPOT.*TB* and QFT-GIT: sensitivity analysis with indeterminate IGRA results included

Test	Number of test results[a]	Sensitivity (95% CI)	Number of test results[b]	Specificity (95% CI)
T-SPOT.*TB*	345	83.3 (78.9 to 86.8)	386	75.4 (71.1 to 79.3)
QFT-GIT	353	69.7 (64.7 to 74.3)	378	71.6 (67.1 to 75.7)
Ratio[c] (95% CI); *p*-value	–	1.19 (1.12 to 1.28); < 0.001	–	1.05 (0.98 to 1.13); 0.1

a Number of test results among those with active TB.
b Number of test results among those without active TB.
c The ratio of the sensitivity (or specificity) of T-SPOT.*TB* to that of QFT-GIT. The natural outputs from GEE models are odds ratios. Ratios of sensitivities (relative sensitivity) and ratios of specificities (relative specificity) were computed post estimation of the models. CIs were obtained using the delta method.

Note
Sensitivities and specificities are presented as percentages.

Appendix 8 Additional T-SPOT.*TB* and QFT-GIT results in human immunodeficiency virus-positive and -negative patients in the main study cohort

TABLE 62 T-SPOT.*TB* and QFT-GIT results by active TB status in HIV-positive patients in the main study cohort

Index test result	Dosanjh category								Total
	1	2	3	4A	4B	4C	4D	4A–D	
T-SPOT.*TB*									
Positive	7	5	0	0	0	4	1	5	17
Negative	4	3	2	0	1	42	28	71	80
Borderline	0	0	0	0	0	3	0	0	3
Indeterminate	2	3	0	1	0	11	16	28	33
Missing	0	1	0	0	0	1	0	1	2
Total	13	12	2	1	1	61	45	108	135
Median SFCs ESAT-6 (range)	10 (0–94)	8 (0–395)	0	0	0	0 (0–42)	0 (0–69)	0 (0–69)	0 (0–395)
Median SFCs CFP-10 (range)	2 (0–136)	0 (0–315)	0	0	0	0 (0–37)	0 (0–44)	0 (0–44)	0 (0–315)
QFT-GIT									
Positive	8	5	0	0	0	5	2	7	20
Negative	5	5	1	1	1	42	36	80	91
Indeterminate	0	2	1	0	0	14	6	20	23
Missing	0	0	0	0	0	0	1	1	1
Total	13	12	2	1	1	61	45	108	135
Median IFN-γ levels (range)	0.95 (0–3.78)	0.01 (0–10)	0.01 (0–0.02)	0.07	0	0.01 (0–3.82)	0.01 (0–1.56)	0.01 (0–3.82)	0.02 (0–10)

TABLE 63 Cross-tabulation of T-SPOT.*TB* and QFT-GIT results in HIV-positive patients in main study cohort

		T-SPOT.*TB*, *n* (%)					
		Positive	Negative	Borderline	Indeterminate	Missing	Total
QFT-GIT, *n* (%)	Positive	13 (76.4)	3 (3.8)	0	4 (12.1)	0	20 (14.8)
	Negative	2 (11.8)	60 (75.0)	3 (100.0)	24 (72.7)	2 (100.0)	91 (67.4)
	Indeterminate	2 (11.8)	16 (20.0)	0	5 (15.2)	0	23 (17.0)
	Missing	0	1 (1.3)	0	0	0	1 (0.7)
	Total	17 (100.0)	80 (100.0)	3 (100.0)	33 (100.0)	2 (100.0)	135 (100.0)

TABLE 64 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in HIV-positive patients in the main study cohort: sensitivity analyses with indeterminate IGRA results included

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	17/24	70.8 (50.8 to 85.1)	15/25	60.0 (40.7 to 76.6)
Culture-positive TB	9/13	69.2 (42.4 to 87.3)	8/13	61.5 (35.5 to 82.3)
Culture-negative TB	8/11	72.7 (43.4 to 90.3)	7/11	63.6 (35.4 to 84.8)
Smear-positive TB	2/5	40.0 (11.8 to 76.9)	3/5	60.0 (23.1 to 88.2)
Smear-negative TB	11/15	73.3 (48.0 to 89.1)	9/16	56.2 (33.2 to 76.9)
Pulmonary TB	6/8	75.0 (40.9 to 92.9)	5/8	62.5 (30.6 to 86.3)
Extrapulmonary TB	9/12	75.0 (46.8 to 91.1)	7/13	53.8 (29.1 to 76.8)
Specificity for a diagnosis of active TB				
Active TB excluded	71/107	66.4 (57.0 to 74.6)	80/107	74.8 (65.8 to 82.0)
Active TB excluded, TST negative, no risk factors for LTBI	28/45	62.2 (47.6 to 74.9)	36/44	81.8 (68.0 to 90.5)
Predictive values				
PPV	17/53	32.1 (21.1 to 45.5)	15/42	35.7 (23.0 to 50.8)
NPV	71/78	91.0 (82.6 to 95.6)	80/90	88.9 (80.7 to 93.9)
Likelihood ratios				
Positive likelihood ratio	–	2.11 (1.46 to 3.05)	–	2.38 (1.51 to 3.76)
Negative likelihood ratio	–	0.44 (0.23 to 0.83)	–	0.54 (0.33 to 0.88)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were included in these analyses as test positives.

TABLE 65 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in HIV-negative patients in the main study cohort: sensitivity analyses with indeterminate IGRA results included

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	270/321	84.1 (79.7 to 87.7)	231/328	70.4 (65.3 to 75.1)
Culture-positive TB	204/233	87.6 (82.7 to 91.2)	176/239	73.6 (67.7 to 78.8)
Culture-negative TB	56/77	72.7 (61.9 to 81.4)	46/78	59.0 (47.9 to 69.2)
Smear-positive TB	51/58	87.9 (77.1 to 94.0)	50/62	80.6 (69.1 to 88.6)
Smear-negative TB	181/214	84.6 (79.1 to 88.8)	150/217	69.1 (62.7 to 74.9)
Pulmonary TB	88/112	78.6 (70.1 to 85.2)	84/117	71.8 (63.0 to 79.2)
Extrapulmonary TB	146/171	85.4 (79.3 to 89.9)	120/172	69.8 (62.5 to 76.1)
Specificity for a diagnosis of active TB				
Active TB excluded	248/316	78.5 (73.6 to 82.7)	224/318	70.4 (65.2 to 75.2)
Active TB excluded, TST negative, no risk factors for LTBI	59/66	89.4 (79.7 to 94.8)	49/66	74.2 (62.6 to 83.3)
Predictive values				
PPV	270/338	79.9 (75.3 to 83.8)	231/325	71.1 (65.9 to 75.7)
NPV	248/299	82.9 (78.3 to 86.8)	224/321	69.8 (64.6 to 74.6)
Likelihood ratios				
Positive likelihood ratio	–	3.91 (3.15 to 4.85)	–	2.38 (1.98 to 2.86)
Negative likelihood ratio	–	0.20 (0.16 to 0.26)	–	0.42 (0.35 to 0.50)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were included in these analyses as test positives.

Appendix 9 Additional T-SPOT.*TB* and QFT-GIT results in patients with diabetes mellitus in the main study cohort

TABLE 66 T-SPOT.*TB* and QFT-GIT results by active TB status in patients with diabetes mellitus in the main study cohort

Index test result	Dosanjh category								Total
	1	2	3	4A	4B	4C	4D	4A–D	
T-SPOT.*TB*									
Positive	13	3	2	0	2	6	1	9	27
Negative	7	1	4	1	4	27	6	38	50
Borderline	1	0	0	0	0	2	0	2	3
Indeterminate	1	1	2	0	0	0	0	0	4
Missing	0	0	0	0	0	4	0	4	4
Total	22	5	8	1	6	39	7	53	88
Median SFCs ESAT-6 (range)	8 (0–123)	12 (1–26)	0 (0–4)	0	0 (0–9)	0 (0–83)	0 (0–22)	0 (0–83)	0 (0–123)
Median SFCs CFP-10 (range)	5 (0–275)	51 (2–103)	1 (0–20)	0	0 (0–20)	0 (0–120)	0 (0–4)	0 (0–120)	1 (0–275)
QFT-GIT									
Positive	12	3	2	0	1	8	1	10	27
Negative	10	2	5	1	5	27	4	37	54
Indeterminate	0	0	1	0	0	2	2	4	5
Missing	0	0	0	0	0	2	0	2	2
Total	22	5	8	1	6	39	7	53	88
Median IFN-γ levels (range)	0.39 (0–10)	1.1 (0.11–5.84)	0.07 (0–3.58)	0	0.01 (0–0.84)	0.01 (0–10)	0 (0–5.92)	0 (0–10)	0.07 (0–10)

TABLE 67 Cross-tabulation of T-SPOT.*TB* and QFT-GIT results in patients with diabetes mellitus

		T-SPOT.*TB*, *n* (%)					
		Positive	Negative	Borderline	Indeterminate	Missing	Total
QFT-GIT, *n* (%)	Positive	18 (66.7)	5 (10.0)	1 (33.3)	2 (50.0)	1 (25.0)	27 (30.7)
	Negative	8 (29.6)	41 (82.0)	2 (66.7)	2 (50.0)	1 (25.0)	54 (61.4)
	Indeterminate	1 (3.7)	4 (8.0)	0	0	0	5 (5.7)
	Missing	0	0	0	0	2 (50.0)	2 (2.3)
	Total	27 (100.0)	50 (100.0)	3 (100.0)	4 (100.0)	4 (100.0)	88 (100.0)

TABLE 68 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in patients with diabetes mellitus in main study cohort: sensitivity analyses with indeterminate IGRA results included

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	19/27	70.4 (51.5 to 84.2)	15/27	55.6 (37.3 to 72.4)
Culture-positive TB	15/22	68.2 (47.3 to 83.6)	12/22	54.5 (34.7 to 73.1)
Culture-negative TB	4/5	80.0 (37.6 to 96.4)	3/5	60.0 (23.1 to 88.2)
Smear-positive TB	7/9	77.8 (45.3 to 93.7)	6/9	66.7 (35.4 to 87.9)
Smear-negative TB	11/16	68.8 (44.4 to 85.8)	7/16	43.8 (23.1 to 66.8)
Pulmonary TB	5/9	55.6 (26.7 to 81.1)	5/9	55.6 (26.7 to 81.1)
Extrapulmonary TB	11/14	78.6 (52.4 to 92.4)	8/14	57.1 (32.6 to 78.6)
Specificity for a diagnosis of active TB				
Active TB excluded	38/49	77.6 (64.1 to 87.0)	37/51	72.6 (59.1 to 82.9)
Active TB excluded, TST negative, no risk factors for LTBI	6/7	85.7 (48.7 to 97.4)	4/7	57.1 (25.0 to 84.2)
Predictive values				
PPV	19/30	63.3 (45.5 to 78.1)	15/29	51.7 (34.4 to 68.6)
NPV	38/46	82.6 (69.3 to 90.9)	37/49	75.5 (61.9 to 85.4)
Likelihood ratios				
Positive likelihood ratio	–	3.14 (1.76 to 5.57)	–	2.02 (1.16 to 3.54)
Negative likelihood ratio	–	70.4 (51.5 to 84.2)	–	55.6 (37.3 to 72.4)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were included in these analyses as test positives.

Appendix 10 Additional results for evaluations of second-generation interferon gamma release assay in the main study cohort

TABLE 69 Diagnostic accuracy of second-generation IGRAs: sensitivity analyses based on all patients in the main study cohort

	Antigens											
	Rv3615c		Rv3879c		Rv3873		Rv2654		ESAT-6		CFP-10	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB												
All TB	274/345	79.4 (74.8 to 83.4)	148/345	42.9 (37.8 to 48.2)	135/345	39.1 (34.1 to 44.4)	146/345	42.3 (37.2 to 47.6)	246/345	71.3 (66.3 to 75.8)	253/345	73.3 (68.4 to 77.7)
Culture-positive TB	208/246	84.6 (79.5 to 88.5)	108/246	43.9 (37.8 to 50.2)	100/246	40.7 (34.7 to 46.9)	108/246	43.9 (37.8 to 50.2)	183/246	74.4 (68.6 to 79.4)	189/246	76.8 (71.2 to 81.7)
Culture-negative TB	57/88	64.8 (54.4 to 73.9)	34/88	38.6 (29.1 to 49.1)	31/88	35.2 (26.1 to 45.6)	3388	37.5 (28.1 to 47.9)	53/88	60.2 (49.8 to 69.8)	56/88	63.6 (53.2 to 72.9)
Smear-positive TB	53/63	84.1 (73.2 to 91.1)	28/63	44.4 (32.8 to 56.7)	26/63	41.3 (30.0 to 53.6)	21/63	33.3 (22.9 to 45.6)	44/63	69.8 (57.6 to 79.8)	47/63	74.6 (62.7 to 83.7)
Smear-negative TB	179/229	78.2 (72.4 to 83.0)	103/229	45.0 (38.7 to 51.5)	88/229	38.4 (32.4 to 44.9)	103/229	45.0 (38.7 to 51.5)	167/229	72.9 (66.8 to 78.3)	169/229	73.8 (67.7 to 79.1)
Pulmonary TB	94/120	78.3 (70.1 to 84.8)	44/120	36.7 (28.6 to 45.6)	37/120	30.8 (23.3 to 39.6)	50/120	41.7 (33.2 to 50.6)	80/120	66.7 (57.8 to 74.5)	82/120	68.3 (59.6 to 76.0)
Extra pulmonary TB	144/183	78.7 (72.2 to 84.0)	84/183	45.9 (38.8 to 53.1)	77/183	42.1 (35.2 to 49.3)	76/183	41.5 (34.6 to 48.8)	134/183	73.2 (66.4 to 79.1)	138/183	75.4 (68.7 to 81.1)
Specificity for a diagnosis of active TB												
Active TB excluded	320/423	75.7 (71.3 to 79.5)	359/423	84.9 (81.1 to 88.0)	366/423	86.5 (82.9 to 89.5)	352/423	83.2 (79.4 to 86.5)	339/423	80.1 (76.1 to 83.7)	333/423	78.7 (74.6 to 82.4)
Active TB excluded, TST negative, no risk factors for LTBI	86/111	77.5 (68.9 to 84.3)	90/111	81.1 (72.8 to 87.3)	92/111	82.9 (74.8 to 88.8)	88/111	79.3 (70.8 to 85.8)	87/111	78.4 (69.8 to 85.0)	90/111	81.1 (72.8 to 87.3)

	Antigens											
	Rv3615c		Rv3879c		Rv3873		Rv2654		ESAT-6		CFP-10	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Predictive values												
PPV	274/377	72.7 (68 to 76.9)	148/212	69.8 (63.3 to 75.6)	135/192	70.3 (63.5 to 76.3)	146/217	67.3 (60.8 to 73.2)	246/330	74.5 (69.6 to 78.9)	255/343	74.3 (69.5 to 78.7)
NPV	320/391	81.8 (77.7 to 85.4)	359/556	64.6 (60.5 to 68.4)	366/576	63.5 (59.5 to 67.4)	352/551	63.9 (59.8 to 67.8)	339/438	77.4 (73.3 to 81.1)	333/425	78.4 (74.2 to 82.0)
Likelihood ratios												
Positive likelihood ratio		3.26 (2.73 to 3.89)		2.83 (2.19 to 3.66)		2.93 (2.23 to 3.86)		2.52 (1.97 to 3.22)		3.59 (2.93 to 4.40)		3.45 (2.84 to 4.19)
Negative likelihood ratio		0.27 (0.22 to 0.34)		0.67 (0.61 to 0.74)		0.70 (0.64 to 0.77)		0.69 (0.63 to 0.77)		0.36 (0.30 to 0.43)		0.35 (0.28 to 0.41)

Note
Sensitivity, specificity and predictive values are presented as percentages.

TABLE 70 Diagnostic accuracy of IGRA combinations with borderline excluded: sensitivity analyses based on all patients in main study cohort

	Antigen combinations							
	ESAT-6 + CFP-10 + Rv3615c + Rv3879c		ESAT-6 + CFP-10 + Rv3615c		CFP-10 + Rv3615c + Rv3879c		CFP-10 + Rv3615c	
Test performance	*n/N*	Estimate (95% CI)	*n/N*	Estimate (95% CI)	*n/N*	Estimate (95% CI)	*n/N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB								
All TB	273/306	89.2 (85.2 to 92.2)	273/306	89.2 (85.2 to 92.2)	263/299	88.0 (3.8 to 91.2)	263/301	87.4 (83.1 to 90.7)
Culture-positive TB	203/216	94.0 (90.0 to 96.4)	203/216	94.0 (90.0 to 96.4)	197/211	93.4 (89.2 to 96.0)	197/212	92.9 (88.7 to 95.7)
Culture-negative TB	60/80	75.0 (64.5 to 83.2)	60/80	75.0 (64.5 to 83.2)	57/78	73.1 (62.3 to 81.7)	57/80	71.2 (60.5 to 80.0)
Smear-positive TB	48/51	94.1 (84.1 to 98.0)	48/51	94.1 (84.1 to 98.0)	47/50	94.0 (83.8 to 97.9)	47/50	94.0 (83.8 to 97.9)
Smear-negative TB	183/207	88.4 (83.3 to 92.1)	183/207	88.4 (83.3 to 92.1)	176/202	87.11 (81.8 to 91.1)	176/204	86.3 (80.9 to 90.3)
Pulmonary TB	88/100	88.0 (80.2 to 93.0)	88/100	88.0 (80.2 to 93.0)	85/97	87.6 (79.6 to 92.8)	85/98	86.7 (78.6 to 92.1)
Extra pulmonary TB	148/167	88.6 (82.9 to 92.6)	148/167	88.6 (82.9 to 92.6)	142/164	86.6 (80.5 to 91.0)	142/165	86.1 (80.0 to 90.5)
Specificity for a diagnosis of active TB								
Active TB excluded	290/368	78.8 (74.3 to 82.7)	296/370	80.0 (75.6 to 83.8)	296/372	79.6 (75.2 to 83.4)	302/372	81.2 (76.9 to 84.8)
Active TB excluded, TST negative, no risk factors for LTBI	82/91	90.1 (82.3 to 94.7)	84/92	91.3 (83.8 to 95.5)	84/93	90.3 (82.6 to 94.8)	86/93	92.5 (85.3 to 96.3)
Predictive values								
PPV	273/351	77.8 (73.1 to 81.8)	273/347	78.7 (74.1 to 82.7)	263/339	77.6 (72.8 to 81.7)	263/333	79.0 (74.3 to 83.0)
NPV	290/323	89.8 (86.0 to 92.6)	296/329	90.0 (86.2 to 92.8)	296/332	89.2 (85.4 to 92.1)	302/340	88.8 (85.0 to 91.7)
Likelihood ratios								
Positive likelihood ratio		4.21 (3.44 to 5.15)		4.46 (3.62 to 5.49)		4.31 (3.51 to 5.28)		4.64 (3.74 to 5.76)
Negative likelihood ratio		0.14 (0.10 to 0.19)		0.13 (0.10 to 0.19)		0.15 (0.11 to 0.21)		0.16 (0.12 to 0.21)

Sensitivity, specificity and predictive values are presented as percentages.

For both the four-antigen (ESAT-6, CFP-10, Rv3615c and Rv3879c) and three-antigen (ESAT, CFP-10, Rv3615), there were 70 test positives out of 94 highly probable TB cases with sensitivity (95% CI) of 74.5% (64.8% to 82.2%). For the other three-antigen (CFP-10, Rv3615c, Rv3879c) and two-antigen combination of CFP-10 and Rv3615, there were 66 test positives out of 93 highly probable TB cases with sensitivity (95% CI) of 71.0% (61.1% to 79.2%). Indeterminate IGRA results were excluded from all analyses.

TABLE 71 Diagnostic accuracy of IGRA combinations: sensitivity analyses based on all patients in main study cohort

	Antigen combinations							
	ESAT-6 + CFP-10 + Rv3615c + Rv3879c		ESAT-6 + CFP-10 + Rv3615c		CFP-10 + Rv3615c + Rv3879c		CFP-10 + Rv3615c	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB								
All TB	312/345	90.4 (86.9 to 93.1)	312/345	90.4 (86.9 to 93.1)	309/345	89.6 (85.9 to 92.4)	307/345	89.0 (85.2 to 91.9)
Culture-positive TB	233/246	94.7 (91.2 to 96.9)	233/246	94.7 (91.2 to 96.9)	232/246	94.3 (90.7 to 96.6)	231/246	93.9 (90.2 to 96.3)
Culture-negative TB	68/88	77.3 (67.5 to 84.8)	68/88	77.3 (67.5 to 84.8)	67/88	76.1 (66.3 to 83.8)	66/88	75.0 (65.0 to 82.9)
Smear-positive TB	60/63	95.2 (86.9 to 98.4)	60/63	95.2 (86.9 to 98.4)	60/63	95.2 (86.9 to 98.4)	60/63	95.2 (86.9 to 98.4)
Smear-negative TB	205/229	89.5 (84.9 to 92.9)	205/229	89.5 (84.9 to 92.9)	203/229	88.6 (83.9 to 92.1)	201/229	87.8 (82.9 to 91.4)
Pulmonary TB	108/120	90.0 (83.3 to 94.2)	108/120	90.0 (83.3 to 94.2)	108/120	90.0 (83.3 to 94.2)	107/120	89.2 (82.3 to 93.6)
Extra pulmonary TB	164/183	89.6 (84.4 to 93.3)	164/183	89.6 (84.4 to 93.3)	161/183	88.0 (82.5 to 91.9)	160/183	87.4 (81.8 to 91.5)
Specificity for a diagnosis of active TB								
Active TB excluded	290/423	68.6 (64.0 to 72.8)	296/423	70.0 (65.4 to 74.2)	296/423	70.0 (65.4 to 74.2)	302/423	71.4 (66.9 to 75.5)
Active TB excluded, TST negative, no risk factors for LTBI	82/111	73.9 (65.0 to 81.2)	84/111	75.7 (66.9 to 82.7)	84/111	75.7 (66.9 to 82.7)	83/111	74.8 (66.0 to 81.9)
Predictive values								
PPV	312/445	70.1 (65.7 to 74.2)	312/439	71.1 (66.7 to 75.1)	309/436	70.9 (66.4 to 74.9)	307/428	71.7 (67.3 to 75.8)
NPV	290/323	89.8 (86.0 to 92.6)	296/329	90.0 (86.3 to 92.8)	296/332	89.2 (85.4 to 92.1)	302/340	88.8 (85.0 to 91.8)
Likelihood ratios								
Positive likelihood ratio		2.88 (2.49 to 3.33)		3.01 (2.59 to 3.50)		2.98 (2.57 to 3.47)		3.11 (2.66 to 3.63)
Negative likelihood ratio		0.14 (0.10 to 0.19)		0.14 (0.10 to 0.19)		0.15 (0.11 to 0.20)		0.15 (0.11 to 0.21)

Note
Sensitivity, specificity and predictive values are presented as percentages.

TABLE 72 Comparison of sensitivity of novel IGRA combinations and T-SPOT.*TB*: sensitivity analysis based on all patients in main study cohort

Test	Number of test results	Sensitivity (95% CI)	Relative sensitivity[a] (95% CI)	*p*-value
T-SPOT.*TB* (ESAT-6 + CFP-10)	345	83.3 (79.0 to 86.9)	–	–
CFP-10 + Rv3615c	345	89.0 (85.2 to 91.9)	1.07 (1.03 to 1.11)	< 0.001
CFP-10 + Rv3615c + Rv3879c	345	89.6 (85.9 to 92.4)	1.08 (1.04 to 1.11)	< 0.001
ESAT-6 + CFP-10 + Rv3615c	345	90.4 (86.8 to 93.1)	1.09 (1.05 to 1.12)	< 0.001
ESAT-6 + CFP-10 + Rv3615c + Rv3879c	345	90.5 (86.9 to 93.2)	1.09 (1.05 to 1.12)	< 0.001

a Sensitivity of the combination test divided by the sensitivity of T-SPOT.*TB*.
Note
Indeterminate test results were included as test positives.
Sensitivity is presented as a percentage.

TABLE 73 Comparison of specificity of novel IGRA combinations and T-SPOT.*TB*: sensitivity analysis based on all patients in main study cohort

Test	Number of test results	Specificity (95% CI)	Relative specificity[a] (95% CI)	*p*-value
T-SPOT.*TB* (ESAT-6 + CFP-10)	423	75.4 (71.1 to 79.3)	–	–
CFP-10 + Rv3615c	423	71.4 (66.9 to 75.5)	0.95 (0.92 to 0.98)	0.002
CFP-10 + Rv3615c + Rv3879c	423	70.0 (65.4 to 74.2)	0.93 (0.89 to 0.96)	< 0.001
ESAT-6 + CFP-10 + Rv3615c	423	70.0 (65.4 to 74.2)	0.93 (0.90 to 0.96)	< 0.001
ESAT-6 + CFP-10 + Rv3615c + Rv3879c	423	68.6 (64.0 to 72.8)	0.91 (0.88 to 0.94)	< 0.001

a Specificity of the combination test divided by specificity of the T-SPOT.*TB*.
Note
Indeterminate test results were included as test positives.
Specificity is presented as a percentage.

Appendix 11 Studies of interferon gamma release assays for the diagnosis of active tuberculosis

TABLE 74 Summary of studies of IGRAs for the diagnosis of active TB published between 1 January 2013 and 16 March 2016

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Adewole *et al.*, 2013[50]	Prospective study of QFT-GIT compared with the TST for diagnosis of pulmonary active TB	61 smear-positive TB cases [mean age 35.1 years (SD 4.3 years)] and 41 healthy disease-free controls [mean age 27.8 years (SD 2.1 years)] were enrolled and analysed	Nigeria (full text was unavailable, so unable to extract more information)	NE	Sensitivity: 76.0% (95% CI 61.8% to 85.2%) Specificity: 63.7% (95% CI 46.0% to 76.0%) Indeterminate rate: 3.5%	Sensitivity: 96.6% (95% CI 88.5% to 98.3%) Specificity: 30.0% (95% CI 20.0% to 56.0%)
Jeon *et al.*, 2013[51]	Retrospective analysis of laboratory and clinical records to evaluate factors associated with indeterminate and negative QFT-GIT results in active TB patients	1301 patients including 168 confirmed active TB cases [mean age of active TB cases 54.8 years (SD 20.1 years)]	Kyung Hee University Hospital, Seoul, Korea (September 2009 and April 2012)	NE	Sensitivity: 76.8% (95% CI 69.8% to 82.5%) Specificity: 58.3% (95% CI 55.3% to 61.4%) Indeterminate rate: 8%	NE
Jia *et al.*, 2013[52]	Prospective study of T-SPOT.*TB* for diagnosis of osteoarticular TB	145 patients were enrolled and all were HIV negative. 18 possible cases and 17 indeterminates were excluded. 86 culture-confirmed or probable patients (age range 18–76 years) with osteoarticular TB and 24 without active TB (age range 16–80 years) were analysed	Beijing Chest Hospital, China (July 2011–June 2012)	Sensitivity: 94.2% (95% CI 87.1% to 97.5%) Specificity: 70.8% (95% CI 50.8% to 85.1%) No indeterminates	NE	NE
Khalil *et al.*, 2013[53]	Comparison of QFT-GIT and the TST for the diagnosis of pulmonary active TB	50 pulmonary active TB cases [mean age 41.8 years (SD 19.0 years)]	Fauji Foundation Hospital, Rawalpindi, Pakistan (July 2011–January 2012)	NE	Sensitivity: 80%	Sensitivity: 28%
Kim *et al.*, 2013[54]	Retrospective study of QFT-GIT for diagnosis of GUTB	57 patients [mean age 52 years (range 17–88 years)] with clinical or radiological features suspicious of GUTB	Urology clinic in Korea (March 2009–August 2011)	NE	Sensitivity: 63.3% (95% CI 45.5% to 78.1%) Specificity: 59.3% (95% CI 40.7% to 75.5%)	NE

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Lagrange *et al.*, 2013[30]	Prospective study of QFT-GIT compared with the TST for diagnosis of TB, stratified by HIV infection status	2213 patients were enrolled. QFT-GIT was performed for 96 patients [median age 38.0 years (IQR 30.5–42.0 years)] with pulmonary active TB and 180 non-active TB cases. Of the 276 patients, a TST was performed in 53 active TB cases and 82 non-active TB cases	Nine centres in India (January 2006–July 2008)	NE	*HIV positive* Sensitivity: 66.7% (95% CI 48.2% to 82.0%) Specificity: 64.8% (95% CI 50.6% to 77.3%) *HIV negative* Sensitivity: 95.0% (95% CI 75.1% to 99.9%) Specificity: 25.0% (95% CI 10.7% to 44.9%) *Total* Sensitivity: 77.4% (95% CI 63.8% to 87.7%) Specificity: 51.2% (95% CI 39.9% to 62.4%) Results above are from 135 patients in which data were available for both QFT-GIT and the TST Indeterminate rate among the 276 patients: 7%	*HIV positive* Sensitivity: 51.5% (95% CI 33.5% to 69.2%) Specificity: 83.3% (95% CI 70.7% to 92.1%) *HIV negative* Sensitivity: 85.0% (62.1% to 96.8%) Specificity: 57.1% (95% CI 37.2% to 75.5%) *Total* Sensitivity: 64.2% (95% CI 49.8% to 76.9%) Specificity: 74.4% (95% CI 63.3% to 82.4%) Threshold: ≥ 10 mm
Lavender *et al.*, 2013[55]	Retrospective review of clinical records of patients with QFT-GIT results for the diagnosis of active TB	415 QFT-GIT requested, of which 120 were excluded. 295 patients [median age 40 years (range 16–90 years)] with and without HIV infection were analysed	Newcastle upon Tyne Hospitals, UK (clinical records of patients who had QFT-GIT requested between 29 June 2005 and 28 October 2010)	NE	Sensitivity: 71.4% (95% CI 59.3% to 81.1%) Specificity: 81.0% (95% CI 75.5% to 85.6%)	NE

continued

TABLE 74 Summary of studies of IGRAs for the diagnosis of active TB published between 1 January 2013 and 16 March 2016 (*continued*)

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Lei *et al.*, 2013[56]	Case–control study of T-SPOT.*TB* for differential diagnosis of intestinal TB and Crohn's disease	88 patients with intestinal TB [mean age 36.2 years (SD 14.1 years)] and 103 with Crohn's disease [mean age 37.0 years (SD 15.7 years)]	Inflammatory Bowel Disease Centre, Zhongnan Hospital, China (2003–11)	Sensitivity: 86% (95% CI 75% to 96%) Specificity: 93% (95% CI 86% to 99%)	NE	Sensitivity: 60% (95% CI 49% to 71%) Specificity: 80% (95% CI 71% to 89%) Threshold: ≥ 10 mm
Liu *et al.*, 2013[57]	Prospective study of T-SPOT.*TB* for diagnosis of pleural TB	168 patients were enrolled, but 70 were excluded because they had no final diagnosis. 98 subjects with pleural effusion and no HIV co-infection were analysed. 55 patients [median age 39 years (range 25–59 years)] had pleural TB and 43 patients [median age 39 years (range 25–59 years)] without pleural TB	Beijing Chest Hospital, China (May 2012–June 2013)	Sensitivity: 92.7% (95% CI 82.7% to 97.1%) Specificity: 62.8% (95% CI 47.9% to 76.0%) Indeterminate rate: 2% (These are results from the analysis of peripheral blood samples)	NE	NE
Lodha *et al.*, 2013[58]	Prospective study comparing QFT-GIT with the TST for diagnosis of intrathoracic childhood TB	362 children [median age 115.5 months (IQR 73–144 months)] were enrolled in a RCT of micronutrient supplementation in children with intrathoracic TB	Two tertiary care hospitals in India (recruitment period not reported)	NE	Sensitivity: 82.0% (95% CI 77.8 to 85.6%)	Sensitivity: 93.0% (95% CI 90.0% to 95.5%) Threshold: ≥ 10 mm
Mahomed *et al.*, 2013[59]	Prospective study screening for active TB in adolescents	6363 adolescents (age range 12–16 years) were screened. 21 had active TB. The TST and QFT-GIT results were available for 5071 and 5524 adolescents, respectively	11 high schools in Worcester, South Africa (2005–7)	NE	Sensitivity: 93.8% (95% CI 78.8% to 100.0%) Specificity: 46.8% (95% CI 36.3% to 57.2%)	Sensitivity: 84.6% (95% CI 61.9% to 100.0%) Specificity: 58.1% (95% CI 49.5% to 66.7%) Threshold: ≥ 10 mm

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Danel *et al.*, 2014[29]	Nested cohort study of QFT-GIT for ruling out active TB in HIV-positive adults	975 adults in Cote d'Ivoire (median age 35 years), 25 with active TB on day 0	Nine clinical centres in Abidjan, Cote d'Ivoire (March 2008–August 2009)	NE	Sensitivity: 88.0% (95% CI 75.3% to 100.0%) Specificity: 66.6% (95% CI 63.6% to 69.6%) Indeterminate rate: 3%	NE
Fei *et al.*, 2014[60]	Case–control study of T-SPOT.*TB* compared with the TST for diagnosis of laryngeal TB	83 patients with laryngeal TB and 52 patients with vocal cord polyps as controls (age range 31–66 years), all without HIV infection	China (August 2007–December 2012)	Sensitivity: 90.4% (95% CI 82.1% 95.0%) Specificity: 92.3% (81.8% to 97.0%)	NE	Sensitivity: 50.6% (95% CI 40.1% to 61.1%) Specificity: 61.5% (95% CI 48.0% to 73.5%) Threshold: ≥ 5 mm
Garazzino *et al.*, 2014[61]	Retrospective multicentre study of children (aged 0–24 months) tested at least once with QFT-GIT and/or the TST for active TB	823 children [median age 13.4 months (range 8.4–18.9 months)], 105 with confirmed TB. Results for both the TST and QFT-GIT were available for 616 children	18 paediatric centres in Italy	NE	Sensitivity: 91.1% Specificity: 98.1% (Results above are for the subset of patients with both the QFT-GIT and TST results) Indeterminate rate in entire cohort with QFT-GIT: 4.3%	Sensitivity: 85.1% Specificity: 97.9% Threshold: ≥ 5 mm
Kim *et al.*, 2014[62]	Prospective study of T-SPOT.*TB* for diagnosis of active TB	134 patients [mean age 55.9 years (SD 20.2 years)] suspected of active TB and 62 healthy adults [mean age 30.6 years (SD 8.5 years)] were consecutively recruited. All had been BCG vaccinated at a very young age. All received T-SPOT.*TB* and 53 had TST results	Hallym University Han-gang Sacred Heart Hospital, Korea (June 2008–June 2010)	Sensitivity: 87.8% (95% CI 74.5% to 94.7%) Specificity: 44.1% (95% CI 34.4% to 54.2%) Specificity in healthy adults: 75.8% (63.8% to 84.8%)	NE	Sensitivity: 70.0% (95% CI 48.1% to 85.5%) Specificity: 48.5% (95% CI 32.5% to 64.8%) Specificity in healthy adults: 40.3% (95% CI 29.0% to 52.7%) Threshold: ≥ 10 mm

continued

TABLE 74 Summary of studies of IGRAs for the diagnosis of active TB published between 1 January 2013 and 16 March 2016 (*continued*)

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Kim *et al.*, 2014[63]	Prospective study of QFT-GIT for diagnosis of miliary TB	44 patients [mean age 64 years (SD 19 years)] with miliary TB	Kyungpook National University Hospital, Daegu, South Korea (September 2009–July 2013)	NE	Sensitivity: 68.2% (95% CI 53.4% to 80.0%) Indeterminate rate: 16%	NE
Park *et al.*, 2014[64]	Retrospective study of the QFT-GIT and TST for diagnosis of smear-negative active PTB	224 sputum smear-negative PTB suspects, 94 confirmed as having active PTB [mean age 46.5 years (SD 20.4 years)] and 130 confirmed as non-PTB [mean age 56.6 years (SD 18.3 years)]. 106 patients received a TST	Gangnam Severance Hospital, Seoul, Korea (October 2007–April 2013)	NE	Sensitivity: 81.9% (95% CI 74.1% to 89.7%) Specificity: 62.3% (95% CI 54.0% to 70.6%)	Sensitivity: 58.1% (95% CI 45.8% to 70.4%) Specificity: 63.6% (95% CI 49.4% to 77.9%) Threshold: ≥ 10 mm
Sauzullo *et al.*, 2014[31]	Prospective study of QFT-GIT in HIV patients with active TB	44 patients infected with HIV [median age 42 years (range 31–62 years)] with active TB	Department of Public Health and Infectious Diseases, Sapienza University, Rome, Italy (September 2008–12)	NE	Sensitivity: 65.9% (95% CI 51.1% to 78.1%) with indeterminates included as negatives Indeterminate rate: 22.7%	Sensitivity: 54.5% (95% CI 40.1% to 68.3%)
Schopfer *et al.*, 2014[65]	QFT-GIT evaluated in children with microbiologically confirmed TB	52,400 children were admitted to hospital, of which 405 children had TB, including 91 children with microbiologically confirmed TB. 81 of these children were tested with QFT-GIT	Jayavarman VII Hospital, a large paediatric referral hospital in Cambodia (July 2005–March 2006)	NE	Sensitivity: 53.1% (95% CI 42.3% to 63.6%)	NE
Wang *et al.*, 2014[66]	Retrospective study of T-SPOT.*TB* for diagnosis of paediatric TB	102 patients with TB, aged ≤ 15 years	China (March 2012–September 2013)	Sensitivity: 58.8% (95% CI 49.1% to 67.9%)	NE	NE

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Wlodarczyk *et al.*, 2014[67]	Study of the QFT-GIT and TST for diagnosis of PTB	126 adult patients admitted with a clinical diagnosis of pneumonia. Of these, 43 were culture positive [mean age 48.6 years (SD 18.2 years)], 37 culture negative [mean age 51.7 years (SD 15.5 years)], and 46 with non-mycobacterial, community-acquired lung diseases [mean age 52.7 years (SD 17.3 years)]	Regional Specialised Hospital of Tuberculosis and Lung Diseases in Tuszyn, Poland (January 2010–June 2011)	NE	Sensitivity (culture positive): 65.1% Sensitivity (culture negative): 55.6% Specificity: 87%	Sensitivity (culture positive): 55.8% Sensitivity (culture negative): 64.9% Specificity: 71.7% Threshold: ≥ 10 mm
Aggarwal *et al.*, 2015[28]	Systematic review and meta-analysis of IGRAs for diagnosis of pleural TB	20 evaluations of T-SPOT.*TB* including 1085 subjects; 14 evaluations of QFT assays including 727 subjects (only one high-quality study; considerable heterogeneity)	China, Taiwan, Egypt, Turkey, South Korea, Italy, South Africa, Norway, Italy, Germany and the Netherlands (included studies published between 2007 and 2014)	NE separately. Results pooled across QFT and T.SPOT.*TB* assays	NE separately	NE
Anwar *et al.*, 2015[68]	Retrospective study of QFT-GIT for diagnosis of active PTB in hospital setting	142 cases of confirmed TB and 226 pneumonia cases in Saudi Arabia (only patients with QFT-GIT result included)	King Abdulaziz Medical City in Riyadh, Saudi Arabia (January 2009–December 2013)	NE	Sensitivity: 74.6% (95% CI 66.1% to 81.7%) Specificity: 76.5% (95% CI 69.9% to 82.2%) Indeterminate rate: 11.4%	NE

continued

TABLE 74 Summary of studies of IGRAs for the diagnosis of active TB published between 1 January 2013 and 16 March 2016 (*continued*)

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Bao *et al.*, 2015[69]	Prospective study of performance of QFT-GIT for diagnosis of TB in children and adults	60 children and 212 adults with suspected active TB. 31 had confirmed TB. HIV-positive patients were excluded	*Children* Shanghai Public Health Clinical Center and Children's Hospital of Fudan University, China (December 2010–August 2011) *Adults* Huashan Hospital of Fudan University (December 2010–December 2011)	NE	*Children* Sensitivity: 83.9% (95% CI 66.3% to 94.6%) Specificity: 88.5% (95% CI 70.2% to 96.8%) *Adults* Sensitivity: 73.7% (95% CI 57.8% to 85.2%) Specificity: 70.4% (95% CI 62.9% to 77.0%) Indeterminate rate: 9.1%	NE
Sali *et al.*, 2015[70]	Retrospective study of QFT-GIT for diagnosis of TB infection or disease in children	621 children with suspected active TB, screened for LTBI or clinically healthy, nationally or internationally adopted children evaluated by a national protocol for immigrants and nationally/internationally adopted children, with or without known history of contact with adult active TB cases; 140 active TB suspects; 19 confirmed TB cases	Paediatric Infectious Disease Unit and Catholic University of the Sacred Heart–A. Gemelli Hospital in Rome, Italy (January 2007–July 2010)	NE	Sensitivity: 87.5% Specificity: 93.6% Indeterminate rate: 4.2%	NE (a TST was performed in less than half of the patients. Hence, TST results were not included in the analysis)
Shin *et al.*, 2015[71]	Retrospective review of clinical records to evaluate use of IGRAs for diagnosing EPTB in suspected cases	418 patients with suspected EPTB in Korea; 324 with confirmed active EPTB. Only 56 had QFT-GIT results	Gangnam Severance Hospital in Seoul, South Korea. (July 2005–June 2012)	NE	Sensitivity: 70.2% (95% CI 56.0% to 81.3%) Specificity: 66.7% (95% CI 35.4% to 87.9%)	Sensitivity: 62.1% (95% CI 42.3% to 79.3%) Specificity: 87.5% (95% CI 52.9% to 97.8%)

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Sun *et al.*, 2015[72]	Prospective study to evaluate utility of T-SPOT.*TB* for diagnosis of paediatric TB in hospital setting	117 children with active TB in China; 413 children with respiratory tract infection	Beijing Children's Hospital (March 2011–June 2014)	Sensitivity: 82.9% Specificity: 96.1% Indeterminate rate: 8.5%	NE	Sensitivity: 67.5% Specificity: 75.3 Threshold: ≥ 10 mm Results were also available for ≥ 5 mm and ≥ 15 mm and combinations of TST with T-SPOT.*TB*
Wong *et al.*, 2015[73]	Retrospective chart analysis for paediatric patients who underwent the QFT-GIT and TST for confirmation of active TB	70 paediatric patients in a population with high uptake of neonatal BCG vaccination; eight children had confirmed TB. 47 children had the QFT-GIT	Intermediate-burden region in Taiwan (January 2008–June 2014)	NE	Sensitivity: 100% (95% CI 63.1% to 100%) Specificity: 97.1% (95% CI 85.1% to 99.9%) Indeterminate rate: 6.4%	Sensitivity: 62.5% (95% CI 24.5% to 91.5%) Specificity: 95.2% (95% CI 76.2% to 99.9%) Threshold: ≥ 10 mm
Xia *et al.*, 2015[74]	Prospective study of the QFT-GIT and TST for diagnosing PTB	300 PTB, 41 disease controls and 59 health community controls were enrolled	Heilongjiang Province and south-east Zhejiang Province, China (May 2010–April 2011)	NE	Sensitivity: 80.9% (95% CI 75.9% to 85.2%) Specificity: 36.6% (95% CI 22.1% to 53.1%) Indeterminate rate: 0.5%	Sensitivity: 79.5% (95% CI 74.3% to 84.4%) Specificity: 36.6% (95% CI 22.1 to 53.1%) Threshold: ≥ 10 mm Results were also available for ≥ 5-mm and ≥ 15-mm thresholds, as well as combinations with QFT-GIT

continued

TABLE 74 Summary of studies of IGRAs for the diagnosis of active TB published between 1 January 2013 and 16 March 2016 (*continued*)

Reference (author and year of publication)	Study design	Population	Setting (recruitment period)	Test		
				T-SPOT.*TB*	QFT-GIT	TST
Uzunhan *et al.*, 2015[75]	Prospective comparison of the QFT-GIT and TST for diagnosis of childhood TB	53 children with TB (16 culture positive); 92 healthy children with no risk factors for TB	Children referred to hospital paediatric clinics in Turkey (recruitment period not reported)	NE	Sensitivity in all TB: 62.3% Specificity: 97.8%	Sensitivity in all TB: 97.8% Specificity: 100%
Azghay *et al.*, 2016[76]	Retrospective analysis of hospital records to analyse contribution of QFT-GIT to TB diagnosis	A total of 395 QFT-GIT assays were performed for suspected TB patients	Jean Verdier Hospital in Bondy, Paris, France (June 2008–June 2011)	NE	Sensitivity: 85% (95% CI 73% to 92%) Specificity: 73.3% (95% CI 68% to 78%) Indeterminate rate: 11.6%	Sensitivity: 78% (95% CI 57% to 91%) Combined test: 92.6% (95% CI 74% to 99%)
Jia *et al.*, 2016[77]	Prospective study to evaluate T-SPOT.*TB* in lymph node TB	405 patients with suspected lymph node TB; 83 with confirmed TB; 282 with TB excluded (21 clinically indeterminate and 19 clinical TB excluded from analyses)	Beijing Chest Hospital, China (July 2011–April 2015)	Sensitivity: 90.4% Specificity: 70.5% Indeterminate rate: 3.8%	NE	NE

EPTB, extrapulmonary tuberculosis; GUTB, genitourinary tuberculosis; IQR, interquartile range; NE, not evaluated; PTB, pulmonary tuberculosis; QFT, QuantiFERON; RCT, randomised controlled trial; SD, standard deviation.

Note

If 95% CIs were not reported and raw data were available, we calculated 95% CIs using the Wilson method.[17,18]

Appendix 12 Key characteristics of patients with indeterminate QFT-GIT and T-SPOT.*TB* results

TABLE 75 Summary of key characteristics studies of patients with indeterminate QFT-GIT and T-SPOT.*TB* results

Characteristic	Test, *n* (%)									
	QFT-GIT					T-SPOT.*TB*				
	Dosanjh category					Dosanjh category				
	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	Total	Culture-confirmed TB	Highly probable TB	Clinically indeterminate	Active TB excluded	Total
All	21	5	6	47	79	12	5	3	37	57
HIV infection status										
Positive	0 (0)	2 (40)	1 (16.7)	20 (42.6)	23 (29.1)	2 (16.7)	3 (60)	0 (0)	28 (75.7)	33 (57.9)
Negative	21 (100)	3 (60)	5 (83.3)	27 (57.4)	56 (70.9)	10 (83.3)	2 (40)	3 (100)	9 (24.3)	24 (42.1)
Diabetes										
Yes	0 (0)	0 (0)	1 (16.7)	4 (8.5)	5 (6.3)	1 (8.3)	1 (20)	2 (66.7)	0 (0)	4 (7)
No	21 (100)	5 (100)	5 (83.3)	43 (91.5)	74 (93.7)	11 (91.7)	4 (80)	1 (33.3)	37 (100)	53 (93)
Immunosuppressive therapy										
Yes	0 (0)	0 (0)	0 (0)	1 (2.1)	1 (1.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
No	21 (100)	5 (100)	6 (100)	46 (97.9)	78 (98.7)	12 (100)	5 (100)	3 (100)	37 (100)	57 (100)
TST status[a]										
Positive	6 (75)	3 (75)	0 (0)	2 (18.2)	11 (44)	6 (100)	1 (100)	2 (100)	1 (14.3)	10 (62.5)
Negative	2 (25)	1 (25)	2 (100)	9 (81.8)	14 (56)	0 (0)	0 (0)	0 (0)	6 (85.7)	6 (37.5)
BCG scar										
Yes	14 (93.3)	4 (100)	4 (66.7)	29 (80.6)	51 (83.6)	10 (100)	3 (75)	3 (100)	21 (67.7)	37 (77.1)
No	1 (6.7)	0 (0)	1 (16.7)	2 (5.6)	4 (6.6)	0 (0)	0 (0)	0 (0)	3 (9.7)	3 (6.3)
Unsure	0 (0)	0 (0)	1 (16.7)	5 (13.9)	6 (9.8)	0 (0)	1 (25)	0 (0)	7 (22.6)	8 (16.7)
BCG vaccinated										
Yes	15 (71.4)	4 (80)	6 (100)	36 (76.6)	61 (77.2)	10 (83.3)	4 (80)	3 (100)	31 (83.8)	48 (84.2)
No	6 (28.6)	1 (20)	0 (0)	11 (23.4)	18 (22.8)	2 (16.7)	1 (20)	0 (0)	6 (16.2)	9 (15.8)

a Test positivity based on the stratified threshold.

Appendix 13 Evaluations of tuberculin skin test

Diagnostic accuracy of tuberculin skin test at different thresholds

At induration thresholds of ≥ 5 mm, ≥ 10 mm and the stratified threshold (≥ 6 mm for unvaccinated and ≥ 15 mm for BCG vaccinated patients), the TST had sensitivities of 95.4% (95% CI 90.9% to 97.8%), 94.8% (95% CI 90.0% to 97.3%) and 93.5% (95% CI 88.4% to 96.4%) in culture-positive and highly probable active TB patients (*Table 76*), respectively. In all non-active TB patients, the specificities for thresholds of ≥ 5 mm, ≥ 10 mm and the stratified threshold were 50.3% (95% CI 42.4% to 58.3%), 55.7% (95% CI 47.7% to 63.4%) and 66.4% (95% CI 58.5% to 73.5%). The TST gave high NPVs at the three thresholds; the NPV at the stratified threshold was 90.8% (95% CI 83.9% to 94.9%). Sensitivities were higher for extrapulmonary TB than pulmonary TB. There were large differences in specificity when analyses were restricted to category 4D non-active TB patients (*Table 76*), but there was very little change in the sensitivities when the analyses were limited to culture-positive TB patients. For further results of sensitivity analyses excluding highly probable (category 2) cases, see *Table 77*.

TABLE 76 Diagnostic accuracy of the TST at different thresholds

	Threshold					
	≥ 5 mm		≥ 10 mm threshold		Stratified[a]	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB						
All TB	146/153	95.4 (90.9 to 97.8)	145/153	94.8 (90.0 to 97.3)	143/153	93.5 (88.4 to 96.4)
Culture-positive TB	98/104	94.2 (88.0 to 97.3)	98/104	94.2 (88.0 to 97.3)	96/104	92.3 (85.6 to 96.1)
Culture-negative TB	44/45	97.8 (88.4 to 99.6)	43/45	95.6 (85.2 to 98.8)	43/45	95.6 (85.2 to 98.8)
Smear-positive TB	23/26	88.5 (71.0 to 96.0)	23/26	88.5 (71.0 to 96.0)	23/26	88.5 (71.0 to 96.0)
Smear-negative TB	106/110	96.4 (91.0 to 98.6)	105/110	95.5 (89.9 to 98.0)	104/110	94.6 (88.6 to 97.5)
Pulmonary TB	42/47	89.4 (77.4 to 95.4)	42/47	89.4 (77.4 to 95.4)	40/47	85.1 (72.3 to 92.6)
Extrapulmonary TB	89/90	98.9 (93.8 to 99.8)	89/90	98.9 (94.0 to 99.8)	88/90	97.8 (92.3 to 99.4)
Specificity for a diagnosis of active TB						
Active TB excluded	75/149	50.3 (42.4 to 58.3)	83/149	55.7 (47.7 to 63.4)	99/149	66.4 (58.5 to 73.5)
Active TB excluded, TST negative, no risk factors for LTBI	16/21	76.2 (54.9 to 89.4)	18/21	85.7 (65.4 to 95.0)	21/21	100.0 (84.5 to 100.0)
Predictive values						
PPV	146/220	66.4 (59.9 to 72.3)	145/211	68.7 (62.2 to 74.6)	143/193	74.1 (67.5 to 79.8)
NPV	75/82	91.5 (83.4 to 95.8)	83/91	91.2 (83.6 to 95.5)	99/109	90.8 (83.9 to 94.9)
Likelihood ratios						
Positive likelihood ratio	–	1.92 (1.63 to 2.27)	–	2.14 (1.78 to 2.57)	–	2.79 (2.21 to 3.51)
Negative likelihood ratio	–	0.09 (0.04 to 0.19)	–	0.09 (0.05 to 0.19)	–	0.10 (0.05 to 0.18)

a According to BCG vaccination status: ≥ 6 mm for unvaccinated and ≥ 15 mm for vaccinated patients.

Note

Sensitivity, specificity and predictive values are presented as percentages.

TABLE 77 Diagnostic accuracy of the TST: sensitivity analyses excluding highly probable (category 2) active TB cases

	Threshold					
	≥ 5 mm		≥ 10 mm		Stratified[a]	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB						
Culture-positive TB	98/104	94.2 (88.0 to 97.3)	98/104	94.2 (88.0 to 97.3)	96/104	92.3 (85.6 to 96.1)
Smear-positive TB	21/24	87.5 (69.0 to 95.7)	21/24	87.5 (69.0 to 95.7)	21/24	87.5 (69.0 to 95.7)
Smear-negative TB	68/71	95.8 (88.3 to 98.6)	68/71	95.8 (88.3 to 98.6)	66/71	93.0 (84.6 to 97.0)
Pulmonary TB	36/42	85.7 (72.2 to 93.3)	36/40	90.0 (76.9 to 96.0)	34/40	85.0 (70.9 to 92.9)
Extrapulmonary TB	49/50	98.0 (89.5 to 99.6)	49/50	98.0 (89.5 to 99.6)	49/50	98.0 (89.5 to 99.6)
Specificity for a diagnosis of active TB						
Active TB excluded	75/149	50.3 (42.4 to 58.3)	83/149	55.7 (47.7 to 63.4)	99/149	66.4 (58.5 to 73.5)
Active TB excluded, TST negative, no risk factors for LTBI	16/21	76.2 (54.9 to 89.4)	18/21	85.7 (65.4 to 95.0)	21/21	100.0 (84.5 to 100.0)
Predictive values						
PPV	98/172	57.0 (50.0 to 64.1)	98/164	59.8 (52.1 to 67)	96/146	65.8 (57.7 to 73.0)
NPV	75/81	92.6 (84.8 to 96.6)	83/89	93.3 (86.1 to 96.9)	99/107	92.5 (85.9 to 96.2)
Likelihood ratios						
Positive likelihood ratio	–	1.90 (1.60 to 2.25)	–	2.13 (1.77 to 2.56)	–	2.75 (2.18 to 3.47)
Negative likelihood ratio	–	0.12 (0.05 to 0.25)	–	0.10 (0.05 to 0.23)	–	0.12 (0.06 to 0.23)

a According to BCG vaccination status: ≥ 6 mm for unvaccinated and ≥ 15 mm for vaccinated patients.

Notes

Sensitivity, specificity and predictive values are presented as percentages. Exclusion of category 2 active TB cases does not affect estimation of specificities. Specificities were included in this table merely for completeness.

Combinations of tuberculin skin test and interferon gamma release assays

The TST and T-SPOT.*TB* had a combined sensitivity of 97.7% (95% CI 93.5% to 99.2%) and specificity of 62.0% (95% CI 53.4% to 69.9%). Results were similar for the combination of the TST and QFT-GIT, with a sensitivity of 96.9% (95% CI 92.4% to 98.8%) and specificity of 59.7% (95% CI 51.1% to 67.8%). *Figure 16* shows sequential likelihood ratios for a testing strategy of a TST followed by either T-SPOT.*TB* or QFT-GIT. A positive result for both the TST and IGRAs gave positive likelihood ratios of 5.80 (95% CI 3.90 to 10.0) for TST followed by T-SPOT.*TB* and 4.77 (95% CI 3.25 to 8.16) for the TST followed by QFT-GIT. When both tests in a combination were negative, the negative likelihood ratios were 0.04 (95% CI 0.01 to 0.09) for the TST followed by T-SPOT.*TB* and 0.05 (95% CI 0.01 to 0 0.11) for the TST followed by QFT-GIT.

In sensitivity analyses excluding highly probable active TB cases (i.e. analyses of only culture-confirmed active TB cases and all non-active TB cases), the sensitivities and specificities of the combinations were largely unchanged. The TST and T-SPOT.*TB* had a combined sensitivity of 97.7% (95% CI 92.1% to 99.4%) and specificity of 62.0% (95% CI 53.4% to 69.9%). For the combination of a TST and QFT-GIT, sensitivity was 96.6% (95% CI 90.5% to 98.8%) and specificity was 59.7% (95% CI 51.1% to 67.8%). As can be seen in *Figure 17*, sequential likelihood ratios were generally similar to those from the analyses including both cultured-confirmed and highly probable active TB cases (see *Figure 16*).

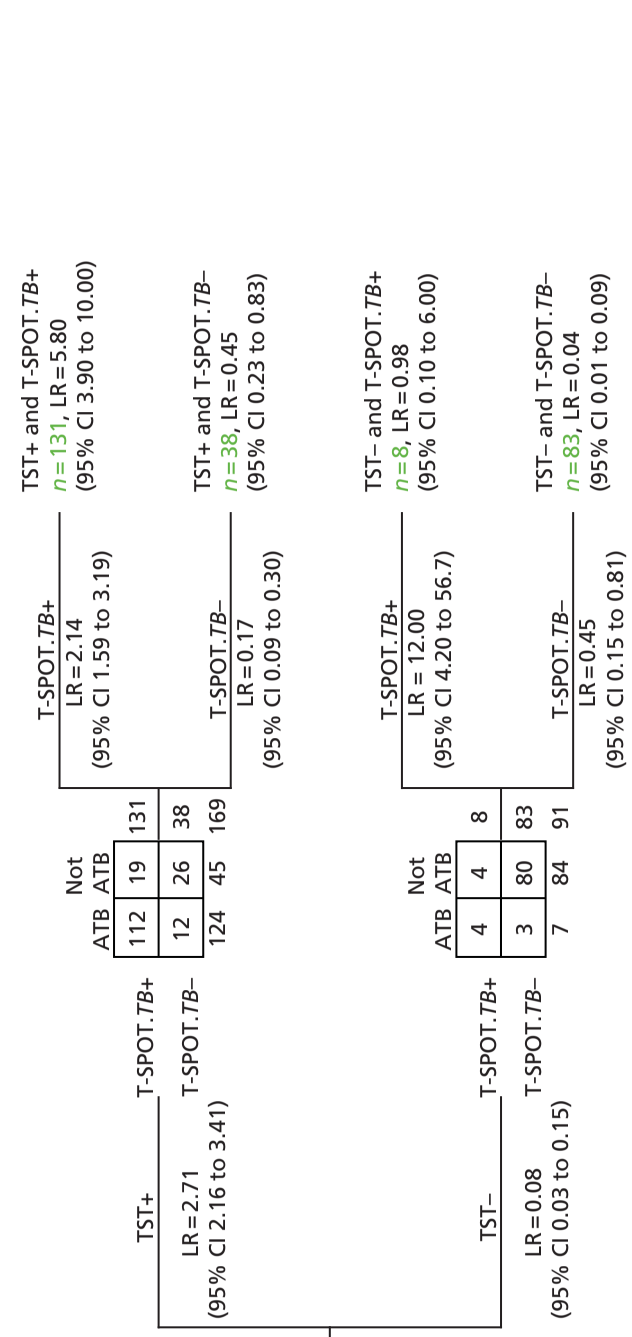

FIGURE 16 Diagnostic performance of combinations of the TST with T-SPOT.*TB* or QFT-GIT. The analyses were based on 260 patients in whom results were available for both the TST and T-SPOT.*TB*, and for both the TST and QFT-GIT. Those with indeterminate IGRA results were excluded from the analyses. The likelihood ratios are presented with 95% CIs. The stratified threshold was used to determine the positivity of the TST. (a) A sequence of testing in which T-SPOT.*TB* follows the TST; and (b) a sequence of testing in which QFT-GIT follows the TST. The difference in the CI of the positive likelihood ratio for the TST is due to stochastic variation in the bootstrap method used for the computation of the likelihood ratios. ATB, active tuberculosis; LR, likelihood ratio. (*continued*)

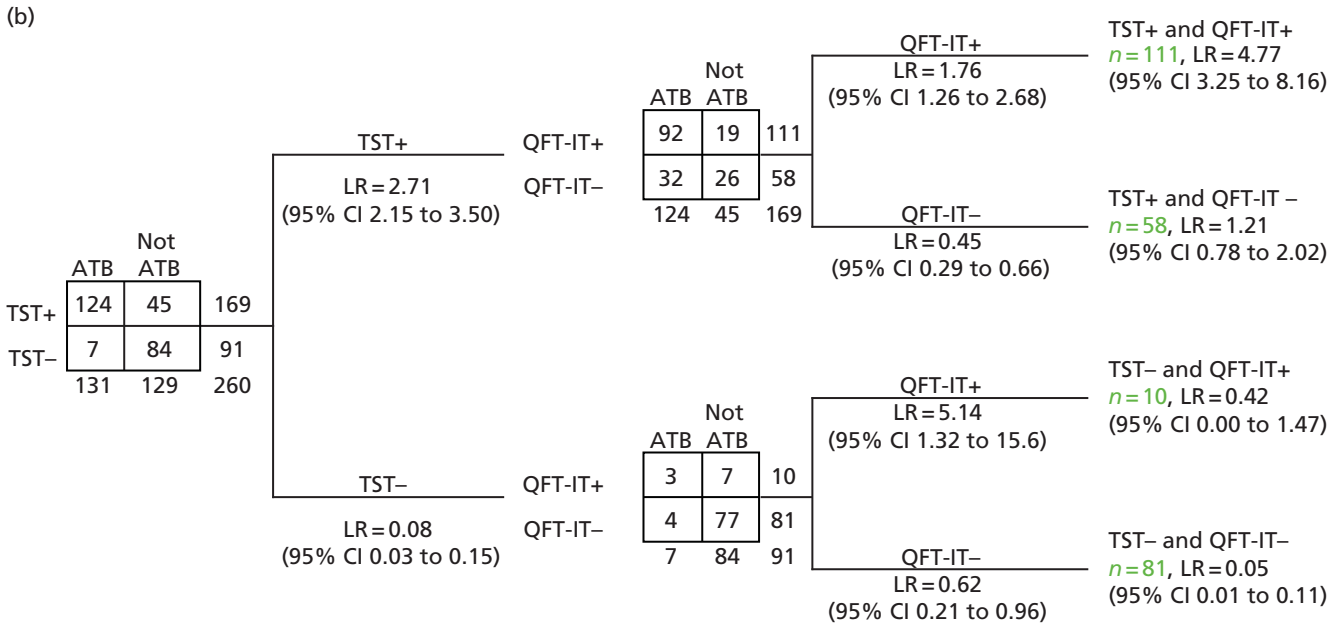

FIGURE 16 Diagnostic performance of combinations of the TST with T-SPOT.*TB* or QFT-GIT. The analyses were based on 260 patients in whom results were available for both the TST and T-SPOT.*TB*, and for both the TST and QFT-GIT. Those with indeterminate IGRA results were excluded from the analyses. The likelihood ratios are presented with 95% CIs. The stratified threshold was used to determine the positivity of the TST. (a) A sequence of testing in which T-SPOT.*TB* follows the TST; and (b) a sequence of testing in which QFT-GIT follows the TST. The difference in the CI of the positive likelihood ratio for the TST is due to stochastic variation in the bootstrap method used for the computation of the likelihood ratios. ATB, active tuberculosis; LR, likelihood ratio.

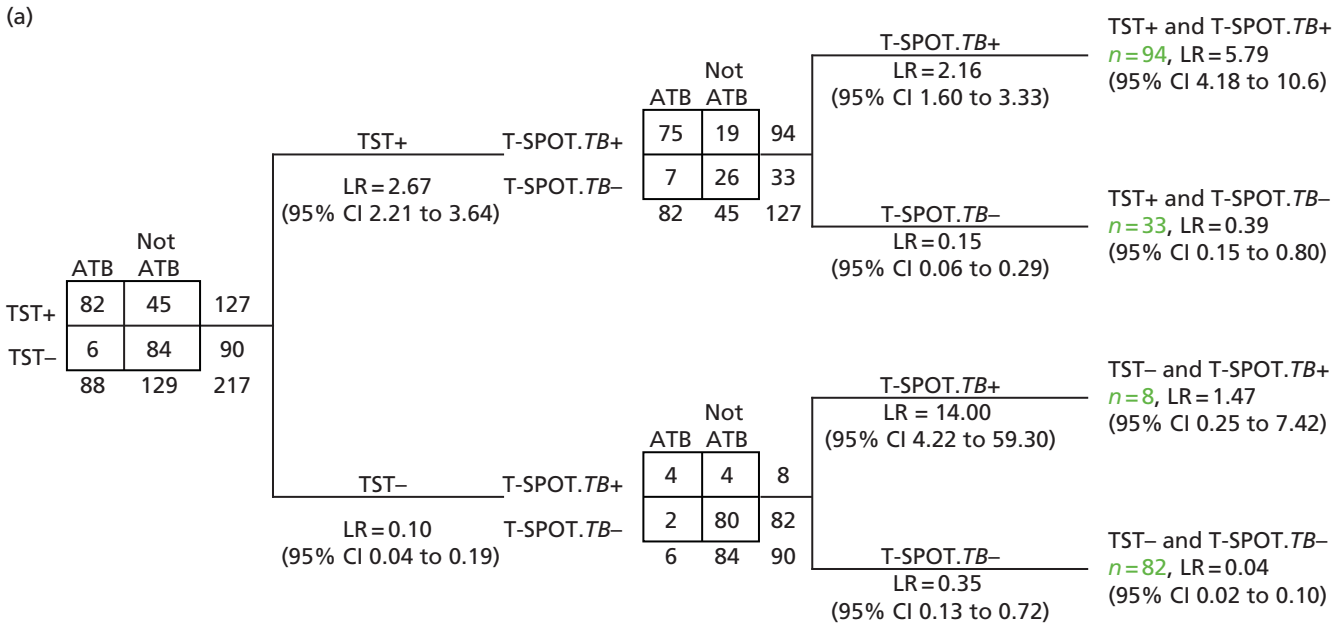

FIGURE 17 Diagnostic performance of combinations of the TST with T-SPOT.*TB* or QFT-GIT: sensitivity analyses excluding highly probable TB cases. The analyses were based on 217 patients in whom results were available for both the TST and T-SPOT.*TB*, and for both the TST and QFT-GIT. Those with highly probable active TB were excluded from the analyses. The likelihood ratios are presented with 95% CIs. The stratified threshold was used to determine the positivity of the TST. The difference in the CI of the positive likelihood ratio for the TST is due to stochastic variation in the bootstrap method used for the computation of the likelihood ratios. (a) A sequence of testing in which T-SPOT.*TB* follows the TST; and (b) a sequence of testing in which QFT-GIT follows the TST. (*continued*)

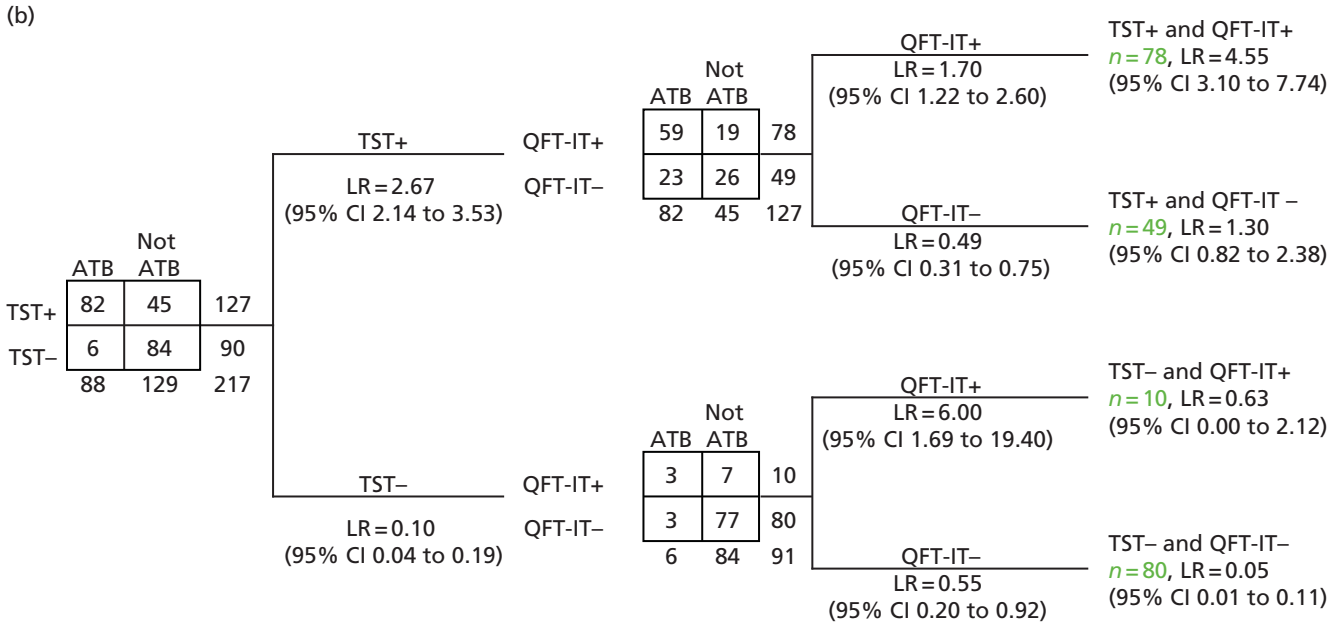

FIGURE 17 Diagnostic performance of combinations of the TST with T-SPOT.*TB* or QFT-GIT: sensitivity analyses excluding highly probable TB cases. The analyses were based on 217 patients in whom results were available for both the TST and T-SPOT.*TB*, and for both the TST and QFT-GIT. Those with highly probable active TB were excluded from the analyses. The likelihood ratios are presented with 95% CIs. The stratified threshold was used to determine the positivity of the TST. The difference in the CI of the positive likelihood ratio for the TST is due to stochastic variation in the bootstrap method used for the computation of the likelihood ratios. (a) A sequence of testing in which T-SPOT.*TB* follows the TST; and (b) a sequence of testing in which QFT-GIT follows the TST.

Appendix 14 Additional results in the human immunodeficiency virus-positive substudy cohort

TABLE 78 Hospital trusts performing T-SPOT.*TB*. The QFT-GIT and/or TST in the diagnostic workup of active TB in the HIV-positive substudy cohort

Hospital trust	Number of tests performed			Number of patients analysed[a]
	T-SPOT.*TB*	QFT-GIT	TST	
Imperial College Healthcare NHS Trust	19	1	4	54
The Heart of England NHS Foundation Trust	3	0	0	6
Chelsea and Westminster Hospital NHS Foundation Trust	24	0	0	40
Royal Free London NHS Foundation Trust	6	3	5	53
St George's University Hospitals NHS Foundation Trust	0	0	0	5
King's College Hospital NHS Foundation Trust	0	0	0	2
University Hospitals of Leicester NHS Trust	0	2	0	23
London North West Healthcare NHS Trust	0	0	3	6
Oxford University Hospitals NHS Foundation Trust	0	0	0	4
Sandwell and West Birmingham Hospitals NHS Trust	0	0	0	1
Barts Health NHS Trust	3	0	0	7
Total	55	6	12	201

a Number of patients included from each centre in the analyses of the IDEA study.

TABLE 79 Cross-tabulation of T-SPOT.*TB* and QFT-GIT results in the HIV-positive substudy cohort

		T-SPOT.*TB*, *n* (%)					
		Positive	Negative	Borderline	Indeterminate	Missing	Total
QFT-GIT, *n* (%)	Positive	19 (65.5)	4 (3.6)	1 (10.0)	5 (11.1)	0	29 (14.4)
	Negative	6 (20.7)	82 (73.9)	9 (90.0)	30 (66.7)	5 (83.3)	132 (65.7)
	Indeterminate	4 (13.8)	24 (21.6)	0	10 (22.2)	1 (16.7)	39 (19.4)
	Missing	0	1 (0.9)	0	0	0	1 (0.5)
	Total	29 (100.0)	111 (100.0)	10 (100.0)	45 (100.0)	6 (100.0)	201 (100.0)

TABLE 80 Diagnostic accuracy of T-SPOT.*TB* and QFT-GIT in the HIV-positive substudy cohort: sensitivity analyses with indeterminate IGRA results included

	Test			
	T-SPOT.*TB*		QFT-GIT	
Test performance	*n*/*N*	Estimate (95% CI)	*n*/*N*	Estimate (95% CI)
Sensitivity for a diagnosis of active TB				
All TB	23/31	74.2 (56.8 to 86.3)	19/32	59.4 (42.3 to 74.5)
Culture-positive TB	13/18	72.2 (49.1 to 87.5)	11/18	61.1 (38.6 to 79.7)
Culture-negative TB	10/13	76.9 (49.7 to 91.8)	8/13	61.5 (35.5 to 82.3)
Smear-positive TB	5/9	55.6 (26.7 to 81.1)	6/9	66.7 (35.4 to 87.9)
Smear-negative TB	14/18	77.8 (54.8 to 91.0)	10/19	52.6 (31.7 to 72.7)
Pulmonary TB	10/13	76.9 (49.7 to 91.8)	9/13	69.2 (42.4 to 87.3)
Extrapulmonary TB	11/14	78.6 (52.4 to 92.4)	7/15	46.7 (24.8 to 69.9)
Specificity for a diagnosis of active TB				
Active TB excluded	101/160	63.1 (55.4 to 70.2)	117/164	71.3 (64.0 to 77.7)
Active TB excluded, TST negative, no risk factors for LTBI	42/66	63.6 (51.6 to 74.2)	54/67	80.6 (69.6 to 88.3)
Predictive value				
PPV	23/82	28.0 (19.5 to 38.6)	19/66	28.8 (19.3 to 40.6)
NPV	101/109	92.3 (86.2 to 96.2)	117/130	90.0 (83.6 to 94.1)
Likelihood ratios				
Positive likelihood ratio	–	2.01 (1.51 to 2.69)	–	2.07 (1.42 to 3.01)
Negative likelihood ratio	–	0.41 (0.22 to 0.75)	–	0.57 (0.37 to 0.88)

Note
Sensitivity, specificity and predictive values are presented as percentages. Indeterminate IGRA results were included these analyses as test positives.

TABLE 81 Comparison of T-SPOT.*TB* and QFT-GIT in the HIV-positive substudy cohort: sensitivity analysis with indeterminate IGRA results included

Test	Number of test results[a]	Sensitivity (95% CI)	Number of test results[b]	Specificity (95% CI)
T-SPOT.*TB*	31	73.4 (55.4 to 86.0)	160	63.2 (55.4 to 70.3)
QFT-GIT	32	59.4 (41.6 to 75.0)	164	71.4 (64.0 to 77.8)
Ratio (95% CI);[c] *p*-value	–	1.23 (0.94 to 1.63); 0.1	–	0.88 (0.77 to 1.03); 0.1

a Number of test results among those with active TB.
b Number of test results among those without active TB.
c The ratio of the sensitivity (or specificity) of T-SPOT.*TB* to that of QFT-GIT. The natural outputs from GEE models are odds ratios. Ratios of sensitivities (relative sensitivity) and ratios of specificities (relative specificity) were computed post estimation of the models. CIs were obtained using the delta method.

Note
Sensitivities and specificities are presented as percentages.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library