

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[†]

¹Institute of Applied Health Research, University of Birmingham, Birmingham, UK

²Tuberculosis Research Centre, National Heart and Lung Institute, Imperial College London, London, UK

³Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, UK

⁴National Institute for Health Research (NIHR) Health Protection Research Unit in Respiratory Infections, Imperial College London, London, UK

⁵NIHR Health Protection Research Unit in Modelling Methodology, Imperial College London, London, UK

⁶Medical Research Council (MRC) Centre for Outbreak Analysis and Modelling, Imperial College London, London, UK

⁷Department of Infectious Disease Epidemiology, Imperial College London, London, UK

⁸Modelling and Economics Unit, Centre for Infectious Disease Surveillance and Control, Public Health England, London, UK

⁹Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

¹⁰St 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*

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

Scientific summary

Interferon gamma release assays for diagnosing active tuberculosis

Health Technology Assessment 2019; Vol. 23: No. 23

DOI: 10.3310/hta23230

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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 smear microscopy, whereas approximately 20% and 25% of non-active TB patients were given a culture test and sputum smear 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.

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@nhr.ac.uk

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

© Queen's Printer and Controller of HMSO 2019. This work was produced by Takwoingi *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

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

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Honorary Professor, University of Manchester, and Senior Clinical Researcher and Associate Professor, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Director, NIHR Dissemination Centre, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

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